DOCUMENT RESUME

ED 269 443 TM 860 277

AUTHOR Pena, Deagelia M.; Henderson, Ronald D.
TITLE Sampling Procedures Used for National Surveys of Public School Teachers--Problems and Possible Solutions.
PUB DATE Apr 86
NOTE 51p.; Paper presented at the Annual Meeting of the 'merican Educational Research Association (70th, San Francisco, CA, April 16-20, 1986).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Adults; Error of Measurement; Longitudinal Studies; *Measurement Techniques; *National Surveys; Opinions; *Public School Teachers; *Research Design; *Research Methodology; Resear n Problems; Sample Size; *Sampling; Statistical Bias; Statistical Studies; Teacher Attitudes; Validity
IDENTIFIERS Sampling Error

ABSTRACT

The sampling of teachers for nationwide surveys offers a challenging endeavor in obtaining a representative and adequate sample to truly represent opinions of the teachers. Ten national surveys of public school teachers conducted betwee" 1980 and 1985 are presented with respect to their sampling design and procedures. Concepts and theoretical considarations provide the background of the critical review. This paper discusses adequacy and representativeness as criteria of a good sample, sample design, bias, sampling error, sampling frame and unit, multistage and stratified sampling, and sample size allocation. Problems associated with prevailing sampling practices include: (1) lack of sampling frame or list of all public school teachers from which a sample may be drawn; (2) the question of assurance that the sample drawn is representative of the target population of public school teachers; (?) the theoretical connection between a sample design and the computation of estimates from the results; (4) the recognition and knowledge of the quantitative relationship between the sample design and the precision of estimates; and (5) reporting results that demonstrate validity, and willingness and ability to sort out the strengths and weaknesses of the survey. Solutions to these interrelated problems are discussed. (PN)

SAMPLING PROCEDURES USED FOR NATIONAL SURVEYS OF PUBLIC SCHOOL TEACHERS -- PROBLEMS AND POSSIBLE SOLUTIONS

BY

DEAGELIA M. PENA AND RONALD D. HENDERSON
NATIONAL EDUCATION ASSOCIATION

PRESENTED AT THE
AMERICAN EDUCATIONAL RESEARCH ASSOCIATION
ANNUAL CONFERENCE
IN SAN FRANCISCO, CALIFORNIA

APRIL 18, 1986

ACKNOWLEDGEMENTS

The authors wish to thank the following persons whose help and contribution by way of important information and suggestions had greatly enhanced this study:

Susan M. Falsey, National Education Association

Charles H. Hammer, Center for Statistics, U.S. Department of Education

Alton B. Sheridan, National Education Association

The Educational Reform movement has thrust education into the national spotlight resulting in an onslaught of innovation in education. One facet of this increased attention in education is a surge in the number of national surveys of teachers (such as those reviewed in this paper). Basically, national and state level policy makers are at times using information from surveys as component of the policy-making process. One major concern however, is whether reported results of teacher surveys truly represent the opinions of the teachers. This concern relates to the representativeness and adequacy of the sample and the application of appropriate sampling techniques and design.

The sampling of teachers for nationwide surveys offers one of the most challenging endeavors in obtaining a representative and adequate sample and in designing surveys that will yield a high degree of precision of estimates. The inherent problems in sampling teachers are the availability of home telephone numbers, the difficulty of accessing teachers through the districts and the absence of nationwide reliable data sets from which samples may be drawn.

PURPOSE AND OBJECTIVES

The purpose of this study was to examine critically the practices and procedures of sampling in national surveys of teachers conducted by at least five organizations which reported or published results in 1980 to 1985. More specifically, the objectives were:

(1) To conduct a critical review of the sampling techniques and sampling designs of five or more national surveys of teachers;

(2) To identify pragmatic, procedural, and theoretical problems associated with the sampling practices prevailing in published teacher surveys;

(3) To explore the implications of good and bad samples to educational policies and decisions; and

(4) To generate solutions and recommendatons that will improve the representativeness and adequacy of teacher samples, and thereby increase accuracy and precision of estimates.

THEORETICAL CONSIDERATIONS

Adequacy of the sample and design has direct bearing on the precision of estimates. Precision is measured by the sampling error which is a function of the variance of estimate of the population parameter (say, mean or proportion) which in turn is a function of the sample size. Thus, the variance of the sample proportion p is

(1) $$S_p^2 = \frac{pq}{n}\left(1-\frac{n}{N}\right),$$

where $q = 1-p$, and $1-\frac{n}{N}$ is the finite population correction (f.p.c.). The standard error, S_p the square root of the right handside of equation (1), is

(2) $$S_p = \sqrt{\frac{pq}{n}\left(1-\frac{n}{N}\right)}.$$

Note that, in sampling from a large population such as in national teacher surveys, n/N would be very small, making f.p.c. very close to one.

So for purposes of discussion, let us drop the f.p.c. and use

(3) $$S_p = \sqrt{\frac{pq}{n}}.$$

Now, the sampling error is the amount subtracted from, and added to the sample proportion, to make an interval estimate. This amount is S_p multiplied by a factor z, the value of which is associated with the confidence level. Three confidence levels and their associated z's are given below:

Confidence Level	z
90%	1.65
95%	1.96
99%	2.58

The interval estimate of the population proportion is usually in this form:

(4) $$p - zS_p < P < p + zS_p \cdot$$

We can see in formula (3) that the larger our sample size n, the smaller S_p becomes; and the smaller S_p is--applied to equation (4)--the smaller the interval. The smaller this interval is, the higher the precision of estimate. Thus, a large sample increases the precision of estimate. The question then is how large a sample should one have? From equation (3) we can solve for n and have

(5) $n = \frac{pq}{S_p^2}$.

To find n we must have values for S_p and p. This is where speculation begins. Suppose I want a precision of estimate, of $\pm$ 3%. It means that in formula (4) I want zS_p to be equal to .03. But z is determined by the confidence level desired. So, at the 95% confidence level $z = 1.96$. Hence

$$zS_p = .03 \quad \text{or}$$

$$1.96S_p = .03 \quad ; \quad S_p = .0153.$$

To ensure that the sample is "adequate" we want the numerator as large as possible.

Now the product pq, or p(1-p), for combinations of p and q such that p + q = 1.0, is largest when p = 0.5. So if we use this value, we are assured that the sample size n would result in a sampling error no higher than 3%. So we substitute p = 0.5 and S_p = .0153 in equation (5) and obtain

$$n = \frac{0.5 \times 0.5}{(.0153)^2} ;$$

$$n = 1068.$$

Thus n = 1068 is an "adequate" sample since it satisfies the maximum sampling error tolerable with the degree of confidence desired. It is presumed that these formulas and procedure are well known to most of the participants in the session, and tables for sample sizes are available in some books. However, this discussion reminds us that there is an explicit procedure to determine exactly how large n should be to be adequate.

Attachment 1 (How to Determine the Sample Size...) is supplied as a handy tool for a quick estimation of n. Our result n = 1068 can also be confirmed by using McCall's (1980) Table H7 (Confidence Level, 95%), p.354, for a population size of 2,000,000.

The <u>design</u> of the sample, on the other hand takes into consideration practical problems--such as the feasibility and efficiency of reaching the sampling units (e.g., the teachers), the cost, and, of course the precision of estimates. Stratified and multi-stage sampling, or a combination of these two designs are examples. These designs are discussed in the latter part of this paper.

Another important consideration, is the representativeness of the sample. The representativeness of the sample affects the degree of accuracy of estimates. A truly random sample is assumed to be representative of the population from which it is drawn. However, it is a fact that in many national surveys, representativeness should become a matter of concern since most surveys depend on a list of teachers which is not complete or up to date, for the drawing of a sample. In fact, a complete list is not available.

The key issue in this paper is that the measure of precision would be meaningless if applied to a sample that is not representative of the population. A secondary issue is that in cases where the sample was representative, the design could still fall short of optimizing the precision of estimates within certain constraints such as cost.

Representativeness and adequacy are then the two basic criteria of a good sampling procedure (Deming, 1966). To get to the fundamental concepts behind these criteria, we need to have a clear notion of a few concepts and recognize the basic problems of sampling. We need to be able to answer such questions as: When is a sample adequate? What factors affect the representativeness of a sample?

The sections that follow intend to clarify these concepts and problems.

Bias and Sampling Error

The sample can be designed to achieve any desired degree of precision. One may wish to estimate a proportion or a mean within 1% or 2% of what would be obtained had the same procedure been applied to the total population. This precision of estimate should not be confused with the validity of results for purposes of prediction. There is a point in conducting polling surveys to forecast election results, as close as possible to the voting time. It should not also be confused with inaccuracies resulting from bias due to non-response and inaccuracies due to nonsampling errors.

Bias results from lack of randomness, e.g., due to incomplete returns or to oversampling one group of respondents, say women. Sampling error under conditions of randomness is a measure of the degree of tolerance of departure of the results from the central measure, with an associated probability. Thus, the smaller the sampling error, the higher the precision. The four possible combinations of the degrees of bias and sampling error are illustrated in Attachment 2.

The generalization of findings from the sample to the population is legitimate only in as far as the sample is properly drawn. This is why the sampling design or plan is very important. In national surveys of public school teachers, the initial critical problem is the source of sampling frame. The frame is merely a list of sampling units (e.g. public school teachers) representing the population being studied (Parel, 1978).

