

A chapter from

A Meeting Place

The History of The American Association for Public Opinion Research

Edited by
Paul B. Sheatsley
and
Warren J. Mitofsky

Copyright © 1992
American Association for Public Opinion Research
All Rights Reserved

ISBN: 0-89089-479-5

For PDFs of other chapters and the entire book see

<http://aapor.org/history>

5

Methodology Issues

Jack Elinson

Author's note: *This chapter presents selected highlights of survey research methodology as these highlights were displayed at AAPOR conferences. It is not intended as a scholarly historical overview of survey research methodology.*

AAPOR members have indeed treated the history of survey research in a comprehensive manner. See, for example, the 1987 volume by Jean Converse of the Survey Research Center at the University of Michigan (Survey Research in the United States: Roots and Emergence 1890–1960, University of California Press.) Before his death in 1985, Herbert Hyman also had been engaged in a historical analysis of the development of survey research from another perspective. Completion of this work, which had been supported by the Russell Sage Foundation.

Centuries before public opinion research had ever been thought of, there were surveys. For example, in 1427 a survey was conducted of 60,000 households covering 260,000 people in the metropolitan area of Florence, Italy. The data, which are registered in 1,000 volumes, were tabulated by social class. Social differences are indicated by the possession of a family name. Sans sample survey methods, and sans computers, the world has waited five-and-a-half centuries for an analysis of the results (Herlihy and Klapisch-Zuber, 1984).

Ambitious as the Florentine survey was, there is no evidence that its designers worried about the methodological issues that have haunted survey researchers in the twentieth century and have been debated at AAPOR conferences since the organization was founded. Three methodological themes were struck at the opening meeting in Central City in 1946: quota versus area probability sampling, the assessment of validity, and the quality of the data collected. The controversies in methodology that raged in the social sciences in the aftermath of the Second World War were to reverberate at the annual meetings of the Association, in the pages of the *Public Opinion Quarterly*, and in the various publications used by AAPOR members for reports of their research. These three themes have persisted as problematic over the forty years

since the birth of AAPOR, testifying to the wisdom of the organizers of the Central City meeting in selecting fundamental issues underlying survey research for their agenda.

Quota versus Area Probability Sampling

Until the advent of probability sampling applied to population surveys in the decade just prior to the organization of AAPOR, quota sampling was the pollster's choice in achieving representativeness. The sample selected was designed to be a miniature of the population to be studied and was thus designed to represent it in essential demographic distributions. Typically, these comprised sex, age, race, socioeconomic status, rural versus urban residence, and geographical region. Pollsters instructed interviewers to obtain respondents having certain characteristics, relying on Census data to create quotas.

During World War II, statisticians engaged in government sample surveys for the Bureau of the Census, the Bureau of Agricultural Economics, and the Research Branch of the U.S. Army's Information and Education Division adopted probability theory as a basis for population sampling. An advantage claimed for probability sampling was its capability of estimating sampling error. Because simple random sampling was not feasible for geographically dispersed populations, statisticians devised ingenious multistage area designs that were both practical and in conformity with probability theory. But such ingenuity had its costs: for example, special enumerations of dwelling units had to be made first, as a basis for the subsequent sampling of these units. Even more costly was the recognition of the social phenomenon of nonresponse along with efforts to contain it. Quota sampling was free of the problem of nonresponse: if a potential respondent refused to be interviewed or was otherwise unavailable, the interviewer simply found someone else with the required characteristics in order to fill the quota. Nonresponse has become a persistent concern of probability sampling.

The first confrontation in AAPOR history between quota samplers and area probability samplers took place at the 1946 Central City meeting. Representing quota sampling was the distinguished social psychologist Norman Meier, director of the Bureau of Audience Research at the University of Iowa. Meier was a consultant to principal pollsters of the time and a productive researcher. He was later to report on empirical tests in Iowa City of the relative performance of quota and probability sampling in the same survey. (Meier and Cletus, 1947).

Representing probability sampling was the noted statistician Morris Hansen, a technical adviser to the U.S. Bureau of the Census.

