CHAPTER 8

New Frontiers in Standardized Survey Interviewing

Frederick G. Conrad
Michael F. Schobert

Sample surveys are the empirical backbone of the social sciences, government policy, political campaigns, and corporate strategy. The method of collecting data that is generally assumed to provide the highest quality information is the standardized interview, either on the telephone or face-to-face. There are exceptions, of course, such as the collection of data about sensitive topics like drug use and sexual activities, for which self-administration (i.e., without an interviewer) is believed to increase the honesty of respondents' answers. However, by and large, if social researchers can afford to conduct interviews, they do so. Interviews generally lead to higher response rates than self-administration (particularly when conducted face-to-face) and allow respondents who cannot read (including those who are visually impaired) to participate.

Standardized interviewing is an approach to collecting survey data in which interviewers read questions exactly as worded to every respondent and are trained never to provide information beyond what is scripted in the questionnaire. The goal is to increase comparability across interviews so that different answers from different respondents cannot be attributed to different question stimuli. In principle, standardized interviewers are interchangeable; the answers should be the same no matter who asks the question.

Uniform question presentation has not always been the industry standard (see Beatty, 1998, for a historical perspective on standardized interviewing). However, since the widespread deployment of telephone interviewing in the 1970s, standardization has been virtually synonymous with scientific social research (see Fowler & Mangione, 1990, for a clear statement of standardized interviewing practice and goals). Nonetheless, standardized interviewing has been criticized because the limits it places on what in-
Theories of Communication and Survey Interviews

Most large-scale surveys try to standardize what interviewers say. As described by Fowler and Mangione (1990), interviewers must read questions exactly as written, they must probe neutrally, and they must never allow their own ideas to influence the respondents' answers. According to Fowler and Mangione, this practice reduces or eliminates what they call interviewer-induced error—systematic effects of particular interviewers on survey responses.

This view grows out of a long tradition in survey methodology of distinguishing between different sources of error and using different methods for reducing each type of error. For example, this view holds that respondent comprehension error is best handled by wording the question in ways that most respondents are likely to understand. By presenting early versions of questions and revising those questions on the basis of the pilot results, misunderstanding can be largely reduced. Standardized interviewers deliver pretested questions in a standardized way—expecting respondents to the same question stimulus (Fowler & Mangione, 1990, p. 14)—then researchers can be confident that differences in the answers stem only from the actual differences between respondents and not from any misunderstanding by some respondents or idiocies and quirks of interviewers' behavior (p. 15).

However, by separating interviewer behavior, respondent behavior, and question wording, this prevailing approach relies on a view of communication that has been discredited, at least for ordinary spontaneous conversations. This view, dating back to John Locke or perhaps even earlier, has been called, variously, the model (Akmajian, Demers, Farmer, & Harwin, 1990), the conduit metaphor (Reddy, 1972/1995), and the meaning-in-world assumption (Schober, 1998). According to this view, speakers encode their thoughts into linguistic messages and send these messages to recipients by speaking, who decode them into their own thoughts. Thought and conceptual material are thus transferred from one head to the other via words.

Although the message model captures most people's intuitions about how communication works, it cannot account for all of what goes on in ordinary conversations (see also Clark, 1992, 1996; Gibbs, 1994; Maynard & Whalen, 1995). The problem is that the message model assumes that the meaning of speakers' words is in the words themselves. But it is not. Rather, speakers can use the same words to express vastly different meanings on different occasions (see Akmajian et al., 1990). Consider the word red (Clark, 1996). It denotes one color when referring to wine, another when referring to hair, another when describing a fire truck, and so on. So even relatively simple words, such as interviewer-agents and respondents (p. 14)—then researchers can be confident that differences in the answers stem only from the actual differences between respondents and not from any misunderstanding by some respondents or idiocies and quirks of interviewers' behavior (p. 15).