In national surveys of public school teachers, the population is approximately 2.4 million teachers in the United States. In sampling, a distinction, however, is made between two definitions of "population."

1. The target population. This is the population for which representative information is desired. Thus, the target population in this discussion is the 2.4 million public school teachers in the United States.

2. The sampling population. This is the population from which a sample will actually be drawn as determined by the sampling frame. The sampling frame is the list of sampling units (e.g., teachers, or in multistage sampling, school systems, schools). If one wishes to know the teachers' opinion on competency testing, the target population could be all K-12 teachers teaching in September 1985. However, if a list was created, and, for various reasons some of these teachers were left out, the remaining teachers would be the sampling population. The list created from this sampling population is the sampling frame. A teacher whose name is in the list (frame) is then a sampling unit.

As we will see later, the sampling frame used by most well known survey organizations is the list compiled by Market Date Retrieval (MDR) of Westport, Connecticut. According to one source in this study, MDR has the most complete list but it includes only about 68% of the 2.4 million teachers. Further, there is no way to get a full list of teachers.

The acceptability of using the sample result to give information about the target population should depend on the reliability of the sampling frame. It follows then that some kind of evaluation of this list is necessary to determine whether it is, in fact, representative of the total population.

Multistage and Stratified Sampling

Some surveys critically reviewed for this paper used multistage sampling at two or three stages, instead of using the MDR list. Because a list of school systems and a list of schools within the system are available. the sampling frame consists of the lists of school systems across the different strata. The sampling unit is the school system. Stratified sampling may also be incorporated in the design.

One example is the stratification of school districts according to enrollment size. One of the organizations surveyed used the stratification shown in Table 1.

Table 1. Stratification of School Districts by Enrollment Size

Strata (enrollment size)	No. of School Districts
100,000 and over	22
50,000-99,999	47
25,000-49,999	118
12,000-24,999	347
6,000-11,999	926
3,000- 5,999	1,856
1,200- 2,999	3,475
300- 1,199	4,836
100- 299	4,004
Total	15,631

A discussion of the related sampling procedure is in the section on the critical review of national surveys conducted these past five years.

The rationale behind the multistage sampling of public school teachers, is that in the absence of a reliable sampling frame, the teachers could be sampled (reached) through their districts.

Why stratify? Generally, stratification improves the precision of estimates. Usually the stratification variable (e.g. enrollment size) is related to the survey variables. For instance, teachers in large school districts may have a different opinion about discipline than teachers in small school districts. In surveys dealing with continuous variables such as household income or age, stratification solves the problem of variability of these characteristics by creating homogeneous subpopulations (strata) before sampling is done. That is, stratification minimizes variances within strata. Stratified random sampling then consists of selecting a simple random sample from each of the strata into which the population has been subdivided (Parel, 1978). The frame would then refer to the lists of teachers from the various strata. On the other hand, stratified sampling with probability proportional to size (pps) uses a measure of size, such as the number of teachers, to determine the probability of selection of a school district.

Estimation from Stratified Sampling

It is difficult to talk about sampling procedures and techniques, without discussing estimation. In determining the precision of an estimate, the variance of the mean is the key parameter in computing the sampling error.

To obtain a measure of precision of the estimate of the population proportion, an estimate, of the variance of the sample proportion is computed by the equation (Parel, 1978)

(5) $$v(p_{st}) = \sum_{h=1}^{L} \frac{N_h^2}{N^2} \left[\left(\frac{N_h - n_h}{N_h} \right) \frac{p_h(1 - p_h)}{(n_h - 1)} \right],$$ or

(6) $$v(p_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} \left[N_h(N_h - n_h) \frac{p_h(1 - p_h)}{(n_h - 1)} \right]$$

where

L is the number of strata
h is the stratum number
N is the population total
N_h is the number of sampling units in the hth stratum
n_h is the sample size in the hth stratum
p_h is the proportion estimate of the population proportion in the hth stratum

$\left(\frac{N_h - n_h}{N_h}\right)$ is the finite population correction for the hth stratum.

Determination and Allocation of the Sample Size

In stratified sampling, the first step is the determination of the overall sample size (n). The second step is allocating n among the L strata. References on sampling methods discuss in detail several methods of allocation. For instance, when the sampling units N_h in the L strata are more or less the same in number, the method is equal allocation. When the total number of sampling units N_h vary from stratum to stratum, sampling is proportional to size.

Thus, in equal allocation,

$$n_h = \frac{n}{L};$$

in proportional allocation,

$$n_h = \frac{N_h}{N} . n .$$

In other cases, allocation is neither equal nor proportional. For instance, in cases where one of the strata needs special analysis, then one may need a sample larger than the size allocated by either the proportional or the equal allocation. Whenever such disproportional unequal allocation is used, adjustments in the overall estimates are made by appropriate weighting.

Sampling Problems

Deming (1966) very aptly placed sampling design within the context of survey design. In designing a survey the basic questions to be answered are:

- What questions should be asked?
- What precision is needed?
- How can the survey best be carried out to provide the information desired, with the desired precision and no more?
- What will the survey cost?
- What do the results mean?
- How can objective measures of sampling errors and biases be obtained, so that the reliability and meaning of the results can be assessed and methods improved?

Thus, in the design of surveys, the major problems of sampling are:

o Specifying the reliability to be achieved, in consideration of the budget and other constraints. How is precision going to be mesured? Would one use coefficient of variation of 1%, 5%, and 15% in some important characteristc? Would this measure of precision require measurements of the differences between various procedures (two or more different ways of stating the questions, different approaches in interviewing, different definitions, etc.) to achieve better data interpretation?

Or is one concerned mainly with estimates of proportions and would use precision in terms of ±3%, ±5%, with 95% confidence for the interval of estimate?

o Designing the survey, (a) so that it will produce the desired precision at the lowest possible cost and (b) with the personnel and other resources available. The design should consider the following pointers:

(1) Consider minimizing non-response by eliminating factors contributing to non-response, such as those related to questionnaire clarity, format and length, interviewing style, and time of day. How much non-response can be tolerated? How much is enough to impair the survey seriously? Can the first wave of response be improved or corrected by interviewing a sample of the people not responding at the first interview or not returning their questionnaires by mail?

(2) If necessary, lay out a few alternative sample designs to show approximately what the costs will be for various degrees of precision;

(3) Determine the maximum allowable sampling error.

o <u>Appraising</u> the precision actually attained in several important characteristics, and evaluating the differences between the various procedures specified for comparison--for instance, an interpenetrating sample in a telephone survey, using the same questions.

Suppose the non-response rates in the two samples are the same, and responses are similar, one may decide to use the cheaper of two methods in the future. On the other hand, suppose the responses to some items or the characteristics of the respondents in the two groups differ, which responses more accurately represent the target population?

It is interesting to remind ourselves that without the theory of probability there would be no way to specify the precision prior to the survey and to evaluate it after the survey is completed.

HOW THE STUDY WAS CONDUCTED

This paper is a critical study of published or reported surveys. The "data" in consideration are the information on sampling procedures used in the surveys. The numeric results of these surveys were outside the primary concern of this study. Eight national surveys of public school teachers were reviewed, in the area of sampling procedure. Four of these surveys were conducted by major survey organizations or agencies, and four by organizations for special purposes such as subject-matter-oriented organizations or minor organizations with limited goals.

Exceptions to reviewing only surveys of all public school teachers were national surveys conducted by two teachers' organizations, where the target population consisted of public school teachers who were members of their respective organizations.

In most of these surveys the description of the sampling procedure lacked enough details to allow an in-depth critical review. In some cases, though, the brief description or the way results were reported was sufficient to raise questions about the sampling procedure. In these instances, the description is quoted verbatim before the critical commentaries so that the reader can formulate his judgment independent of this review. Whenever possible, the source of the publication was contacted usually by telephone. The survey persons were usually very cooperative and many of the clarification questions were answered. In addition, we were able to get information on the practical difficulties encountered in obtaining a good sample.

In one instance where the survey is still in progress and the report had yet to be written, the write-up on the sampling design was mailed to us. This was one survey that demonstrated how a survey/sampling design was carefully delineated and documented!

CRITICAL REVIEW OF TEN SURVEYS

The 10 survey organizations are not identified in this paper. They are named A to J. In this section, each survey organization's sampling procedure will be presented. The description will be followed by a critical commentary.

It will be evident that the survey designs especially those of large organizations hinged on how to reach that sample of teachers across the country, and achieve an acceptable level of precision of estimates and representativeness. The effort to achieve these goals (though not explicitly stated as goals) of precision and representativeness already deserve some credit, as they would pave the way to improvements on design and implementation in future surveys.

Most of the surveys depended on some available lists or existing data sets compiled within a given year. One list used by at least two organizations was that compiled by Market Data Retrieval of Westport Connecticut. Two organizations on the other hand used stratified multistage sampling and teachers were randomly selected at the final stage. Ironically, these intricate designs where much can be learned were not published; apparently, the details which were highly technical were not intended for the general reader of the report or publication.

As a summary, the ten surveys are presented in a matrix of the ten surveys across the major characteristics of the sampling design. The succeeding sections present these 10 surveys (not in alphabetical order).

Survey A

This survey is a mailed questionnaire survey. A "representative sample" of public school teachers was obtained from a list provided by Market Data Retrieval.

> "A sample of 2,000 teachers was selected to reflect the total national population of teachers." The sample was stratified proportionately by region and by teaching level."