At this first confrontation, Meier expressed doubts that area probability sampling was superior to quota sampling for opinion research. He was supported by Elmo Wilson, who felt that "opinion polls still had reason to hold to quota methods, given the problem of costs, the need for speed, and the lower level of accuracy they required." (Converse, 1987:231) Hansen argued that probability selection methods provided a theoretical basis for estimating sampling error and the yield of information, in statistical terms, per dollar spent.

The controversy between advocates of quota sampling and probability sampling was played out in the pages of the *Public Opinion Quarterly* in the ensuing several years. The articles by Norman Meier and by Philip Hauser (1950) are illustrative. Who won? It is clear that probability sampling of populations became dominant in surveys sponsored or conducted by the federal government, not the least of which were the sample surveys conducted by the Bureau of the Census itself. Governmental surveys required a degree of precision for policy purposes less demanded by soundings of public opinion.

Assessment of Validity

The second methodological theme struck at the Central City conference was that of the assessment of validity. It was first addressed by Herbert Hyman. Hyman (1944–1945), asking, "Do They Tell the Truth?," described empirical studies comparing respondents' replies to questions of fact, such as whether they had purchased war bonds, with independent records. He noted that persons with more schooling were not always more likely to tell the truth and that social class values operate in the reporting of facts: people with more schooling are more likely to overreport the possession of library cards while those with less schooling are more likely to overreport the possession of drivers' licenses. Hyman's studies would later be replicated by the National Opinion Research Center (NORC) in other contexts, using factual items such as home ownership, telephone ownership, library card possession, driver's license possession, voting registration, and the like (Parry and Crossley, 1950). The approach used by Hyman and later by NORC in assessing the reporting of facts was analogous to that of the courtroom, in which testimony is compared with other sources of evidence for the same facts in order to assess truthfulness. Some thirty years later, Al

Biderman (1981) was to lead a session at AAPOR introducing the work of cognitive psychologists in an effort to understand how people organize their memory in responding to factual questions.

In some instances, attempts to assess validity in the reporting of attitudes are assisted by the use of independent records. Not a new issue in psychology, this had been addressed almost a half century earlier by Terman, in intelligence testing when he correlated test scores with school grades and teacher ratings where a reasonable degree of association was to be expected. The notion of internal consistency had also been introduced by psychometrists (Kuder, Richardson, et al.), in which responses to parts of a set of items were correlated with other parts of that same set, odds with evens, for example. The argument was that if items of the same set correlated highly with each other they could be interpreted as measuring the same thing. These measures of internal consistency came to be called measures of reliability.

A few years before the Central City conference the work of Louis Guttman in the Armed Forces Morale Attitudes Research Branch had produced a new approach. Guttman (1944) provided a mathematical method for testing the hypothesis that a set of items purporting to measure a single, specified attitudinal dimension were drawn from the universe of items that theoretically populated that dimension. His formulation differed from that of the earlier psychometrists. A measure of the unidimensionality, or reproducibility, of a set of attitude items was proposed by Guttman. These notions of Guttman were elucidated at the Central City conference by Jack Elinson, a colleague of Guttman's in the Research Branch of the U.S. Army headed by Samuel Stouffer during World War II. Stouffer was to become one of the earliest presidents and award winners (Julian Woodward Award) of AAPOR.

These two approaches to validity, addressed at the Central City conference in 1946, persist forty years later, with variations, of course. The planners of the Central City conference got to the heart of the matter at the beginning.

Data Collection

The third methodological theme taken up at the inaugural Central City conference was that of data collection, more specifically, the quality of interviewing and interviewer training. At issue was the procedure of hiring and supervising by remote control (mail, telephone, and telegrams) from a single location (research headquarters), versus the di-

rect, face-to-face recruiting, training, and supervising of interviewers in the field. Again, as in the case of sampling, speed and cost were major issues. A not infrequent mode of carrying out a field survey was for the research director to engage or otherwise contract the services of social science professors in colleges around the country who would find the requisite number of students to do the interviewing. Thus, everyone would benefit, keeping costs and time spent per survey at a minimum.

