Standardization in Survey Research:
How Did it Get Here and How Far Should it Go?

Paul Beatty
National Center for Health Statistics, Centers for Disease Control and Prevention
6525 Belcrest Rd., Room 915
Hyattsville, MD, USA
pbeatty@cdc.gov

Standardization provides survey research with some of its great strengths, while simultaneously serving as one of its greatest liabilities. It is a strength because standardization makes it possible to seriously consider surveys as measurement tools. Studying humans is always difficult and involves some subjectivity, but the potential for *measurement* gives survey research great scientific credibility. It also allows for efficiency in handling data, provides for clear quality control criteria, and minimizes random error introduced by the interviewer.

At the same time, standardization puts unusual demands on the interaction between the interviewer and the respondent. The interviewer=s speech is highly constrained-- ideally she merely reads questions verbatim. When the respondent fails to provide answers that conform to the desired response format, interviewers are expected to rely upon a small set of probes to return the interview to its intended course. Similarly, if a respondent asks for clarification, the interviewer=s possible responses are generic and prescripted so that there is little chance of introducing variation into the interaction. While the survey interview ostensibly resembles an everyday exchange of information, the parallels have limits: most interactions between individuals allow both participants to ask questions and to Anegotiate@ about the meaning of utterances in a relatively unfettered manner, resources that are absent from surveys. Some have suggested that these constraints make the survey interview so contrived that it is no longer viable for collecting information of any value about social phenomena (Suchman and Jordan 1990).

In spite of these criticisms, most statisticians remain skeptical about approaches not based on standardization. Such approaches would introduce an unknown degree of interviewer variation, which is precisely what standardization is designed to prevent. Furthermore, even when statisticians acknowledge the communication difficulties imposed by the survey interview, they may not be convinced that these difficulties actually affect the quality of their statisticsB at least not on a scale that would justify allowing new variation that could undermine the sanctity of the measurement process. From this perspective, the burden of proof that changes are necessary is on advocates of alternatives.

Debates on this subject will not be resolved in this paper; nor will this paper provide statistical evidence that standardization adversely affects the validity of statistics. Rather, it will illustrate a partial disconnect between the current practice of standardized interviewing and its original design, philosophy, and assumptions. I do this not to demonstrate that standardization is Abad@ or impossible to implement, but rather to show that standardization may have evolved into something different than originally intended. A closer look at the evolution of standardized interviewing practice suggests a few modifications that may be necessary to make it Awork@ effectivelyB without throwing out the overall paradigm.

Survey researchers originally embraced interviewers for characteristics that are shunned todayB their ability to tailor their communication as appropriate to particular interviews. In the 1930's, interviewing methods pioneered by Rensis Likert served as a model for emerging practices of survey data collection. In earlier years, questions had not been formally written at allB interviewers simply worked from a list of topics. The advantages of formally scripted questions became obvious, but even when adopted, interviews were somewhat conversational. Questions were open-ended in order to avoid what Likert referred to as Aartificially focused alternatives.@ He also argued that Athe words may vary so long

as the question in each case represents exactly the same idea@ (cited in Converse, 1987, p. 158). Interviewers also maintained flexibility in determining when questions had been answered, and were empowered to rephrase questions under some circumstances.

From a modern perspective, this sort of interviewing seems risky, but Likert saw this freedom as absolutely necessary. He was an outspoken critic of alternative methods based on completely closed questions with fully scripted behavior, as practiced in the polling industry. One concern had to do with communicationB he did not believe it was possible to eliminate the interviewer=s role in obtaining accurate answers. Questions and respondents were rarely a perfect fit; it took a human interviewer to negotiate between them. A second but related concern had to do with rapport. Telling respondents that they needed to Arestrict their thinking... to the dimensions seen by the experimenter@ suggested that the researcher was uninterested in their perspective. As a result, respondents= commitment to meaningful participation would be severely diminished (Likert 1947, p. 199). Likert=s position was influential, particularly at the University of Michigan=s Survey Research Center, which he helped to establish. Interviewers there were trained that Agood rapport must be maintained throughout the interview to insure full and valid information@ and to Anot obviously read@ each question in order to avoid Aan atmosphere of interrogation.@ Rather, the interviewer should set the tone that she was Aextremely interested in having the respondent=s ideas@ (Survey Research Center, 1954, pp. 31 and 34).