The report contained a detailed accounting of the original 2,000 questionnaires, which resulted in a 41% or 813 completed returns. Sampling error was not published but, as computed for this paper, was $\pm$ 3.4% at the 95% confidence level.

An effort was made in Survey A to determine representativeness that may have been affected by incomplete returns. A telephone survey of 100 teachers who had not answered the mailed questionnaire was conducted and the results showed that the sample of nonrespondents "closely paralleled" the responding group both in terms of attitudes and socio-economic and demographic characteristics. We analyzed this claim of representativeness as reported. In one of the attitude questions, for instance, forty-three percent of the mailed-questionnaire group (with n = 813) answered "yes" while 40% of the telephone group (with n = 100) answered "yes". How similar are these two groups? Computations of sampling error resulted in an interval estimate of 40 to 46% and 33 to 53% at the 95% confidence level, for the mailed survey and the telephone groups respectively. This analysis confirmed the report's statement on representativeness, although the small sample size of 100 resulted in a large sampling error. A test of difference between two proportions will produce the same result--that there is no significant difference in proportions between the two groups.

It appears that the design was given careful consideration with respect to precision and representativeness. The remaining question pertains to the sampling frame. How representative and up to date is this MDR list which comprised only 68% of the total population of teachers? A minor point, in addition is one statement that may need clarification: "The sample was stratified proportionately by region and by teaching level." According to this statement, stratification was made after the sample was drawn. Normally, stratification is done before drawing the sample. However, this does not preclude the need to do otherwise for practical reasons. If the sample before stratification was a self weighting simple random sample where each teacher had the same probability (sampling fraction) of being selected, then a post-sampling stratification would still approximately be proportionate by region and level. By self weighting sample is meant that, since the probability of selection was applied to each teacher in the list the sample will end up, more or less, proportinately representing the various categories into which the sample may be broken down.

Survey B

This survey also selected its sample from the list provided by MDR. There were 4,822 in the original sample, of which 1846 were completed personal interviews. Again, there is a vague statement here: "Sample sizes for completed interviews were set for each state based on the proportion of elementary and secondary public school classroom teachers in each state." If this is so, were some selections discarded to meet the known proportion? Or does the statement mean that the original sample allocations were proportional to the numbers of teachers in the two levels - elementary and secondary in each state. The phrase "based on the proportion of elementary and secondary public school classroom teachers in each state" is also not clear, whether "proportion" refers to the state's proportion based on the US total; or whether it meant separate proportions of secondary and elementary teachers. Sampling error was not published but as computed for this paper was ±2.2% at 95% confidence level.

It was never the purpose of this paper to critique the reporting language. However, an appraisal of the design is only as good as the understanding of what actually happened in the process.

One indication of representativeness is the 32% male in the sample (591/1846) close enough to the known percentage of 33.1% in 1981 published by the National Center for Education Statistics (Grant, 1984).

Survey E

The sampling frame was not described nor the source named. According to the report, one in 700 K-12 teachers were drawn from the population of 2.1 million public school K-12 teachers. Two follow-ups resulted in a 45% response or 1346. Arithmetical calculations showed that the original sample size must have been 3000 (1/700 of 2.1 million), and the completed return of 1346 was approximately 45% of 3,000.

Sampling errors ranged from 1.4% to 22.4% for the overall sample proportions. The survey article presented the table for the 90% confidence limits for subgroup sizes 100 to 1299. This table is shown here as Table 2.

Table 2 Confidence Limits for Percentages

APPROXIMATE NUMBER OF PERCENTAGE POINTS TO BE ADDED TO AND SUBTRACTED FROM THE OBSERVED SAMPLE PERCENTAGES TO OBTAIN THE 90 PERCENT CONFIDENCE LIMITS FOR THE POPULATION PERCENTAGES

	Observed percentage near				
Size of subgroup	10% or 90%	20% or 80%	30% or 70%	40% or 60%	50%
100-199	5.0	6.6	7.6	8.1	8.3
200-299	3.5	4.7	5.3	5.7	5.8
300-399	2.9	3.8	4.4	4.7	4.8
400-499	2.5	3.3	3.8	4.0	4.1
500-599	2.2	2.9	3.4	3.6	3.7
600-699	2.0	2.7	3.1	3.3	3.4
700-799	1.9	2.5	2.9	3.0	3.1
800-899	1.7	2.3	2.7	2.9	2.9
900-999	1.6	2.2	2.5	2.7	2.7
1,000-1,099	1.6	2.1	2.4	2.5	2.6
1,100-1,199	1.5	2.0	2.3	2.4	2.5
1,200-1,299	1.4	1.9	2.2	2.3	2.4

The design was apparently a simple random sample from a national list.

The survey indicated that the representativeness of the survey sample returns was tested by comparing the "basic demographic data with" the few other known sources of data.

Though the information on sampling procedure is not complete, there is every indication that representativess and precision were taken seriously into consideration. The weakness, again is the sampling frame. How reliable and up to date was this sampling frame?

Survey F

Nothing can be said about this survey, because no mention was made of the sampling procedure or size. The article started by saying that nearly two-thirds of U.S. teachers endorse..." The next paragraph described the survey as a "nationwide, statistically representative sample of U.S. teachers..."

Such statements--as "A clear majority, 62% of teachers agree..." or "nearly two-thirds of the U.S. teachers endorse..."--imply exact percentages of the 2.1 million teachers. We know it is not so.

In another section, the article states that "62.7 percent of the teachers who responded to our article agree that..." This survey did not discuss how the sample was obtained. Whether the sample size was 100 or 10,000, it is not known.

The article mentioned the availability of a full report at a certain price. Since there is a full report which we had not examined, there is no commentary on the sampling procedure. The percentages, however, were reported in sweeping statements which to our judgment was misleading, especially if no sampling procedures or errors of estimates were discussed.

Survey G

The organization surveyed 252 elementary teachers "from all geographic regions of the U.S."

How the teachers were sampled was not clear in the report. To make the published figures work, some guesses had to be made. Thus, the appraisal of this procedure was based on the best interpretation one could make of the sampling procedure as published.

According to the report 100 elementary school principals were "randomly identified" by the ______________(name of Association) from a cross section of their membership. "Each principal was mailed 10 copies of the questionnaire and was asked to distribute them to 10 elementary teachers ..."

There were problems of interpretation, obviously. The phrases subject to question were:

"randomly identified".
"cross section of their membership."

The random selection, if at all it was one, was left to persons who may not know or care about sampling. There was no mention of instructions given for selection.

The questions involved are:

- How random was the principal's selection?
- How were the principals selected?
- Does this imply a two-stage simple random sample?
- If the number of teachers was constant (10) in each sample school, what were the probabilities of selection and how were these probabilities taken into account in making estimates?

With regard to the sample size, it appeared that if 10 questionnaires were given to each 100 principals, then the size of the sample was 1,000 to start with. Further assumptions or interpretations were that no follow up was made, and that 252 was the number of respondents--a reasonable 25% if no follow-up was made.

The article attempted to show that the sample was representative, under the section "Description of the Sample" by stating that "teachers were rather evenly distributed from grades K through 6," the group for whom the survey was conducted.

The problems in giving validity to this survey encompass all criteria of a good sample design. In summary:

- The sample size 252 "from all geographic regions" is too small to yield reasonably precise estimates.
- "geographic regions" were not defined.
- A two-stage sample was implied but not explicitly described.
- There was no mention of probabilities of selection.
- An "even distribution of teachers" across grades K to 6, regardless of what it meant, does not necessarily indicate representativeness.
- Sampling errors were completely left out and the source of sampling frame was not mentioned.
- The article left the impression that selecting 100 schools (through principals) and sending 10 questionnaires to each principal was, apparently, a design for practical convenience, rather than for sampling purposes.

Survey H

The article indicated that the author wrote principals at 500 schools. The principals were given a set of rules to draw four teachers at random.

The 500 schools consisted of 100 private and 400 public schools. It was found out that the article was based on a report yet to be completed. Further information was obtained beyond that given in the article. According to the information obtained from the source of the article, MDR was a source of sampling frame. However, it was not clear how the two-stage sample was drawn--whether the first stage was the school, and what MDR had to do with the sample if the principals did their own selection.

All that can be said is that there was nothing in the article that indicated that a well planned sampling procedure was designed. No sampling errors can be computed. There is a big question on the representativeness of this sample.

Surveys C and D are the last two surveys to be discussed because of their sophisticated stratified multistage sampling procedure that required extensive description and discussion of details.

Enough deta ls of the two designs were discussed in order to clearly present to the readers two different models of stratified multistage sampling of public school teachers.

Survey D

The design was a "self weighting stratified, disproportionate two-stage cluster with unequal sampling rate in the second stage." This design was based on McCall's (1980) description of the process, involving nine steps (pp. 273-277). Table 3 (McCall, 1980, p.276) illustrates the process.

Knowing that there were 2,024,000 teachers in the whole U.S., it was desired to sample 2,025, or .001, or 1/1000 of the total population. With two stage sampling, the probability of selecting a teacher is equal to the probability of selecting a stratum times the probability of selecting a teacher within stratum.

