UNDERSTANDING THE STANDARDIZED/NON-STANDARDIZED INTERVIEWING CONTROVERSY

Paul Beatty, The University of Michigan
Institute for Social Research, Box 1248, Ann Arbor, MI 48106-1248

Almost fifty years ago, Paul Lazarsfeld addressed a growing controversy between two seemingly irreconcilable schools of thought in the survey field. Some researchers believed the only way to obtain useful survey information was through loosely structured "in depth" interviews; others countered that there was no advantage to this cumbersome method and more structured surveys were highly preferable. Lazarsfeld's "Negotiation" between the two viewpoints (1944) is widely regarded as a classic, finding common ground that future researchers could work together on. Ironically, today we are witnessing a controversy that is, although different in some important ways, strikingly similar in others. Some researchers claim that the standardized interview has failed: after many years of attempted developments, standardized interviewing remains too flawed to be useful, inherently riddled with interviewer-respondent communication problems. While a great majority continue to support standardized interviews, admitting imperfections while defending the general strength of the methodology, the debate seems to be growing. Some wonder if the standardized interview has run its course and will gradually fade from view. Many agree that the future of standardized interviewing is difficult to predict, but that in any case, there is no room for compromise between the two opposing schools of thought. This view, I argue, is incorrect.

If one understands the conflict between the two viewpoints, there is a quite logical compromise. However, understanding the evolution of this conflict is a somewhat tricky matter. Looking at methodological pieces from the past is insufficient, as they provide little insight into how interviewing was actually conducted; likewise, technical manuals from the past explain little of the basis for the strategies they employed. To completely understand the forces that have moved surveys toward standardization, as well as those that have resisted, it is necessary to look at both works together. When this is done, the path that standardization has taken becomes much clearer; it is fairly simple to

understand what issues have been left unresolved in the standardization debate and see what common ground exists between the two schools of thought. Though no one can predict the future with certainty, the most logical directions that will be followed seem surprisingly obvious.

Standardization was developed to gain tighter control of the errors produced by interviewers, through the precise scripting of questions and development of standardized behavior appropriate for an interviewer to follow in all situations. The goal, as noted by Groves (1989), is "nothing less than the elimination of the interviewer as a source of measurement error." This has not, however, always been a priority in survey research; in fact, this goal developed quite gradually.

Changes in the level of question scripting and interviewing standardization were motivated by the same debate: should interviewers be "knowledgeable experts" who "conversationalized" respondents through interviews, using questionnaires only as guidelines; or should interviewers be non-experts who obtained information on a mass-scale using standardized techniques? This debate goes back at least as far as the 1930's when interviewers were first used for academic research. Most researchers viewed interviewers as a means to escape the restrictions of closed questionnaires (as used in mail surveys) and were anxious to use all the potential freedom that interviewers provided. Still, no one denied that at least a little "fixation" was required for the research to be useful. At the Bureau of Applied Social Research, Lazarsfeld believed that this fixation should apply only to the meanings of questions, not the questions themselves. Other research centers acknowledged the need for standardization more quickly; by 1945, NORC had made major strides toward standardizing survey interviews and questionnaires-- interviewers there were given very explicit instructions regarding standardized question asking. The University of Michigan's SRC fell somewhere in the middle, holding a compromise position throughout this prevailing but gradual move to standardization (Converse, 1987).

An interviewing manual from SRC (1954) provides some insight into the continuing

compromise position there, balancing such statements as "the questionnaire helps to standardize the interview," with instructions such as "Use the questionnaire, but use it informally"; another set of instructions reads "Ask the questions exactly as worded on the questionnaire", while a few pages later, the interviewer is empowered to "reword the question slightly. . . as a last resort" to clear up misinterpretations. The manual instructs interviewers work in a dual role, as a "technician who applies standard techniques to each interview," while maintaining the role of "human being who builds up a permissive and warm relationship with each respondent." On the one hand, we see a genuine desire to embrace standardization-- even though interviewers are given some leeway, they are instructed to use it with great caution. There are many mentions throughout the manual that standardization is a key component of surveys' usefulness. "Obviously," states the manual, "if a question is differently worded for different respondents, it will not yield comparable results." Given a statement such as that, it is surprising that interviewers maintained as much discretion as they did. It is also clear, however, that another issue is considered of greater importance for the success of survey interviewing: rapport between the interviewer and respondent. Furthermore, maintaining rapport and standardization were often seen as contradictory goals.

