Standardized Survey Interviewing

In a standardized survey interview, the interview proceeds according to a script that is intended to minimize any potential impact of individual interviewers' behavior on respondents' answers and the resulting data. Standardized interviewing procedures for sample surveys were developed over several decades of the twentieth century as evidence accrued that even seemingly minor differences in how interviewers behaved in interviews sometimes affected answers and data quality. Interviewers' biases or assumptions about particular types of respondents could creep into the interview through subtle changes in wording or tone that could lead respondents to interpret questions or the situation differently than they would have with a different interviewer. Even without intending to influence answers, interviewers who attempted to increase rapport by rephrasing a question the second time they asked it, or politely didn't present all the response alternatives to a question because they judged that some alternatives wouldn't fit a particular respondent's circumstances, could harm the quality of the data. It also became clear that interviewers could, in all innocence and with good intentions, introduce bias through how they reacted when a respondent expressed reservations or uncertainty about an answer; interviewers could subtly encourage respondents to give answers that fit the interviewers' preconceptions rather than the respondent's actual circumstances or opinions.

Standardized survey interviewing procedures are designed to circumvent these problems, and to ensure that the data from all respondents are fully comparable because all respondents have answered the same questions under the same procedures.

Standardizing the interviewing procedures is intended to address the measurement error due to interviewers, which is assumed to be independent of measurement error due to question wording (which can be addressed through better question pretesting) and measurement error due to respondents (which can't easily be addressed by survey researchers).

Ideally, interviewers adhering to standardized procedures read (either from paper or from a computer screen) survey questions and all response alternatives precisely as worded by the designers of the survey, and they repeat the full question and all response alternatives when asked to repeat the question. In the strictest forms of standardized survey interviewing, interviewers also leave the interpretation of questions entirely up to respondents, and only respond to any requests for clarification with neutral probes like "whatever it means to you" and "let me repeat the question." The logic is that if only some respondents receive clarification or help with answering, then the stimulus (the question wording and response alternatives) is different for different respondent, and thus there is no guarantee that the data are comparable.

The broad consensus is that standardized interviewing procedures are the most desirable for sample surveys, and that more idiosyncratic or ethnographic forms of interviewing that are useful for other research purposes are risky or undesirable in surveys. But this consensus can manifest itself somewhat differently in different survey organizations, where the precise procedures that count as standardized in one center can differ from those in other organizations. For example, organizations differ on whether providing clarification to a respondent counts as non-standardized or standardized, and on whether repeating only the most appropriate response alternatives is better than repeating

them all. Survey organizations can also vary in how extensively they train and monitor their interviewers for adherence to the standardized procedures they advocate, which means that in practice some standardized surveys turn out to be less standardized than others.

Controversies

Starting in the 1990's, methodological researchers who closely examine interview recordings began documenting that strictly standardized procedures sometimes can create uncomfortable interactions that not only frustrate respondents but lead to demonstrably poorer data quality. When interviewers shoehorn respondents into answers that do not reflect their circumstances, when they repeat information that the respondent already knows, when they ask for information the respondent has already provided, or when they refuse to clarify what their questions mean, respondents can become alienated and recalcitrant, and they can provide incorrect answers. When viewed through this lens, the practical results of strict standardization can, on occasion, run counter to the intended effects. Perhaps, in the attempt to standardize wording in the interview, survey researchers are failing to standardize what really ought to be standardized: the meaning of the questions. On this view, to serve the goal of making sure that respondents' answers are truly comparable, interviewers should instead work to make sure that respondents are interpreting questions in the same way—even if this means deviating from a script and tailoring clarification and probes to individual respondents.

The jury is still out on the implications of this work and what it will mean for the future of standardized interviewing. There are a number of important considerations. First, much is unknown about how often problematic interactions occur and how often respondents' interpretations of questions differ from survey designers' in different domains of questioning; if problems and misinterpretations are too frequent, then alternate ways of implementing standardized interviews are worth investigating, but if they are quite rare then revamping procedures on a large scale is not clearly worthwhile. Second, the survey enterprise currently relies on a particular model of hiring and training interviewers, and it is not yet clear how training and monitoring interviewers who implement a less scripted version of standardization would work. Based on some reports, some interviewers would prefer to implement less strictly standardized interviews, but whether a different kind of interviewer (for example, with higher levels of education or subject matter expertise), and different levels of compensation, would be needed to do this is unknown.

The larger question about the best way to achieve the goals of standardization remains: should survey researchers standardize the stimulus (question) and the subsequent interaction? Or should survey researchers standardize the respondent's *interpretation* of the question and *experience* of the interaction?

This question is put into sharp relief by comparing standardized interviewing with self-administered surveys. In a sense, a self-administered survey (whether administered via paper and pencil, via clicking on a textual web survey, or via listening to a recorded voice on a laptop computer) is the ultimate in standardization: all respondents are presented with precisely the same stimulus, with no variation in what the "interviewer"

(words on the page, audio recording) does. One could argue that the future of standardized interviewing is to migrate human interviews into automated self-administered interviews, because, in essence, this creates one interviewing agent for the entire survey sample, and so the differential impact of different interviewers disappears. (The less quantified aspects of human contact and rapport-building would also, of course, disappear).

But the larger question about what it takes to make survey responses comparable and to minimize the effects of the interviewer/interviewing agent would still not be addressed, even with full self-administration. For example, in a paper and pencil survey, the interpretation of the questions is left entirely up to the respondent; but should it be? Would the data be more comparable if respondents were all assisted in interpreting the questions in the same way (e.g., via clarification dialogue in a web survey)? In an audio self-administered survey, will different respondents interpret the recorded interviewer's vocal tone differently in ways that affect their answers? Would the data be more comparable if the interviewer's voice were variably tailored so as to be experienced as more similar across respondents?

Theories and practices of standardization—and thus questions about standardization—are thus expanding rapidly with new interviewing technologies. Debates about and data from studies of human interviews are likely not only to inform those new technologies, but the careful manipulation of features of human interaction that will come from new technologies will inform ongoing debates about standardization in human interviewing.

Michael F. Schober

See also Measurement error, interviewer-related error, (I would need to see the full list of headwords to be able to do this correctly, but they don't seem to be available on the web site)

Further Readings

Beatty, P. (1995). Understanding the standardized/non-standardized interviewing controversy. *Journal of Official Statistics, 11*, 147-160.

Conrad, F.G., & Schober, M.F. (2000). Clarifying question meaning in a household telephone survey. *Public Opinion Quarterly, 64*, 1-28.

Fowler, F.J., & Mangione, T.W. (1990). *Standardized survey interviewing: Minimizing interviewer-related error.* Newbury Park, CA: SAGE Publications, Inc.

Houtkoop-Steenstra, H. (2000). *Interaction and the standardized survey interview: The living questionnaire.* Cambridge, UK: Cambridge University Press.

Maynard, D., Houtkoop-Steenstra, H., Schaeffer, N.C., & van der Zouwen, J. (Eds.) (2002). *Standardization and tacit knowledge: Interaction and practice in the survey interview*. New York: Wiley.

Suchman, L., & Jordan, B. (1990.) Interactional troubles in face-to-face survey interviews. *Journal of the American Statistical Association, 85*, 232-253