The public school systems in the United States were stratified by enrollment size as shown in table 3. There were nine strata. The source of the list of school systems with information on enrollment size and number of teachers was the data base of the surveying organization supplemented by information from the NCES--National Center for Education Statistics, U.S. Department of Education. The total number of systems within each stratum is shown in the table. The disproportionate sampling rate is shown in column (5) of the table. Notice that all the 22 systems i stratum one were selected for the sample giving (a sampling rate of 1:1. In the next stratum, 27 systems out of the 47 were selected, giving a sampling rate of 1:1.7. The (unequal) sampling rates for teachers are those rates which when multiplied respectively by the sampling rate for systems will give 1/1000. This process, with disproportionate allocations of school systems and unequal sampling rates (for teachers) at the second stage, ultimately resulted in equal probability of selecting a particular teacher, one in 1,000. In summary, we have the following characteristics of this sample design:

Table 3

EXHIBIT 15.2.–DATA USED IN SELECTING A SELF-WEIGHTING STRATIFIED DISPROPORTIONATE TWO-STAGE CLUSTER SAMPLE OF CLASSROOM TEACHERS BY USING A LISTING OF PUBLIC SCHOOL SYSTEMS IN THE UNITED STATES STRATIFIED BY THE NUMBER OF PUPILS ENROLLED, 1979-80

System stratum	Stratum limits (enrollment)	Total number of systems	Number of systems in sample	Sampling rate for systems (1st stage)	Estimated number of teachers in population	Sampling rate for teachers (2nd stage) (1,000 ÷ Col. 5)	Expected number of teachers in sample	Probability of selecting a particular teacher
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	100,000 & over	22	22	1:1	207,000	1:1,000	207	.001
2	50,000-99,999	47	27	1:1.7	134,000	1:588	131	.001
3	25,000-49,999	118	34	1:3.5	170,000	1:286	171	.001
4	12,000-24,999	347	42	1:8.3	254,000	1:120	256	.001
5	6,000-11,999	926	51	1:18.2	358,000	1:54.9	359	.001
6	3,000- 5,999	1,856	45	1:41.2	364,000	1:24.3	363	.001
7	1,200- 2,999	3,475	42	1:82.7	336,000	1:12.1	336	.001
8	300- 1,199	4,836	34	1:142.2	169,000	1:7.0	170	.001
9	1- 299	4,004	32[a]	1:125.1	32,000	1:8.0[a]	32[a]	.001
Total		15,631	339	–	2,024,000	–	2,025	.001

Source:

NEA Research: Data used in 1980 Nationwide Teacher Opinion Poll, February 1980.

[a] The teachers in the many small school systems selected in the first stage from the stratum were combined as if in a single system and the appropriate number of teachers was selected by use of random numbers.

- Design: "Self-weighting stratified disproportionate two stage cluster sampling, with unequal sampling rate at the second stage."
- Target Population: Public School Teachers in the U.S.
- Sampling Population: First Stage: List of School Systems
  Second Stage: Lists of teachers within systems
- Sampling Frame: Same as sampling population
- Sampling Unit: First stage: School system
  Second stage: Teacher
- Stratification variable: enrollment size
- No. of strata = 9
- Disproportionate selection of school systems (first stage)
- Unequal sampling fraction of teachers (second stage)
- Equal probability of selection of teachers
- Parameters: N = 2,024,000
  n = 2,025
  P_T = 1/1000 probability of selecting a teacher
  $P_{s(i)}$ 1,...,9 = variable sampling rate for the systems

The survey being reviewed is in progress; thus the figures in this paper were from a similar survey five years ago. In the current survey, school systems directories were used for the systematic sampling of teachers.

As discussed by McCall (1980) the self-weighting two-stage samples have merits in that the actual sample data yield unbiased (or slightly biased) and consistent estimates of population parameters without weighting the strata or the clusters.

A word of caution for a very well defined sampling design should be said in reference to the final stage of selecting teachers at random. If this stage is haphazard, as when no control is exercised on the procedure, then, like, a deck of cards, the whole design, could crumble, and the estimates made from the sample data not meaningful. Upon further investigation of this process, care was in fact exercised, so that the teachers selected were truly random. This design should rate as excellent in both theory and practice.

Survey C

In order to meet the objectives of the survey, a sample of public elementary and secondary schools was drawn and the school questionnaire administered to the principals in the sample schools. Within the sample schools, a sample of teachers was drawn, and a different survey questionnaire was administered.

The sampling frame for the selection of the sample schools was the "Common Core Data" (CCD) 1983-84 universe of public elementary and secondary schools. Information on this frame allowed for stratification by geography, by size and by level of the school.

The schools were stratified by level:

Stratum 1	Grades 10, 11, 12
Stratum 2	Grades 9 and below
Stratum 3	Ungraded, vocational, special education and alternative schools

Within level the schools were further stratified by region and further stratified by the size of the school measured by its Fall membership. The report did not give the number of regions or the number of size strata. Let R be the number of regions and S the number of size strata. Then the total number of strata was 3 x R x S. Suppose there were four regions and three size strata, then the total number of strata is 36.

The probabilities of selection of the schools were equal to the square root of the number of teachers at the school.

With the school as the unit of observation, the design was a one-stage stratified sampling with pps (probability proportional to size) where the MOS (measure of size) was the sqaure root of the number of teachers in the school. As a survey of public school teachers in the U.S.--the subject of this paper--the design may be described as a stratified two-stage sampling of teachers. The sampling frame at the first stage consisted of the list of schools within the strata, and the sampling unit was the school.

At the second stage, a list of all teachers who taught at the sample school was the sampling frame. This list was obtained from the state education agency, and if not available, from the local education agency.

These sample teachers were further stratified into three strata by field of assignment: (1) general or elementary education, (2) mathematics and sciences, and (3) all other fields. The sampling fraction was computed for each of the three teacher strata.

The scheme was somewhat complicated, hence the succeeding procedure is reported here verbatim from the description of the design, with permission of the author. The reference used is a design drawn up before conducting the survey, hence, the future tense.

Table 5 gives an illustrative determination of the sampling fraction.

Table 5. Determination of the Sampling Fraction (Survey C)

Teacher Strata	Number of Teachers from the Sample	Desired Sample of Teachers	Sampling Fraction	Weight
Elementary	40,000	4,000	0.10	5
Math & Science	10,000	2,000	0.20	10
All Other	50,000	4,000	0.08	4
Total	100,000	10,000	0.10	

After the sampling fractions have been established a probability selection of teachers will be made. An equal number of teachers will be selected from each school where possible. If the school has fewer teachers than the required sample size, then all the teachers in that school will be taken in the sample. Otherwise, a probability sample will be drawn following the procedure outlined below. Suppose a sample of four teachers is desired in each school. To obtain the overall sample size in each teacher stratum the sampling fractions will be applied thru the use of a weight variable to assign differential probabilities of selection. The above illustration has such a weight variable included. If school A has 15 teachers, of which eight are elementary teachers, two are math/science teachers and the remaining five are stratified in the other category, then the sampling would proceed in the following fashion.

Weight up the counts in each stratum by the appropriate weight variable (using the weights from the illustrative example above), producing counts 8X5=40 (the count of teachers in the stratum multiplied by the stratum weight), 2X10=20, and 5X4=20. The total weighted count is 80 (40+20+20=80). The sampling interval is the weighted count divided by the sample size (80 divided by 4=20). A random start between 0 and the sampling interval is chosen to obtain the first selection number;, the sampling interval is added to the random start to find the second selection number, etc.

The selection numbers are then compared to the cumulative weighted counts to find the number of teachers to be selected in each stratum. The application of the procedure to determine the number of teachers to be selected from each stratum in a school is illustrated in the example school described earlier. See Table 6.

TABLE 6. School A (Survey C)

Teacher Stratum	Listed Teachers.	Weight	Weighted Counts	Cum Weighted counts	Sample Size
Elem.	8	5	40	40	2
Math & Science	2	10	20	60	1
All Other	5	4	20	80	1
Total	15		80		4

The design proposed two alternative selections of teachers. Systematic sampling is demonstrated as follows, referring to Table 6.

The design Interval = 80 divided 4 = 20
Random start = 5
First selection = 5
Second selection number = 25
Third selection number = 45
Fourth selection number = 65

The verbatim description of the design ends here.

By this systematic sample scheme it was clear that the numbers were applied to the weighted counts column, and to a list of teachers sorted by subject matter as shown in the table. Each elementary teacher would then be given a count of 5, (the weight), math and science, a count of 10, and all other, a count of 4.

Thus, the table should look like Table 7.

Table 7. Systematic Selection of Weighted List (Survey C)

Teacher Stratum	Weight	Serial Numbers of the 15 Teachers	Weighted Numbering	Systematic selection
Elem.	5	1	1-5	#5
		2	6-10	
		3	11-15	
		4	16-20	
		5	21-25	#25
		6	26-30	
		7	31-35	
		8	36-40	
Math & Science	10	9	41-50	#45
		10	51-60	
All other	4	11	61-64	
		12	65-68	#65
		13	69-72	
		14	73-76	
		15	77-80	

With simple random sampling, two elementary teachers of the eight listed, one math and science teacher of the two listed and one of the five listed other teachers were to be randomly selected.

This procedure gave the appropriate allocation for each school. As the procedure was applied to the sampled schools, the overall desired sample size was achieved.

The estimation procedure took into account the weight produced by this sampling design. It was beyond the scope of this paper to describe this procedure in this particular survey.

One important feature of this design is that the objectives of the survey were taken into consideration in determining the sample size, which in turn was going to affect the precision of estimate. To determine the precision of estimates ahead of time, this survey obtained the approximate numbers of teachers by stratum, in the universe, from the "1979-80 Survey of Teacher Demand and Shortage." Table 8 shows the approximate universe and sample sizes.