The word "rapport" appears dozens of times throughout the manual, most prominently in this statement of purpose: "Your goal is 'rapport' with the respondent... [This] designates the personal relationship of confidence and understanding between the interviewer and respondent which provides the foundation for good interviewing. Elsewhere are whole sections designed to help maintain rapport ("the interviewer recognizes that good rapport must be maintained throughout the interview to insure full and valid information") that explain its many benefits (such as "good rapport stimulates discussion"). In short, the 1954 Manual for Interviewers stresses the importance of rapport above virtually anything else. Indeed, methodological pieces agreed that the concept was important: Hyman's Interviewing in Social Research, published the same year, emphasized that rapport helped to ensure response validity. It is also critical to understand that too much standardization was considered to be at odds with rapport building. Consider this excerpt from the 1954 manual:

You should keep the questionnaire in sight

during the interview, glancing at it before asking each question. Put each question to the respondent in a natural and conversational tone of voice, not obviously reading it. Try to avoid drawing too much attention to the questionnaire since your goal is to set up a friendly relationship between yourself and the respondent; too much obvious attention to the questionnaire makes for an atmosphere of interrogation, which is something you want to avoid. Each question should be asked in a manner implying that it presents an interesting topic, and that you are extremely interested in having the respondent's ideas on it.

Thus, although standardization was seen as more desirable over time, the higher importance of rapport stood as a barrier to its development.

Nevertheless, very little hard evidence existed to show that rapport was necessary (Hyman neglected to do so) -- this was considered axiomatic. Actually, no one had a clear definition of what rapport was, and the fuzziness of the concept allowed it to be attacked eventually. Goudy and Potter (1975), citing various definitions used over the years, noted that no matter how rapport was defined, there was no evidence that it improved interviewing productiveness; in fact, citing an observation by Gordon (1969), they questioned whether rapport might actually be destructive: "Often the neophyte thinks he has conducted an excellent interview because 'rapport was perfect' and the respondent was 'completely at ease, talked spontaneously, and documented that she had enjoyed the interview.' Yet when the interview is analyzed for the amount and clarity of relevant data, it is found to be incomplete, superficial, and ambiguous." Goudy and Potter therefore concluded:

When we realize that problems exist in reading questions correctly, recording information, and coding that information [they cite multiple sources here] then studying rapport seems less important. Unless agreement on a conceptual and operational definition can be reached and unless elements in the interviewing situation purporting to constitute rapport can be isolated and tested, further empirical studies of rapport may be useless. . . [Researchers should] abandon the concept [of rapport], admit that interviewing is an art that may contain

certain scientifically controllable elements, and work to attain reliability and validity in interview data through interview effects currently measurable.

Thus, by the mid-seventies, survey researchers were questioning the wisdom of the rapport-building approach, turning their attention to "interview effects currently measurable". Research on controlling interviewer variation, previously hindered by concerns for rapport, had new opportunities to flourish.

Up until this point, "looser" forms of interviewing (and questions) had served two purposes: first, the format provided a diagnostic capability for communication problems, and second, it facilitated rapport-building. Of these purposes, rapport arguably attracted the most attention; with that gone, there was little standing in the way of standardized interviews. Researchers had always been curious about the errors introduced by interviewers; Hyman had devoted methodological attention to it, while others such as Kahn and Cannell (1957) studied interviewer variation from a more theoretical framework. Now, with the shift in priority, a wide variety of methodological work was created with the goal of stamping out interviewing error altogether. Such efforts focused on the development of training techniques and precise prescriptions for what interviewers are told to do (Groves, 1989). The results of such research may be examined most clearly in their practical application, though the 1983 interviewing manual from SRC, General Interviewing Techniques (GIT).