Of course, the logistical advantages of standardization were understood even then. Actual practice represented a compromise between the twin needs of efficiency and accuracy, and an equilibrium continued for some time. Yet eventually the advantages of increased standardization started to outweigh the disadvantages. This was partially due to technological changes. The growth of telephone interviewing in the 1970s (and later, computer-assisted interviewing) made it possible to tabulate and analyze data with much greater speed. Simultaneously, data needs became much more specific, moving from open-ended questions to very precise ones (e.g., about number of doctor visits during the past month, including trips to an emergency room but not counting overnight hospital stays). It became increasingly important to read questions exactly as written in order to capture all relevant details.

These potential advantages put an increased burden on advocates of less standardized interviewing to empirically demonstrate its advantages. For the most part, they failed to convincingly defend their position. The concept of rapport came under particular attack. Goudy and Potter (1975) argued that there was no evidence that rapport improved data quality, and called for an alternative focus on Ainterview effects currently measurable.@ Weiss (1968) provided evidence that good rapport might actually decrease data quality. The demise of rapport opened the door for the efficiency that standardization offered. Simultaneously, it became increasingly possible to monitor and enforce standardized behavior in centralized interviewing facilities. (For a more thorough overview of these developments, see Beatty, 1995).

Yet standardization had become accepted in survey research with some tacit assumptions about how complex questions would actually be. While no one actually laid out guidelines of Amaximum allowable complexity,@ even ardent advocates of standardization realized that there was some point where its effectiveness would break down. Even when interviewers were instructed to read questions verbatim, it was generally recognized that minor bending of the rules might take place. Standardization as recently practiced puts communication demands on participants that were not fully anticipatedB partially because the rules were designed for either general or simple questions, and partially because strict enforcement was not anticipated.

Most recent criticisms of standardized interviewing focus on the difficulty to Amediate uncertainties of relevance and interpretation@ given its constraints (Suchman and Jordan, 1990, p. 232). Non-standardized paradigms (e.g., proposed by Mishler, 1986) ostensibly overcome these deficiencies. Unfortunately, the alternatives are inadequate for collecting timely, population-based statistical data. The

resulting debate has been somewhat of a stalemate, with neither side proposing solutions that meet the needs of the other. In a recent exception, Schober and Conrad (1997) proposed a Aflexible@ interviewing strategy in which the interviewer has some freedom to provide non-standardized guidance. Such interviewing improved the accuracy of respondent reports in some circumstances. Yet Schober and Conrad present their Aflexible@ interviewing style as a stronger departure than it actually is. Their innovation is in correcting communication problems within the interview while maintaining as much standardization as possible. Further progress can be made if researchers increasingly focus on identifying actual communication problems within standardized interviews.

Clearly, some problems are the result of badly written questions. In less standardized interviews, the burden of clear communication rests upon the interviewer, but standardization shifts the burden to the questions. Unfortunately, knowledge about writing good questions has not always kept pace with changes in interviewing practice. In fact, some criticisms of standardization can be more accurately viewed as criticisms of standardized interviewing with poor questions. But recently, great progress has been made in establishing procedures for diagnosing questionnaire problems in cognitive laboratories. When properly applied, these techniques improve questionnaires and make standardization more viable.

Yet it is unlikely that *all* problems with standardization can be resolved in this manner. One problem is that many survey questions are straightforward for most respondents, but ambiguous for a particular subset. The question AAre you currently being treated by a doctor for arthritis?@ is one example. It is straightforward for respondents who are undergoing current treatment (e.g., frequent doctor visits) and for those who have no arthritis at all. But it is more complicated for people whose circumstances may or may not qualify as Acurrent treatment,@ e.g., those who received guidance from a doctor about arthritis a year earlier. Adding detailed definitions might solve such difficulties, but this might also make the question more complicated or burdensome for the majority who have no difficulties with the original.