Table 8. Universe and Sample Size of Teachers (Survey C)

Stratum	Universe Size	Sample Size	Sampling Fraction
1. Elem.	90,000	4,000	0.004
2. Math & Science	245,000	2,000	0.008
3. Other	1,055,000	4,000	0.004
Total	2,200,000	10,000	0.005

The coefficient of variation for a proportion, p was used as a measure of precision:

(7) $CV(p) = (D \cdot q / np)$
np

where CV(p) is the coefficient of variation for a proportion, p

D is the design effect,
n is the effective sample size.

According to Survey C, the CV's based on a design effect of 1.5 and a response rate of 95% for the teachers in respondent schools, for various proportions of teachers with a characteristic are given in Table 9.

Table 9. Approximate CV's for Survey C

Proportion of teachers with a characteristic	Estimated approximate CV for proportion
0.75	0.008
0.50	0.013
0.35	0.018
0.20	0.027
0.10	0.040
0.05	0.058
0.02	0.093

Tables 8 and 9 were lifted from Survey C's paper with the author's permission.

Like Survey D, this design should rate excellent in sampling theory.

Membership Surveys -- I and J

Surveys I and J were conducted by two teacher organizations, directed only to their respective members. Surveys whose target population is not the entire public school teachers, as a whole, or surveys on selected school characteristics such as level, or subject matter are beyond the scope of this paper. However, these two reports were reviewed to demonstrate the complexity c design and difficulty of obtaining a random sample of teachers in the case of sampling all teachers where no complete list was available. In membership surveys such as I and J, design does not present the problems discussed earlier.

For Survey I, the sampling error was predetermined to compute the sample size. The selection was strictly random with equal probability of selection of members. Of course, like in Surveys A to H, the question of nonresponse presents a problem in representativeness. Several follow ups were made to maximize returns--thus reaching a 75% response.

In survey J, the sample size was 800. No mention was made of representativeness, and of sampling errors. The sampling frame was the membership file. Teachers were selected by systematic sampling.

It is clear that compared to the survey of the total public school teachers membership surveys could more easily minimize nonresponse. They have very well defined sampling frames--namely the membership list or file.

Hence, a good sampling design can easily be implemented. A sample random sampling design (one-stage) was found sufficient for purposes of reaching the teachers to be sampled.

Matrix Summary of the Sampling Procedures of the Ten Surveys

Table 10 is a summary of the features of the ten surveys for qualitative comparison. An attempt was made to secure independent "ratings" of the eight surveys of public school teachers. It was not feasible to ask the intended raters to read eight reports for purposes of ratings. Secondly, the ratings would not have much validity since the description of the sample came in different formats--articles, unpublished reports, unfinished survey reports, telephone conversation, personal interview, etc.

Thus, this tabulated summary would serve as a summary to give the reader a birds-eye view of the current surveys going on. Table 11 is the rating system to be used but is now suggested as a possible evaluation sheet that the reader of this paper may want to use to appraise the surveys reviewed or to use as a form to plan and evaluate their own survey/sample design.

Table 10. Sampling Procedures of Ten National Surveys of Public School Teachers in 1980 to 1985

SURVEY	Sample Size[1]	Representativeness	Precision/Sampling Error: Value	Precision/Sampling Error: How to determine	Source of Sampling Frame	Sampling Design	Questionnaire
A	(1994) 813	° 41% responded. ° NR[2] bias resolved	± 3.4% at 95% CL (not reported)	$e^2 = \frac{Z^2(p(1-p))}{n}$ N large	Market Data Retrieval, West port, CN	"The sample was stratified prop.by region & teaching level"	Mailed Q.
B	(4822) 1,846 N=1.2 mil	None cited; but has 32% male; approx., same prop., as pop.	± 2.2% at 95% CL (not reported)	-ditto-	MDR	Sample proportional to teaching level, for each state; cited similarity of data w/ reputable survey group	Personal interview
C	2,800 sch 10,000 teachers	Report not completed but sample was designed to ensure representativeness	sch= .014-.172 Teachers= .008-.093	CV[3] = $(D \times \frac{q}{np})^{1/2}$ (See text)	(name) = universe of public elem. & secondary schools	° Multi stage stratified sample: Level-Region-School-Teacher. ° Probability selection, weighted data	
D	2025 (1877) 1198 1326 N=2.2 mil	Design ensures representativeness but no discussion of nonresponse	Varies by the values of the mean and s_x	At 90% CL 1.645 x S_*	"From a comprehensive file of these systems"	Two-stage Prob.Sampling °Public School System stratified by size °Systematic sample 1 in 1000 = combined prob. of selection of teacher.	Mailed Q
E	1 in 700 out of 2.1 mill. (3000) 1346	Demo data compared w/known sources of similar data; finding=rep.	± 1.4% to ± 8.3%	Used table w/ Confidence limits for prop. from 10%-50%.	Not given	Apparently simple random from a national list	Mailed Q 2 follow-ups
F	No description of sampling procedure and related info. Simply reported percent results.						Mailed Q.

BEST COPY AVAILABLE

Table 10. continued

SURVEY	Sample 1 Size	Representativeness	Precision/Sampling Error Value	Precision/Sampling Error How to determine	Source of Sampling Frame	Sampling Design	Questionnaire
G	252 "From all geog. regions of U.S."	Statement on representativeness but unacceptable.	None	None	None	100 Elem.Princ"randomly identified ..."from a cross section of their membership	Mailed Q No follow-up
H	100 private 400 public	None	Report not finished	Report not finished	MDR	Two-stage °"Rand.Samp.of Private & public schools °Have set of rules for princ.to draw 4 teachers at random	Mailed Q
		SELECTED TEACHERS BELONGING TO AN ORGANIZATION -					
I	1783	Representative by Sample Design	± 4%	e^2	Membership data base	Simple Random from comprehensive list. Equal prob. selection.	Mailed Q w/follow up
J	(800)	None mentioned	None	None	Membership files	Systematic Sampling	Mailed Q

1/ n is the effective sample size; the number in parenthesis is the original sample size.

2/ NR = nonresponse 3/Coefficient of Variation

Table 11. Please rate each study according to some aspects of sampling by encircling 1, 2, or 3. 1/ When rating, bear in mind the relationship between a good/bad sample and the reliability of results.

SURVEY	Sample Size	Representativeness	Precision/Sampling Error: Amount	Precision/Sampling Error: How Determined	Source of Sampling Frame	Sampling Design
A	1 2 3	1 2 3	1 2 3	1 2 3	1 2 3	1 2 3
B	1 2 3	1 2 3	1 2 3	1 2 3	1 2 3	1 2 3
C	1 2 3	1 2 3	1 2 3	1 2 3	1 2 3	1 2 3
D	1 2 3	1 2 3	1 2 3	1 2 3	1 2 3	1 2 3
E	1 2 3	1 2 3	1 2 3	1 2 3	1 2 3	1 2 3
F	1 2 3	1 2 3	1 2 3	1 2 3	1 2 3	1 2 3
G	1 2 3	1 2 3	1 2 3	1 2 3	1 2 3	1 2 3
H	1 2 3	1 2 3	1 2 3	1 2 3	1 2 3	1 2 3

1/ Rating:

1 = Not Acceptable

2 = Acceptable but results should be evaluated and/or reliability examined.

3 = Highly Acceptable and theoretically sound.

PROBLEMS

One of the objectives of this paper was to identify problems associated with the sampling practices prevailing in national surveys of public school teachers, and to offer solutions to these problems. As gleaned from these surveys the problems were:

1) lack of sampling frame or list of all public school teachers from which sample may be drawn;

2) the assurance that the sample drawn, (or, in effect, the respondents in incomplete returns of a survey) is representative of the target population of public school teachers;

3) the theoretical connection between a sample design and the computation of estimates from the results;

4) the recognition and knowledge of the quantitative relationship between the sample design and the precision of estimates; and

5) reporting results that demonstrate validity, and willingness and ability to sort out the strengths and weaknesses of the survey.

These problems are all interrelated, hence the discussion and proposed solutions do not correspond one-to-one to the enumerated problems in this section.

Discussion and Proposed Solutions

Determining Representativeness of the MDR list

As mentioned earlier the most comprehensive list of teachers is that compiled by Market Data Retrieval of Westport Connecticut. The list which contains only 68% of the 2.4 million teachers indicates a large discrepanc between the target population and sampling frame. Still many surveys relied on this list.

One survey tested representativeness of incomplete returns (though no in connection with the MDR list) by comparing demographic, socioeconomic and attitudinal variables of the respondents with those of the nonrespondents who were interviewed by telephone, using the same questionnaire. Similarly, if the MDR list will continue to be used, a study to test its representativeness ought to be conducted. One way would be comparing a sample drawn from this frame with a sample drawn directly from district or school lists according to a comprehensive multistage design as used by Surveys C and D, assuming that the within stratum selection is truly random. Otherwise, the problems of lack of sampling frame, and representativeness are best resolved by designs that eventually achieve a sampling of teachers without necessarily drawing them from a national list.

Representativeness may also suffer from nonresponse in mail surveys, since the nonrespondents may be different from respondents in the variable(s) being studied. Edgar (1986) conducted a review of studies correcting for nonresponse bias. In her current work, she is developing a model to estimate from partial returns, values of specific variables for a complete response.