In the early efforts to improve the quality of sample surveys there was no way to escape higher costs. Sampling theoreticians like Stevens Stock and Lester Frankel (1939) would argue that the efficiency of a sampling design could be measured by calculations of cost per unit of information obtained, a way of assessing costs that was not the same as out-of-pocket expenditures. Stock and Frankel argued from a statistical perspective that the appropriate question to answer in a sample survey was not simply "How much was spent per interview?," but rather, "How much did it cost to reduce the error variance?"

For many years, one of the original Central City participants, Paul Sheatsley, contributed a great deal to the art and science of interviewing in survey research. His chapter (with Arthur Kornhauser) on "interviewing in social research," the standard reference to the subject, appeared in successive editions of *Research Methods in Social Relations*, the textbook sponsored by the Society for Psychological Research on Social Issues (Kornhauser and Sheatsley, 1976). CATI (computer-assisted telephone interviewing), for all its vaunted economic advantages and efficiencies, still needs to confront the basic elements of the interviewer-respondent interaction first elucidated by Sheatsley. The art of questionnaire construction as enunciated by Sheatsley appeared as recently as 1983 as "Questionnaire Construction and Item Writing," in Peter Rossi, James Wright, and Andy Anderson, editors, *Handbook of Survey Research* (Academic Press, New York, 1983).

Technological changes have made possible variations in data collection procedures by interviewing. The most important of these is the increased use of the telephone (Waksberg, 1978; Groves and Kahn, 1979). It is ironic that the famous 1936 *Literary Digest* poll failed primarily because it depended on a mail survey of telephone subscribers and automobile owners. The failure of the *Literary Digest* poll made way for the rise of the fathers of modern-day polling, Archibald Crosley, George Gallup, and Elmo Roper. These pioneer pollsters, devotees of quota sampling, eschewed telephones in favor of face-to-face interviewing. Today access to the public telephone is sufficiently universal to encourage the use of telephone interviewing for many kinds of sample

surveys. Indeed, access by telephone can be superior to face-to-face interviewing (in homes or at the doorstep), especially in the hazardous neighborhoods of large cities. Although some households protect their privacy with unlisted telephone numbers, this can be circumvented by random-digit dialing. Telephone surveys also came to be used in surveys of elite members of the population—for example, medical doctors.

Studies of interviewing were in great vogue in the years immediately following the establishment of AAPOR. Among the best known were those carried out by Herbert Hyman and his colleagues (Hyman, et al., 1954) at the National Opinion Research Center. At the Bureau of the Census, influential contributions were made by Eli Marks and others (Mauldin and Marks, 1950) in their demonstration that sample surveys conducted by relatively small numbers of well trained and supervised interviewers could be more accurate than censuses conducted by large numbers of less well trained and supervised enumerators. The post-enumeration sample surveys conducted by Marks and his colleagues at the Bureau of the Census were used as a check on the accuracy of the census. It was through these sample surveys that the disproportionate underenumeration of segments of the population, such as young black adult males, was discovered. This demonstration by government statisticians contributed greatly to the reputation of sample surveys.

Data Analysis

Methods of data analysis were hardly discussed at Central City. This was to be left for later years and blossomed with the advent of computers, which replaced electrically driven desk calculators and punch-card sorting and tabulating machines. The breakdown of poll results by demographic segments of the population was, however, early regarded as important for understanding and prediction. Demographic analysis produced a dramatic moment at Central City when Joe Belden raised an ethical issue: should he publish the results of his pre-election Texas poll that showed a high proportion of blacks in favor of the liberal candidate, when he feared that by so doing the liberal candidate might lose votes in the actual election?

Question-Wording

AAPOR members discussed problems of question-wording vigorously at annual conferences, in the pages of the *Public Opinion Quar-*

terly, in numerous scholarly articles in social science journals, and in some noteworthy books. Pollsters felt the pressure of social psychologists and sociologists disposed to use multiple questions and attitude scales. Seeking to get the most information with the fewest possible questions, Gallup (1947) proposed the "quintamimensional design." Using only five questions, the q.d. attempted to get at respondents' knowledge about an issue, interest in it, attitude toward it, reasons for the attitude, and strength of opinion.