The GIT manual differs from the 1954 manual mainly in its emphasis on examples designed to develop precise, standardized behavior from interviewers. Gone is the 1954 manual's friendly prose describing the ultimate relationship between the respondent and interviewer, motivated by explanations about how the interviewer fits into the big picture. GIT is an ambitious technical document, three times the size of its 1954 counterpart, filled with expanded sections on question asking, clarification techniques, probing, feedback, and so on. Its purpose is not general description; it is intensely specific, containing dozens of examples and seven lengthier exercises for trainees. Consider the case of probing: the 1954 manual devotes 8 pages to the topic, offering a few possible probes to use in case rapport alone is insufficient. The 1983 GIT devotes 30 pages to the topic, providing a complete list of acceptable probes, with practice exercises to demonstrate

proper applications. Looking at such training materials, the attempted suppression of any interviewer variation is quite clearly documented. Nor does the attempt to control interviewer variation stop with initial training: it continues through procedures such as interview monitoring, brush up training, and so on. The level of standardization that has been reached is so great that some researchers have questioned whether such restrictions have gone too far. Less structured interviewing, as I mentioned previously, had two advantages. The most important of them, rapport, had been discredited. Yet the other, diagnosing communication problems, had never been adequately addressed.

It was, however, quite a simple problem to forget, considering the advantages that standardization brought. First of all, researchers learned how much faster standardization made the survey process, particularly important in an age when surveys are used to make election and economic forecasts, develop business strategies, and make other time-sensitive decisions. No one in academia wants to wait for survey data any longer than he(she) has to either; without standardization, it is difficult to conceive the amount of processing that would be necessary before data would be available for large academic research programs. Furthermore, standardization allowed more statistical precision, which enabled increasingly elaborate statistical methods to be applied (compare Stouffer's cross-tabulations from the 1950s to the latest LISREL analyses and log-linear models, which require much stronger assumptions about the data involved). Also, as we have seen, standardization gave survey researchers new power to study and attack interviewer error-- perhaps more attention than is desirable, considering all the possible sources of survey errors.

Lately, researchers have become increasingly concerned that standardized interviewing has strangled communication with respondents. The heart of the current problem is that interviewers have no discretion to exercise whatsoever, which is the basis of current attacks on the usefulness of the standardized interview. A possible alternative type of interviewing, as described by Groves (1989), dictates that "interviewers should be trained in the concepts inherent in the questions and be allowed to probe, rephrase, and adapt the questionnaire to individual respondent needs." Suchman and Jordan (1990) provide one of the better known arguments in favor of such de-standardization. They claim that the conflict between interviews as conversations and

interviews as data-collection instruments has never been adequately resolved; as they stand, surveys rely on conversational norms to succeed (i.e., asking and answering questions) while suppressing "interactional resources that routinely mediate uncertainties of relevance and interpretation [in conversations]." As such, they argue that "the standardized interview has become such a fragile, technical object that it is no longer viable in the real world of interaction"-- a harsh indictment indeed. Yet it should not be too surprising. The whole controversy regarding the most appropriate style of interviewing is due to a resurgent interest in diagnosing communication problems-- the second advantage of less standardized interviewing. In the frenzy to wipe out interviewer error, diagnosing communication problems has not received adequate attention. Suchman and Jordan, and others like them, are merely bringing the issue back to the forefront.

Critics of Suchman and Jordan might counter that though the survey interview does have certain problems, their representation of them is highly overstated. The examples they provide surely represent worst-case scenarios: rambling, pointless interactions between interviewers and respondents in which the interviewer follows no observable conventions for clarification of misunderstandings (such as repeat of question, providing definitions or "Whatever it means to you" guidelines, and neutral probes) and keeping the respondent focused (through feedback to emphasize the proper role of the respondent). Suchman and Jordan have not provided examples of the failure of proper interviewing to yield useable survey data; they have merely provided examples that interviewing, when performed poorly, can yield unusable data, which is hardly the same thing. Yet, though their specific examples miss the mark, their general notions cannot be as easily dismissed. As a result of the strict rules of GIT, de-emphasizing any intelligent contribution from interviewers, it is possible that meaning may not be standardized even though the literal wording of the question has been properly conveyed. For example, Suchman and Jordan describe one situation in which a respondent is very confused about what constitutes an "alcoholic beverage" (the respondent failed to treat wine as an alcoholic beverage, which seemingly contradicted previous answers); though the interviewer exacerbates the situation by following procedures that most survey researchers would consider inadequate, it is not clear that by using GIT techniques, an interviewer could have completely eradicated the problem-- at least given the way the questionnaire was written. There was clearly a problem in communicating correct meaning to the respondent that standardized interviewing did not address. Furthermore, some researchers argue that even if interviewers did have adequate discretion to address ambiguities and misunderstandings, they are unable to pay attention to such matters because these tasks have lower priority than following overly-detailed rules. All of this has led to the speculation that the standardized interview has run its course and outlived its usefulness, soon to be replaced by more conversational interviews. Theoretically, if interviewers were allowed more leeway to diagnose communication problems and clarify the meanings of questions, the goals of current survey research would be more successfully met: to standardize the meanings of questions in such a way that we may infer something about the population in question. Are we, then, seeing the beginning of the end of survey research as we know it, closing a period that future researchers might look back on as some "dubious flirtation with standardized interviewing"?