Sampling Design to Resolve the Problems of Lack of Sampling Frame and of Representativeness

In surveys C and D, the multistage stratified sampling design resolved these problems on sampling frame and representativeness, as well as problems of costs.

In stratification by student enrollment in Survey D, there was an implicit assumption that the ratio of students to teachers would be uniform across the different enrollment sizes. Therefore, in an earlier similar survey (not reviewed in this paper because of identical design), enrollment, for which data were readily available was used to determine the numbers of teachers in school systems--for which data were not available. A regression equation estimating the number of teachers from enrollment size was used according to the information source for Survey D.

In the current study, enrollment size was again used for stratification but the actual number of teachers within school systems which was available after selecting school systems for the sample, was used in determining the probability of selection.

Stratification also saves time and costs in selecting the sample. In the stratum of large school systems, all systems were included in the sample (Survey D). In the strata of smaller systems, if all systems were included, at the most, only one teacher from a school system would be selected. This would require correspondence with a large number of school systems. In this design fewer systems were selected; and with the probability of selection of system entering the formula, more than one teacher were sampled from the selected systems, to maintain the sampling fraction of 1/1000.

This is a design that solved practical problems and minimized costs, without sacrificing randomness and reliability. Precision may probably suffer to a slight degree depending on the homogeneity of the sampling units within strata.

Discrepancy Between Actual and Excepted Sample Size in Stratified Sampling--Practical Problem

Before going into estimation, the first thing to do after the samples in the L strata have been selected is to examine the closeness of the actual number of teachers drawn from the strata to the expected number as determined by the design. These expected sample sizes are shown in column (8) in Table 3 which illustrates the stratified, disproportionate, two-stage, cluster sampling. In the actual implementation of the design the systems that had been sampled in Survey D were contacted and given uniform instructions for systematic sampling, using the sampling rate for the teachers at the second stage (Column (7), Table 3). The gap between the time when statistical data (used in the sample design) were collected and the time when school systems were contacted for selecting teachers would of course produce a discrepancy between the sample size determined at the planning stage (estimated sample size) and the actual number sampled. The actual procedure was to give the precalculated sampling interval to the person in the school system. Thus, it would not be realistic to expect exactly the same number as estimated. However, the respective stratum sample sizes should not be too different from the expected. Table 12 shows the estimated sample sizes by State and by Stratum. At the bottom of the table are actual sample sizes by stratum, totalling 1998. Such a table as this, which shows the distribution of the nine strata among the states would bring the reader to the reality of sampling which is presented in a different light from the theory from which its process emerged.

Table 13 shows the actual total number of teachers N in each stratum, the actual stratum sample size n and the prior estimate (n). This table illustrates the deviation of the implementation from the plan. The discrepancy and its percentage as a fraction of the estimate were also presented.

Table 13. Actual and Estimated Sample Sizes in the Nine Strata of Survey D

Stratum	Actual No. of Tchrs in the n_h	Total No. of Tchrs N_h	Estimated No. of Tchrs in the Sample n_h'	Discrepancy d $d = n_h - n_h$	Per Cent Discrepancy $\frac{\lvert d \rvert}{n_h'}$
1	192	175,028	175	+ 17	.097
2	182	159,788	160	+ 22	.138
3	179	198,272	198	- 19	.096
4	266	276,713	277	- 11	.040
5	350	372,147	372	- 22	.059
6	323	348,258	348	- 25	.072
7	232	345,413	345	- 13	.038
8	198	189,558	190	+ 08	.042
9	76	58,873	59	+ 17	.288
U.S.: ALL Strata	1,998	2,124,050	2,124	-26	.012

Table 12. Distribution of Teachers Sampled by Survey D Across State and Strata, 1985

		STRATA 1	2	3	4	5	6	7	8	9	ALL
S	AK	0	0	0	0	9	0	0	0	0	9
T	AL	0	9	5	9	7	0	10	0	0	40
A	AR	0	0	0	0	6	0	8	6	7	27
T	AZ	0	0	5	6	0	8	0	0	0	19
E	CA	35	12	30	35	33	10	8	7	2	172
	CO	0	10	0	6	0	0	0	5	1	22
	CT	0	0	0	8	19	8	0	0	1	36
	DC	4	0	0	0	0	0	0	0	0	4
	DE	0	0	0	4	0	0	0	0	0	4
	FL	37	22	14	6	7	8	0	0	0	94
	GA	0	25	11	0	9	0	8	0	0	53
	HI	9	0	0	0	0	0	0	0	0	9
	IA	0	0	0	6	6	0	0	10	0	22
	IL	22	1	5	6	7	16	16	22	4	99
	IN	0	5	0	7	17	8	27	0	0	64
	KS	0	0	3	8	0	8	0	9	2	30
	KY	0	7	0	0	7	9	16	0	0	39
	LA	0	0	5	0	6	0	0	0	0	11
	MA	0	4	0	12	6	6	14	5	1	48
	MD	14	14	0	6	0	0	0	0	0	34
	ME	0	0	0	0	0	0	0	5	1	6
	MI	11	0	10	5	13	9	17	10	0	75
	Mn	0	0	6	7	0	6	2	5	1	27
	MO	0	4	1	13	7	7	9	[illegible]	1	47
	MS	0	1	0	0	6	8	0	0	0	15
	MT	0	0	0	0	0	0	9	21	32	62
	NC	0	10	10	0	7	11	0	0	0	38
	ND	0	0	0	0	0	0	0	0	2	2
	NE	0	0	5	6	9	0	0	11	6	37
	NH	0	0	0	0	0	0	0	0	1	1
	NJ	0	0	0	0	8	19	0	10	1	38
	NM	0	6	0	0	0	10	0	0	0	16
	NV	0	5	0	0	0	10	0	1	0	16
	NY	22	0	5	0	27	10	8	5	1	78
	OH	0	10	10	0	14	34	37	6	0	111
	OK	0	0	0	6	0	0	8	17	5	36
	OR	0	0	0	7	8	10	0	0	0	25
	PA	0	0	0	1	27	32	[illegible]	1	0	77
	SC	0	8	5	12	20	9	0	0	0	54
	SD	0	0	0	6	0	0	0	0	1	7
	TN	9	5	5	12	0	5	0	6	0	42
	TX	21	11	19	33	23	37	0	5	2	151
	UT	0	5	0	6	7	1	0	0	0	19
	VA	8	0	20	6	14	6	0	5	0	59
	VT	0	0	0	0	0	0	0	9	2	11
	WA	0	0	0	14	0	8	0	0	1	23
	WI	0	8	0	6	7	10	16	5	1	53
	WV	0	0	5	7	14	0	3	0	0	29
	WY	0	0	0	0	0	0	0	7	0	7
	ALL	192	182	179	266	350	323	232	198	76	1998

The total sample size of 1998 was just one percent less than the estimated 2,124. By strata, the discrepancies in proportion to the estimated sample size, n_h' range from 4% (strata 4,7,8), to 29% oversampling in stratum nine. Stratum nine by design was the group of school systems with the smallest enrollment--1 to 299. It is interesting to note (see discrepancy column, d) that the smallest (strata 8 & 9) and the largest (strata 1 and 2) school systems were over sampled, while those in the middle (Strata 3 to 7) were undersampled. The authors of this paper were not attempting to speculate on the explanation but these are the types of information that may be used by the sampling practitioner to improve on the next design.

Sample Design and Estimation--Theoretical Problem

In most of the survey reports reviewed, the sample design was relegated to the background. Apparently, the important thing for most researchers was the assurance of an adequate, representative and reliable sample, or at the very most, an explanation of flaws in this area. Results were reported, without consideration of the method of sampling. For instance, in using clusters or strata are the variances within strata homogenous? If not, how would the separate estimation of variances and mean, by stratum, be accomplished, and how should they be summarized? Books on sampling present a wide range of methods of sample allocation--such as equal, proportional, optimum (when cost enters the formula/or Neyman (when the stratum variance S_h^2 or proportion P_h is expected to vary from stratum to stratum (Parel, 1978).

In order to demonstrate the need for this connection between sample design and estimation, let us consider Survey D. McCall (1980, p.274) in discussing sample size determination for stratified two-stage sampling, pointed to the problem of lack of adequate preliminary information regarding the nature of the population to be surveyed and the expected variances to be encountered. In estimating the sample size, one has usually no recourse but to make certain assumptions. In the course of completing the design, or after a pilot survey, additional information may come up and thereby one could make a more realistic determination of sample size. After data have been collected, all prior estimates or assumptions should be discarded and estimation made, in accordance with the design.

To illustrate one procedure of mean and variance estimation for each stratum after data from the survey have been collected, let us use the relevant columns of Table 14. Table 14 shows how to compute the estimates of the stratum mean and the stratum variances.

The illustrative data are as realistic as possible. Thus, the mean salary for each stratum was estimated from the states' average salaries of instructional staff, 1983-84 (NEA, 1984). Thus, the data across strata appeared more homogenous than they probably would be, had the salary data for each individual teacher in the sample been available.