Bias in question-wording, often laid at the door of issue advocates, versus neutrality in question-wording, as striven for by pollsters, occasionally led to absurd positions, such as: a neutral question is one that divides the population 50%-50%! In contrast, based on the idea that any question was a sample question drawn from a universe of questions pertaining to a single dimension, Louis Guttman and Edward Suchman (Suchman and Guttman, 1947) proposed an empirical, objective, and invariant method of determining the most neutral item in an attitude scale. Applying the second component of attitude scale analysis in Guttman's schema, that is, "intensity," Guttman and Suchman suggested that the most neutral item or question was one that tapped an attitude scale position that was held, on average, with the lowest intensity. Efforts to frame neutral, unbiased questions still continue. The goal appears to be: what is the most unbiased way one can frame a question on an issue for the purpose of a public referendum?

The same issues of question-wording effect that were addressed by Stanley Payne, and earlier by AAPOR award winners Hadley Cantril (1947) and Daniel Katz (1944) continue to engage social scientists. Among the more recent examinations of this subject are *Measurement of Subjective Phenomena*, edited by Denis Johnston (Johnston, 1981), and *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*, by Howard Schuman (an AAPOR president from 1985 to 1986) and Stanley Presser (1981). Volumes intended to bring "the art of asking questions" up to date were produced by Patricia Labaw, author of *Advanced Questionnaire Design* (1980); by Seymour Sudman (AAPOR president from 1981 to 1982) and Norman Bradburn, authors of *Asking Questions: A Practical Guide to Questionnaire Design* (1982) and by Charles Turner and Elizabeth Martin, in *Surveying Subjective Phenomena* (1985).

In a debate organized by Harry Alpert (an AAPOR president), Samuel Stouffer (an AAPOR president) and Rensis Likert argued the relative merits of closed and open questions. This was such a serious methodological issue that Paul Lazarsfeld (an AAPOR president and award

winner) had offered to negotiate the matter in an often-quoted article in the *Public Opinion Quarterly* (1944). Furthermore, while Stouffer, Lazarsfeld, and Alpert all became presidents of AAPOR, Likert, not a frequent attendee of AAPOR meetings, did not. Indeed, until relatively recently, the Institute of Social Research at the University of Michigan, founded and led by Likert for many years, maintained a somewhat aloof attitude toward AAPOR's pollsters, its market researchers, and its advocates of closed-ended questions. (Likert, like Lazarsfeld and Stouffer before him, was nevertheless an AAPOR award winner.)

For a number of years, a highlight of an AAPOR conference was its exciting "author meets the critics" sessions. Not infrequently the critic was Shirley Star, a posthumous AAPOR award winner. Little published herself, Star often performed the role of caustic critic of acclaimed published works. Among the authors subjected to her devastating methodological criticism were AAPOR award winner Stanley Payne and his *The Art of Asking Questions* (1951); Angus Campbell and his *The Quality of American Life* (1976); and Howard Freeman and Ozzie Simmons and their prize-winning (Hogg Foundation) work, *The Mental Patient Comes Home* (1963). In one characteristically flamboyant display of scholarly chutzpah, Star selected a statistical table and claimed that her analysis would show that the data provided support for an opposite hypothesis!

While crosstabulations in data analysis have not been completely abandoned, we must acknowledge that the computerization of data processing has led AAPOR members to different, and higher, levels of methodological sophistication in this arena. The displays of computer machinery commonly seen at AAPOR's latter-day conferences were initiated by Robert Lee (an active AAPOR member and holder of various Executive Council offices), who had had a long stint at IBM. The hardware and statistical software packages displays, occasionally accompanied by didactic sessions on their comparative advantages and disadvantages, are currently normal expectations of AAPOR conference attendees. Computer salespeople and their wares have provided AAPOR conferences with a distinctly commercial, as well as educational, aspect. The ease of applying standard statistical routines has led many an AAPOR member into the woods of multiple regression, analysis of variance, factor analysis and even LISREL, with a sometimes casual regard for the statistical assumptions implied in the application of these methods.