I would speculate that such a prediction is highly premature at best. If for no other reasons, the benefits of practical simplicity, timeliness, and greater statistical power have cemented standardization into the core of modern surveying. To deny this is to ignore the most practical facts of the matter.

This does not mean, however, that compromise is impossible and will not occur. We are, after all, interested most of all in reducing the total amount of error in surveys. For all the good that interviewing standardization brings, its benefits can be misleading. If attacking the slightest interviewer deviation brings about modest reduction of interviewer error-- but at the same time causes a greater increase in error from the respondent, who is not able to draw on the communicative resources of an informed, intelligent interviewer-- then the strategy is self-defeating. The reason we do not rely completely upon computers to actually carry out interviews over the phone or through increased self-administered methods is not because the technology is unavailable (CASIC, 1990); it is because the survey interaction remains an interaction between humans, which has certain advantages. It also has disadvantages. It is essential for survey researchers desiring the "total elimination" of interviewer error to accept that this is impossible-- at least without compromising research quality in more important ways. Deciding exactly how much and what kind of

interviewer freedom can be allowed is tricky business; certainly, following precise interviewing rules remains critical in order to collect the rewards standardization brings. However, standardization advocates must recognize that tightly prescribed interviewer behavior does have an ultimate point of diminishing returns.

On the other hand, those researchers who most desire the de-standardization of surveys should redirect their energies away from "interviewer liberation." The greatest strides in eliminating communication problems will be made not from freeing interviewers; that strategy merely trades one set of problems for another. Rather, the real question researchers must answer is: how can we solve communication problems while harnessing the full benefits of standardization?

Although no one has all of the answers to this question, some possible solutions are hardly new; others will require more thought and fine-tuning among survey users. First, the importance of pretesting, which should never be underestimated, will only grow. Through a pretest, one can assess communication difficulties in a questionnaire through responses from dozens of people with fresh perspectives-- but too often a pretest can be regarded as a convenience for "when time allows." This view must be continually discouraged if standardized surveys are to be successful. In addition, there are several techniques that obtain similar input through more systematic methods: cognitive interviewing, designed to help researchers understand how respondents interpret questions, will certainly receive even more attention. The volume of research recently published indicates that enthusiasm in this area continues to increase. Also, even more systematic, quantitative procedures such as behavior coding will grow (at least to the extent that such labor-intense activities can be carried out in a timely manner). Some researchers who design and analyze large-scale surveys have only cursory knowledge of these procedures, so there is hope that their impact on survey quality will continue to expand. Perhaps there will soon be additional pre-survey strategies to assess respondent interpretation problems.

In addition, researchers need to re-evaluate the sort of behavior they identify as "good interviewing" and reward interviewers for accordingly. They should keep in mind that the ultimate goal of standardized survey interviewing is to obtain complete answers to questions that are both administered in a uniform manner, and understood. Often, however, interviewer success during an interaction with a respondent is measured solely by how well the interviewer conforms to specifically prescribed behaviors: was that an acceptable probe? was that acceptable feedback given to the respondent? While these are important-- directive probes and inappropriate feedback can certainly damage data quality-- this approach is incomplete. It is possible that communication problems are increased because interviewers spend too much time making sure that they conform to highly specific criteria (which is what they are rewarded for), and not enough time evaluating whether the respondent understands and answers the question correctly (which is what they should be rewarded for). I suggest that this weakness can be overcome within the constraints of standardized methodology-- not by allowing much "interviewer leeway," but by making sure we are not overzealous in enforcing the smallest letter of the law when clear answers were obtained through generally acceptable means.