Table 14. Computation of the Sample Mean and Variance of the Nine Strata

Stratum	Total No. Sampling units in in the Stratum[1]	Sample taken from the Stratum[1]	Estimate of Stratum Salary[2]	Est. of Stratum variance	Est. of Stratum Stan. dev.	Variance of the mean
	N_h	n_h	$\bar{x}_h$	s_h^2	s_h	$\frac{s_h^2}{n_h}$
1	175,028	192	$25,030	18,443,896	4,295	96,062
2	159,788	182	21,991	9,421,297	3,069	51,765
3	198,272	179	23,285	19,331,124	4,397	107,995
4	276,713	266	22,881	15,187 728	3,897	57,097
5	372,147	350	23,341	14,523,[illegible]4	3,811	4,495
6	348,258	323	22,752	8,671,700	2,945	26,84
7	345,413	232	23,037	11,242,630	3,353	48,460
8	189,558	198	21,949	15,938,899	3,992	80,499
9	58,873	76	21,144	8,028,796	2,834	105,642
Total	N=2,124,050	n=1,998				

[1] Data from Survey D

[2] Estimated from state averages. If this were an actual computation, $\bar{x}_h$ would be the mean salary of the sample teachers in stratum h, considered as estimate of the mean salary of all teachers in that stratum.

The estimate $\bar{x}_{st}$ of the population mean, $\bar{x}$ is

(8) $$\bar{x}_{st} = \frac{\sum_{h=1}^{L} N_h \bar{x}_h}{N}$$ where $\bar{x}_h$ is the estimate of the mean of the hth stratum and L is the number of strata.

Computing from the table,

$\bar{x}_{st}$= $23,063.
This would be the estimated average salary of teachers in the whole U.S. (illustration only, not fact).

The estimate of the variance of the sample mean is

(9) $$v(\bar{x}_{st}) = \sum_{h=1}^{L} \left[\frac{N_h^2}{N^2} \cdot \frac{(N_h - n_h)}{N_h} \cdot \frac{s_h^2}{n_h} \right]$$

or

(10) $$v(\bar{x}_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} \left[\frac{N_h (N_h - n_h) s_h^2}{n_h} \right]$$

where,

s_h^2 is the sample variance in the hth stratum

$\frac{N_h - n_h}{N_h}$ is the finite population correction (f.p.c.) for the hth stratum.

This f.p.c. may be ignored if it is greater than or equal to 0.95, or if n_h/N_h is less than or equal to 0.05 (Parel, 1978).

Since in national teacher surveys, we are sampling from a very large population, n_h/N_h would be very small--in this example--approximately 1/1,000. We can then ignore

(11) $(N_h - n_h)/N_h$ in equation (9). This reduces equation (9) to

(12) $$v(\bar{x}_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} \left(N_h^2 \cdot \frac{s_h^2}{n_h} \right).$$

Using the values in Table 14, we obtain

$$v(\bar{x}_{st}) = 4,870.$$

What is shown here is the connection between design and estimation. When estimating the population

statistic such as the mean, N or the total number of sampling units in the stratum enters the computation as the stratum weight. Likewise, the variance of the mean would be the statistic in computing the reliability (or sampling error) of the estimate of the mean.

Thus if the strata in an actual survey are not very homegeneous as measured by $V(\bar{x}_{st})$--a characteristic normally expected or desired in stratified sampling--and this procedure is ignored, by using the unweighted mean of all sampled units in the strata, then this estimate would not be accurate.

For want of better data to illustrate the connection between design and estimation, we have used the data that were collected for a design involving cluster sampling.

It is at this point that we consider it more helpful to distinguish between stratified and cluster sampling.

If the population consists of groups with great variation in their means, it is usually recommended to divide or stratify the population into more or less homogeneous subpopulation or strata. This design improves efficiency, that is, minimizes the variance. The characteristics of a stratified population usually, is that the variance <u>within</u> strata is small, while the variance <u>between</u> strata is large. One can see, that with this rationale in mind, it becomes necessary to make separate stratum estimates of the parameters as we have shown, before estimating the population parameters.

Similar to the strata in stratified sampling, clusters are mutually exclusive subpopulations which together comprise the entire population. Unlike strata, clusters are preferably formed with heterogeneous rather than homogeneous elements, so that each cluster will be typical of the population (Parel 1978).

The advantages and disadvantages of stratified random sampling and of cluster sampling, as discussed below will be more appreciated in the light of the examples given.

<u>Advantages and Disadvantages of Stratified Random Sampling (Parel, 1978)</u>

The Advantages of stratified random sampling are:

(1) it is more efficient than random sampling if the population had been so stratified that the stratum variance S_h^2 are small relative to the overall population variance S^2;

(2) it allows for more comprehensive data analysis since information is provided for each stratum or subpopulation; and

(3) it is administratively convenient

The disadvantages of stratified random sampling are:

(1) the stratification of the population may mean the need for additional prior information about the population and its subpopulations; and

(2) a separate frame is needed for each stratum.

Advantages and Disadvantages of Cluster Sampling (Parel, 1978)

One advantage of cluster sampling is that there is no need to construct a list of elements in the population as one must in random or stratified sampling; the frame for cluster sampling is simply a list of the clusters.

One disadvantage of cluster sampling is that it is not so efficient as random or stratified sampling. Thus, adjacent schools and teachers in these schools may have more similar characteristics than those distantly apart. For cluster sampling to give precise estimates the elements within clusters should be heterogeneous with respect to the characteristic being measured, so that the sample will be more representative of the population.

Going back to Survey D, without knowledge at this point, of the characteristics of the subpopulations, it was not easy to determine whether the intent was to stratify or to form clusters; or whether the design was motivated by the feasibility of sampling or administrative convenience. Nevertheless, the very well defined procedure and the availability of comprehensive lists of school system and teachers within systems, as well as the information on the numbers school systems within strata, making estimates using formulas for stratified sampling would not hurt.

The best interpretation we can give of Survey D is that its procedure is that of stratified sampling where the stratification variable is enrollment size. Then it became apparent that if one considers the variable salary, the clusters of teachers may be homogeneous with respect to salary; and if so, the nine strata may be considered as clusters.

In planning on complex designs, one should have prior assurance that the final unit of analysis, the teachers could be selected randomly to continue to ensure representativeness. Much of the operational problems lie in the final sampling of teachers. For instance, the school or school system may not furnish a list, and/or the selection is left to another school personnel with vague or without instruction on random selection. There were cases from the surveys reviewed, where no instruction was given.

When one fails at this tail-end of the process, then the sample design and its implementation may be compared to a castle built from a deck of cards where in the last card at the top topples the whole deck.

CONCLUSION, IMPLICATIONS, AND EDUCATIONAL IMPORTANCE

The concern over the use of approriate statistical procedures is not limited to education-related surveys. Neter (1986) in his presidential address at the 1985 Annual Meeting of the American Statistical Association, discussed concerns over the uses of statistical sampling techniques in a wide array of disciplines; the theme in that meeting was Statistics and Public Policy.

It may be small comfort to know, that national surveys of teachers are not the only projects experiencing problems in sample surveys. Johnson and Smith (1969), in describing the tremedous impact of sampling procedures throughout the modern world pointed to the need of solving problems of nonsampling errors and large sampling errors among others. Their book, a compilation of papers by experts in the field provided a broader view of the theoritical aspects of survey sampling.

It seems that surveys of the general public like that conducted by The Gallup Organization (Gallup, 1985)--"A Gallup Survey on Education"--and by the Census Bureau (1980) in its 1980 census sample, had rigorous sample designs, practically beyond question in theory and practice and in reporting. Gallup's design of the sample--"a replicated probability sample down to the block level in the case of urban areas, and to segments of townships in the case of rural areas"--took serious consideration of reliability and implemented preventive measures that ensured representativeness, as well as technical measures that would reduce the sample bias. The Census Bureau (1980) on the other hand, took into account both sampling and nonsampling errors. Such concepts as sampling unit, confidence intervals, standard errors, were deeply entrenched into the operational aspects of the design; estimation procedures and control of nonsampling error were clearly connected with the design.

Thus, if a certain standard of credibility is being maintained for these public surveys, why can't the same standards be applied to national surveys of public school teachers? With educational reforms on the top agenda of policy makers on the one hand, programs, laws and regulations and other educational decisions on the other hand, are rising in equal fast tempo. Serious problems arise when persons in charge make decisions, formulate policies and develop programs based on faulty data and bad samples.

The crucial role of surveys in shaping policies and decisions on educational matters cannot be overemphasized. A survey that misleads the public on the opinions of teachers, on account of nonrepresentative and inadequate samples can do a great deal of harm in making decisions about education.

On the other hand, a survey that truly represents the opinions of our educators would have contributed toward the total input into the process of determining what is good for the students, the teachers and the school, all inseparable in the shaping of the future in education.

It is expected that this presentation would draw more attention to the importance of appropriate sampling procedures in the reliability of estimates and representativeness of the sample. For so[illegible] survey organizations, the solutions offered could be of help in improving their traditional sampling approach.

REFERENCES

Bureau of Census, Census of Population and Housing, Washington,D.C., U.S. Department of Commerce, February 1983.

Deming, William Edwards, Some Theory of Sampling. Dover Publications, Inc., New York, N.Y., 1966.

Edgar, Suzanne H., Nonresponse Bias in Demographic and Attitudinal Responses from Mail Surveys, University of Maryland, 1986 (Study in progress).

Fleiss, Joseph L., Statistical Methods for Rates and Proportions, John Wiley and Sons, New York, 1973.

The Gallup Organization, Inc., A Gallup Survey on Education, The Gallup Organization, Inc., 53 Bank Street, Princeton, New Jersey 08542, June, 1985.

Grant, W. Vance and Thomas D. Snyder, Digest of Education Statistics 1983-84, National Center for Education Statistics, 1984.