However we write the question, interviewer intervention will be necessary to help some respondents. If respondents express uncertainty of meaning, interviewers traditionally resort to neutral statements such as Awhatever it means to you.@ Of course, this also has problems: respondents with identical circumstances could have different interpretations, and provide different answers to the question. The approach is also socially awkward. A slight shift in practice would entail training interviewers to identify, acknowledge, and respond directly to the source of confusion. In the example above, the respondent is uncertain whether his previous visit qualifies as Acurrent@ treatment. The best solution would be to informally help the respondent reach a decision based on an understanding of the question=s meaning.

Another survey asked respondents to agree or disagree with a series of statements about Awhether the following were reasons you did not get a flu shot during pregnancy.@ One such statement was AI wanted to avoid medications during pregnancy.@ As in the example above, many respondents had clear yes or no responses. Other respondents= circumstances were more complicated: they did have a general reluctance to avoid medications during pregnancy, but were never actually offered a flu shot and did not think about getting one. An Aagree@ response accurately reflects their attitudes, but a Adisagree@ response accurately reflects their decision processes. Here also, interviewers are usually expected to provide neutral statements. Yet an interviewer who was informed about the intent of the question might solve the problem quickly given some flexibility (e.g., AIf you actually decided not to get this shot because of your pregnancy, I=d say your answer should be yes. If this decision didn=t actually come up in this way, I=d say to answer no.@) The exchange would still qualify as measurement, but would rely on some limited conversational resources to resolve the ambiguity.

Another survey question asks ADo you have difficulty reading newspaper print even when wearing glasses or contact lenses, if that is how you see best?@ Often, respondents give unusable answers such as

Aactually, glasses make my reading worse.@ Common interviewer instructions call for repeating the question to get around the impasse. Yet under norms of conversation, the interviewer should directly address the source of the problem (an objection to the premise) rather than to point out the inadequacy of the response. A more constructive reply might be: ASo, you read best without glasses. When you=re not wearing them, do you have difficulty reading newspaper print?@ This solution acknowledges the problem and redirects communication based on the previous exchange. It does call for some improvisation, but is not leading.

Survey interviewers have a clear mandate to successfully complete interviews, and learn how to do so. When one observes actual field or telephone interviews B especially in a non-supervisory capacity B it becomes obvious that these adaptive behaviors already occur more frequently than many researchers realize. If we acknowledge this, then we can study what interviewers actually do and use it to our advantage, and help them to perform in a non-leading manner. Those who view this as an undesirable new introduction of error fail to recognize that strict standardization also causes errors which may be *more* serious. Modest adaptations do not represent a radical departure of interviewing style, but are consistent with the original objectives of standardized interviewing B to balance measurement and communication. They acknowledge the need for verbatim questions and non-leading behavior, but employ interviewers as intelligent observers and solvers of communication problems. As a form of interpersonal communication, survey interviewing cannot be completely scripted. Acknowledging this could lead to improvements in data quality while still maintaining an overall paradigm of measurement.

REFERENCES

Beatty, P. (1995). Understanding the Standardized/Non-Standardized Interviewing Controversy. Journal of Official Statistics, 11, 147-160.

Converse, J.M. (1987). Survey Research in the United States: Roots and Emergence 1890-1960. Berkeley: University of California.

Goudy, W.J., and Potter, H.R. (1975). Interviewer Rapport: Demise of a Concept. Public Opinion Quarterly, 39, 530-543.

Likert, R. (1947). The Sample Interview Survey: A Fundamental Research Tool of the Social Sciences. In Current Trends in Psychology, ed. W. Dennis. Pittsburgh: University of Pittsburgh.

Mishler, E.G. (1986). Research Interviewing. Cambridge: Harvard University Press.

Schober, M. and Conrad, F. (1997). ADoes Conversational Interviewing Reduce Survey Measurement Error?@ Public Opinion Quarterly, 61, 576-602.

Suchman, L., and Jordan, B. (1990). Interactional Troubles in Face-to-Face Survey Interviews. Journal of the American Statistical Association, 85, 232-241.

Survey Research Center, The University of Michigan (1954). Manual for Interviewers. Ann Arbor: University of Michigan.

Weiss, C.H. (1968). Validity of Welfare Mothers' Interview Responses. Public Opinion Quarterly, 32, 622-633.