Methodology and the Mass Media

AAPOR can point with pride to its role in contributing to the methodological sophistication of the mass media in the reporting of polls and sample surveys, and through the media to the education of the public. Nonetheless, the level of reporting on methodology by the media remains lower than what some methodologists would like to see. For example, the media rarely report response rates.

The better newspapers now carry little boxes with their poll stories ritually detailing sample sizes and expected degrees of statistical confidence that should be attached to the reported results. The networks, particularly CBS News, always include the items of disclosure called for by the AAPOR Code of Professional Ethics and Practices. Not infrequently the exact wording of questions used is also stated. The better newspapers and networks supply background reports to interested scholars and archive their polls at Michigan's Interuniversity Consortium for Political and Social Research (ICPSR) and Connecticut's Roper Public Opinion Research Center.

In general, the relationship between journalists and pollsters has improved quite a bit. Reporters seem to be less threatened by pollsters than they used to be. Straw polls, that is, "unscientific polls," are more likely not to be reported, and if they are, usually with disclaimers as to their scientific import.

Some poll reporters, like Adam Clymer of the *New York Times* and Barry Sussman, formerly of the *Washington Post*, became regular attendees, and even participants, at AAPOR conferences. Philip Meyer, a former journalist and currently Kenan Professor of Journalism at the University of North Carolina, was an AAPOR president, as was Warren Mitofsky of CBS News. While the little boxes carried by obedient newspapers may be a far cry from what Stuart Dodd had proposed to AAPOR as standards for sample survey reporting, they are half a loaf, or at least a few slices, of methodologically sound bread.

Methodology and Social Science

Many of AAPOR's presidents have made major contributions to methodology that have affected social science research beyond sample surveys and polls. Among them are: Herbert Hyman (1954, 1955) on interviewing and analysis; Bernard Berelson (1952) on content analysis; Howard Schuman (1981) on question-wording; Warren Mitofsky

(1970) on the Mitofsky-Waksberg telephone sampling method; Samuel Stouffer (1950) on research design; and Paul Lazarsfeld (1938, 1950) on panel analysis and latent attribute analysis. Stouffer and Lazarsfeld were also recognized for their contributions to social science by also serving as presidents of the American Sociological Association. All, it is worth noting, while methodologically sophisticated, are equally well known for their contributions in substantive areas of social science.

References

- Berelson, Bernard (1952) *Content Analysis in Communication Research*. Glencoe, IL: Free Press.
- Biderman, Albert D. (1981) "Survey research meets cognitive psychology: Dalliance or alliance?" Paper presented at the annual meeting of AAPOR, Buck Hill Falls, Pennsylvania.
- Campbell, Angus, Philip E. Converse, and Willard L. Rogers (1976) *The Quality of American life: Perceptions, Evaluations, and Satisfaction*. New York: Russell Sage Foundation.
- Cantril, Hadley (1940) "Experiments in the wording of questions." *Public Opinion Quarterly* 4:330–332.
- Cantril, Hadley and Associates (1944) *Gauging Public Opinion*. Princeton: Princeton University Press.
- Colombotos, John (1969) "Personal versus telephone interviews: Effect on responses." *Public Health Reports* 84:773–782.
- Converse, Jean M. (1987) *Survey Research in the United States: Roots and Emergence, 1890–1960*. Berkeley: University of California Press.
- Freeman, Howard E. and Ozzie G. Simmons (1963) *The Mental Patient Comes Home*. New York: John Wiley and Sons.
- Gallup, George (1947) "The quintamensional plan of question design." *Public Opinion Quarterly* 11:385–93.
- Groves, Robert M. and Kahn, R.L (1979) *Surveys by Telephone*. New York: Academic Press
- Guttman, Louis (1944) "The basis for scaling qualitative data." *American Sociological Review* 9:139–150.
- Hauser, Philip M. (1950) "Some aspects of methodological research in the 1950 census." *Public Opinion Quarterly* 14:5–13.
- Herlihy, David and Christiane Klapisch-Zuber (1984) *The Tuscans and Their Families: A Study of the Florentine Catasto to 1427*. New Haven, CT: Yale University Press.