Jessen, Raymond J., Statistical Survey Techniques. John Wiley and Sons, New York, 1978.

McCall, Chester H., Sampling and Statistics Handbook for Research in Education, National Education Association, Washington, D.C. 1980.

National Education Association (NEA), Data Search: Estimates of School Statistics, 1983-84, As Provided by the State Departments of Education, NEA, 1984.

Neter, John, Boundaries of Statistics--Sharp or Fuzzy, Journal of the American Statistical Association: 81, 1-8, March, 1986.

Parel, Cristina P., et all, Social Survey Research Design, The Statistical Center, University of the Philippines Systems, Quezon City, Philippines, 1978.

Snedecor, George W. and William G. Cochran, Statistical Methods, The Iowa State University Press, Ames, Iowa, 1967.

Sudman, Seymour, Applied Sampling, Academic Press, New York, 1976.

Pena/Henderson
AERA, 1986

How to Determine The Sample Size in Estimating a Proportion

- Confidence Level and associated z

(Frame1)

Confidence Level	z
90%	1.65
95%	1.96
99%	2.58

Note

Confidence Interval is the interval of estimate, measured by z multiplied by the standard deviation s. The confidence level may also be referred to as the reliability in percentage values.

- Sampling error = measure of precision; e.g. ± 3%, ± 5%, applied to a sample estimate of the population proportion.
- Example:

(Frame2)

Given: Desired Con ' nce Interval: 95% or .95
Desired precision: ±3% or ±.03

Problem: Determine sample size.

Solution: CI at 95%, is associated with z = 1.96

	given		unknown		given
Part A	1.96	x	s	=	.03
	z	x	stand.dev.	=	acceptable error

(Note: For a quick mental computation, use 2 instead of 1.96)

Therefore:
(Frame 3)

$s_p = \frac{.03}{1.96}$ $\qquad s_p = \frac{\text{desired precision}}{z}$

$s_p =$ .0153 (put this aside)

Part B

The formula for the standard deviation of a proportion, s_p is

$$s_p = \sqrt{\frac{p\,(1-p)}{n}}$$; and for the variance, is

$$s_p^2 = \frac{p(1-p).}{n}$$

Let $p = .5$ (or 50%)
$1-p = .5$

Note: .5 yields the highest s_p in a range of values from .0 to 1.0; hence it is used in determining the sample size n.

$$S_p = \frac{.5 \times .5}{n} ; \qquad n = \frac{.25}{s_p}$$

Squaring both sides we have

(Frame 4)

$$n = \frac{.25}{s_p^2}$$

$s_p = .0153$ (Part A)

$$n = \frac{.25}{(.0153)^2}$$

$n = 1067.9$ or 1068, the required sample size to achieve the desired precision of ± 3%

Comment: (1) The computations above are applicable to sampling from a large population. The correction factor (N-n)/N is omitted in Part B, since for a large population N, and a relatively small n, the value is close to 1.

(2) Note that a quick computation of sample size can be achieved using only simple arithmetical calculations in frames 3 and 4.

(3) The total information needed in performing the calculations are in Frames 1 to 4. The text in between are explanatory and may be ignored.

ATTACHMENT 2

BIAS AND SAMPLING ERROR

Concepts and Illustrations

Using situations commonly encountered in NEA surveys these two concepts will be clarified by the illustrations of four possible combinations:

1. Sample is biased and has low precision (large sampling error)
2. Sample is unbiased but has low precision
3. Sample is biased but has high precision (small sampling error)
4. Sample is unbiased <u>and</u> has high precision

Illustrations:

1. Biased, low precision

   - You know that 40.0% of the membership you are surveying are men.

   - Your returns indicate only 17.4% are men.

   - Your results could be <u>biased</u> (toward the women), especially if the question or item tends to elicit responses from women, different from those of men.

   - Further, you have a very small sample so that your sampling error at the 95% confidence level is ± 8%; your estimates have a <u>low precision</u>

   - Thus your survey results are not only <u>biased</u> but also have a <u>low precision</u>

<u>Example</u>: 74% favored a given issue. With a sampling error of ± 8%, you are "95% confident that the proportion is between 66% and 82%."

Suppose further that the breakdown by sex is given in the table below, based on 1000 sample respondents.

Table 1 Males underrepresented

	Total		Female		Male	
Yes	737	(74%)	697	(82%)	37	(25%)
No	263	(26%)	153	(18%)	113	(75%)
Total	1,000	(100%)	850	(100%)	150	(100%)

We mentioned earlier that males were under represented. You will also note from the table that men tend to say no (75%) while women tend to say yes, on the issue. Therefore, the overall proportion saying "yes", 74% could be an over estimate, because males were under-represented.

2. <u>Unbiased, low precision</u>

Suppose this time that your sample proportion of men and women represent the membership proportion 40-60. The table would be:

Table 2: Correct Proportion of Male members

	Total		Female		Male	
Yes	592	(59%)	492	(82%)	100	(25%)
No	408	(41%)	108	(18%)	300	(75%)
Total	1,000	(100%)	600	100	400	(100%)

Note that there are 400 males, 600 females. Looking at the "Total" columns in Table 1 and Table 2, one can see that the "yes" proportion in Table 1, of 74% is an overestimate, compared to the yes column in table 2, of 59%.

Note further that even if the 95% C.L. of ± 8% was applied to table 1, the more accurate proportion saying "yes", 59% is still below the confidence interval of 66 to 82 in table 1.

3. <u>Biased, high precision</u>

Suppose we have a higher precision in Table 1, ± 3%; then our estimates (based on a biased sample) of those saying yes would be 74 + 3% or 71% to 77%. The precision has increased or improved, but note that because of the bias our smaller interval of estimate moves the unbiased 59% further away from the interval.

4. <u>Unbiased, high precision</u> (Ideal combination)

Suppose we have a higher precision for Table 2, ±3%; then our estimates based on an unbiased sample is 59 ± 3% or 56% to 62%. Thus, we can safely assume, with 95% confidence that the proportion of the total population answering "yes" to the question is between 56% and 62%.

Sampling Biases

<u>Sampling Biases</u> (mostly applied to NEA Surveys)

- Human failures

  Wanted active classroom teachers in 1985-86, who were members as of August 1985 but used membership records of January 1986 after nonrenewals have been purged. Your sampling frame would <u>not</u> include nonrenewals.

  - Intentional or nonintentional preferences - e.g. in evening calls, accessing only those members who answer their phones may result in a higher proportion of younger married members.

- Bias of estimating procedure
  - The estimating procedure used has built-in bias e.g. erroneous use of the table of random numbers, such that each member is not given an equal chance of being selected.
  - Use of the wrong formula of estimation
- Bias from failure to randomize starting points in systematic sampling
  Bias from inadequate definition of the sampling frame.

References

Deming, William Edwards; Some Theory of Sampling, Dover Publications, Inc., 180 Varick Street, New York, N.Y. 10014, 1966.

Jenning, Raymond J. Statistical Survey Techniques, John Wiley and Sons, New York, 1978.

Sudman, Seymour, Applied Sampling, Academic Press, Inc. 111 Fifth Avenue, New York, New York 10003, 1976.

Pena/Henderson
AERA 1986

BIAS AND SAMPLING ERROR

Concepts and Illustrations

Using situations commonly encountered in NEA surveys these two concepts will be clarified by the illustrations of four possible combinations:

1. Sample is biased and has low precision (large sampling error)
2. Sample is unbiased but has low precision
3. Sample is biased but has high precision (small sampling error)
4. Sample is unbiased <u>and</u> has high precision

Illustrations:

1. Biased, low precision

   - You know that 40.0% of the membership you are surveying are men.
   - Your returns indicate only 17.4% are men.
   - Your results could be <u>biased</u> (toward the women), especially if the question or item tends to elicit responses from women, different from those of men.
   - Further, you have a very small sample so that your sampling error at the 95% confidence level is ± 8%; your estimates have a <u>low precision</u>
   - Thus your survey results are not only <u>biased</u> but also have a <u>low precision</u>

<u>Example</u>: 74% favored a given issue. With a sampling error of ± 8%, you are "95% confident that the proportion is between 66% and 82%."

Suppose further that the breakdown by sex is given in the table below, based on 1000 sample respondents.

Table 1 Males underrepresented

	Total		Female		Male	
Yes	737	(74%)	697	(82%)	37	(25%)
No	263	(26%)	153	(18%)	113	(75%)
Total	1,000	(100%)	850	(100%)	150	(100%)

We mentioned earlier that males were under represented. You will also note from the table that men tend to say no (75%) while women tend to sav yes, on the issue. Therefore, the overall proportion saying "yes", 74% could be an over estimate, because males were under-represented.

2. Unbiased, low precision

Suppose this time that your sample proportion of men and women represent the membership proportion 40-60. The table would be:

Table 2: Correct Proportion of Male members

	Total		Female		Male	
Yes	592	(59%)	492	(82%)	100	(25%)
No	408	(41%)	108	(18%)	300	(75%)
Total	1,000	(100%)	600	100	400	(100%)

Note that there are 400 males, 600 females. Looking at the "Total" columns in Table 1 and Table 2, one can see that the "yes" proportion in Table 1, of 74% is an overestimate, compared to the yes column in table 2, of 59%.

Note further that even if the 95% C.L. of ± 8% was applied to table 1, the more accurate proportion saying "yes", 59% is still below the confidence interval of 66 to 82 in table 1.