Researchers also need to spend more time teaching interviewers the concepts that are inherent in the questions they ask-- not so that interviewers can create conversations, as some have suggested, but so they can judge the appropriateness of the respondents' answers. Question-by-question specifications (Q-x-Q's) should be written more consistently, and more time should be spent to make certain that interviewers understand them. Too often, interviewers read the Q-x-Q's for the first time when a respondent indicates that (s)he does not understand a question. In our desire to standardize, we have essentially trained interviewers to follow algorithms in order to elicit responses. In the process, we may have deluded ourselves with unwarranted optimism that interviewers can do their job successfully with very little understanding of the questions they ask. These are several ways to address the problems noted by non-standardization advocates while maintaining standardization; continued thought and discussion within the survey research community will be important in developing similar strategies along these lines.

These are the areas where the most energy should be spent in our efforts to control communication error. Many of them take place before interviewers are even involved on a large scale. The others involve the same fundamentals of interviewing, but more attention to creating informed interviewers. Understanding miscommunication between interviewers and respondents remains a major priority. Correctly realizing that attention to this problem is needed, some have forecast the end of the standardized

1052

interview. While such arguments are overstated, they are also understandable. Over the last fifty years, survey researchers have radically changed their views of the ideal interaction between interviewers and respondents. Interviewers were initially valued for the freedom they brought to the research process, though researchers increasingly realized that freedom went hand in hand with interviewer-introduced error. Paralyzed for many years by the concept of rapport, researchers finally confronted interviewer error head-on as rapport faded into the background; their efforts were further fueled by new awareness of the benefits that standardization allowed, such as timeliness and statistical sophistication. The result was a massive attempt to completely wipe out interviewer error, which, due to the extremeness with which it was attempted, started to backfire. The proper conclusion, however, is not that standardized interviewing is a failure, but that it is time for a new compromise between its advocates and enemies. It is rather ironic that almost 50 years ago, Lazarsfeld saw the need for compromise on a similar debate: whether "various non-directive means of stimulating full discussion in the interviewing situation" (Skott, 1943) were preferable to "more objective methods of research" (Lazarsfeld, 1944). Granted, the specifics of the debates are not identical, but they cover much common ground; and in any case, it is definitely time for negotiation. Standardization will not go away, though it may allow interviewers some minor flexibility in clarifying communication problems; at the same time, standardization advocates must realize that interviewing is not the only source of survey error, and our efforts to reduce interviewing errors do have practical limits. Let us hope that within the next 50 years, all us who rely on surveys as research method can come to more of a consensus on this matter.

REFERENCES

"Computer Assisted Survey Information Collection" (CASIC). Working paper prepared by Statistical Policy Office (OMB), May, 1990.

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

Gordon, Raymond L. (1969). Interviewing: Strategy, Techniques, and Tactics. Homewood, IL: Dorsey. Cited in Goudy and Potter (see below).

Goudy, W.J. and H.R. Potter (1975). "Interviewer rapport: demise of a concept." Public Opinion Quarterly, 39: 530-543.

Groves, Robert M. (1989). Survey Errors and Survey Costs. New York: Wiley.

Guenzel, Pamela J., Tracy R. Berckmans, and Charles F. Cannell (1983). General Interviewing Techniques. Ann Arbor: The University of Michigan.

Hyman, Herbert (1954). Interviewing in Social Research. Chicago: University of Chicago Press.

Kahn, R. and C.F. Cannell (1957). The Dynamics of Interviewing. New York: Wiley.

Lazarsfeld, Paul F. (1944). "The controversy over detailed interviews-- an offer for negotiation." Public Opinion Quarterly, 8: 38-60.

Skott, Hans E. (1943). "Attitude research in the Department of Agriculture." Public Opinion Quarterly, 7: 280-292. Cited in Lazarsfeld, op. cit.

Suchman, Lucy and Brigitte Jordan (1990). "Interactional troubles in face-to-face survey interviews." Journal of the American Statistical Association, March 1990, Vol. 85.

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