- Hyman, Herbert H. (1944–1945) “Do they tell the truth?” *Public Opinion Quarterly* 8:557–559.
- Hyman, Herbert H. (1955) *Survey Design and Analysis: Principles, Cases and Procedures*. Glencoe, IL; Free Press of Glencoe.
- Hyman, Herbert H. et al. (1954) *Interviewing in Social Research*. Chicago: University of Chicago Press.
- Johnston, Denis F. (1981) Measurement of subjective phenomena. (Special Demographic Analyses CDS 80–3) Washington, D.C., U.S. Bureau of the Census.
- Katz, Daniel (1944) “The measurement of intensity.” In Hadley Cantril and Associates: *Gauging Public Opinion*. Princeton: Princeton University Press.
- Kornhauser, Arthur and Paul B. Sheatsley (1976) “Questionnaire construction and interview procedures.” In Claire Selltitz, Lawrence S. Wrightsman, and Stuart W. Cook, *Research Methods in Social Relations*. Third edition. New York: Holt, Rinehart and Winston.
- Labaw, Patricia J. (1980) *Advanced Questionnaire Design*. Cambridge, MA: Abt Books.
- Lazarsfeld, Paul F. (1944) “The controversy over detailed interviews—an offer for negotiation.” *Public Opinion Quarterly* 8:38–60.
- Lazarsfeld, Paul F. (1950) “The logical and mathematical foundations of latent structure analysis.” In Samuel A. Stouffer et al., *Measurement and prediction*. (Studies in Social Psychology in World War II.) Princeton: Princeton University Press.
- Lazarsfeld, Paul F. and Marjorie Fiske (1938) “The ‘panel’ as a new tool for measuring opinion.” *Public Opinion Quarterly* 2:596–512.
- Mauldin, W. Parker and Eli S. Marks (1950) “Problems of response in enumerative surveys.” *American Sociological Review* 15:649–657.
- Meier, Norman C. and Cletus J. Burke (1947) “Laboratory tests of sampling techniques.” *Public Opinion Quarterly* 11:586–593.
- Mitofsky, Warren (1970) “Sampling of telephone households.” *Unpublished CBS memorandum*.
- Parry, Hugh J. and Helen M. Crossley (1950) “Validity of responses to survey questions.” *Public Opinion Quarterly* 14:61–80.
- Payne, Stanley L. (1951) *The Art of Asking Questions*. Princeton: Princeton University Press.
- Riley, Matilda White (1973) “Aging and cohort succession: Interpretations and misinterpretations.” *Public Opinion Quarterly* 37:35–49.
- Schuman, Howard and Stanley Presser (1981) *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording and Content*. New York: Academic Press.

- Sheatsley, Paul B. (1983) "Questionnaire construction and item writing," pp. 195–230 in Peter H. Rossi, James D. Wright and Andy B. Anderson (eds.), *Handbook of Survey Research*. New York: Academic Press.
- Stock, J. Stevens and Lester R. Frankel (1939) "The allocation of sampling among several strata." *Annals of Mathematical Statistics* 10:288–293.
- Stouffer, Samuel A. et al. (1950) *Measurement and prediction*. (Studies in Social Psychology in World War II, Volume 4.) Princeton: Princeton University Press.
- Suchman, Edward A. and Louis Guttman (1947) "A solution to the problem of question bias." *Public Opinion Quarterly* 11:445–455.
- Sudman, Seymour and Norman M. Bradburn (1982) *Asking Questions: A Practical Guide to Questionnaire Design*. San Francisco: Jossey-Bass.
- Terman, Lewis M. (1919) *The Intelligence of School Children*. Boston: Houghton Mifflin Co.
- Turner, Charles F. and Elizabeth Martin (1984) *Surveying Subjective Phenomena*. (2 volumes) New York: Russell Sage Foundation.
- Waksberg, Joseph (1978) "Sampling methods for random digit dialing." *Journal of the American Statistical Association* 73(361):40–46.