However, by separating interviewer behavior, respondent behavior, and question wording, this prevailing approach relies on a view of communication that has been discredited, at least for ordinary spontaneous conversations. This view, dating back to John Locke or perhaps even earlier, has been called, variously, the model (Akmajian, Demers, Farmer, & Harwin, 1990), the conduit metaphor (Reddy, 1972/1995), and the meaning-in-world assumption (Schober, 1998). According to this view, speakers encode their thoughts into linguistic messages and send these messages to recipients by speaking, who decode them into their own thoughts. Thought and conceptual material are thus transferred from one head to the other via words.

Although the message model captures most people's intuitions about how communication works, it cannot account for all of what goes on in ordinary conversations (see also Clark, 1992, 1996; Gibbs, 1994; Maynard & Whalen, 1995). The problem is that the message model assumes that the meaning of speakers' words is in the words themselves. But it is not. Rather, speakers can use the same words to express vastly different meanings on different occasions (see Akmajian et al., 1990). Consider the word red (Clark, 1996). It denotes one color when referring to wine, another when referring to hair, another when describing a fire truck, and so on. So even relatively simple words, such as red, do not have one unchanging meaning that everyone knows. Rather, the meaning of red depends on the context in which it is used.
standing or misunderstanding, and they adjust their utterances, content by moment, to ensure that their addressers understand them well enough for current purposes. Addressers, by providing such evidence, help mold the utterances speakers produce.

In this view, no utterance is complete until it has been ground—until both participants have accepted that it has been understood. Understanding a reference in any particular utterance requires active participation by both speaker and addressee, and this can take several turns. Note that the point is not that words do not have conventional meanings; in fact, the conventional meanings of words provide important constraints on speakers’ meanings. But speakers regularly use words in idiomatic ways that go far beyond dictionary definitions (see Clark, 1991; Clark & Gerrig, 1983). Speakers produce utterances based on their common ground with their conversational partners—that is, what they presume that they and their conversational partners mutually know, believe, and assume.

**Experimenting with Alternatives to Standardized Interviewing**

What might the collaborative view of language use imply for survey interviewing? As we indicated earlier, standardized interviewers are prohibited from grounding the meaning of questions (although the rationale is not expressed in these collaborative terms). If a respondent asks a standardized interviewer to clarify a question (e.g., “What do you mean by ‘household’?”) the interviewer can reply only by administering one of a small set of “neutral” or nondirective probes, such as restating the question or indicating that the interpretation of the question is the sole obligation of the respondent (e.g., “Whatever it means to you.”) By withholding substantive clarification (as in restating the question) and even encouraging respondents to interpret the question in their own way, the interviewer relinquishes control over what the question means to the respondent. The obvious cost of this practice is that respondents may not understand the questions the way the survey researchers intended and thus answer them inaccurately, jeopardizing the validity of the research.

We have conducted a series of experiments in which we have evaluated response accuracy and interaction patterns for strictly standardized and more collaborative “conversational” interviews, as well as for a range of interviewing techniques in between. In these studies, professional interviewers telephoned naive respondents either in the laboratory (Schober & Conrad, 1997; Schober et al., 2004) or at home (Conrad & Schober, 2000 and asked questions from large U.S. government surveys. In the laboratory studies, the “respondents” answered the questions on the basis of scenarios that described the work, housing, and purchases of fictional people. Because we created the scenarios, we knew the correct answers—according to official definitions—for each question-scenario combination, and so we could determine response accuracy. In the household study, the respondents answered questions about their own lives; we evaluated the response accuracy by less direct re-interview methods.

In each study, respondents participated in either strictly standardized interviews, following Fowler and Maugonce’s (1990) prescriptions, or less standardized, more collaborative interviews. In all interviews, respondents were first posed exactly the same questions, worded, underlining the shared commitment to standardization under the two approaches; the difference was whether it was wording or underlying meaning that was standardized. In the more collaborative interviews, interviewers were then encouraged to ground understanding of question meaning, for example, by providing scripted definitions when respondents explicitly asked for them (Schober et al., 2004) or by using whatever words interviewers chose to make sure respondents understood the questions as intended (Schober & Conrad, 1997).

In all studies our basic question is which interviewing technique leads to more accurate responses. Of course, even in purely standardized surveys, interviewers can affect responses. What we have tested here is how the kinds of influences that occur in more collaboratively standardized interviews affect response accuracy as compared with the kinds of influences that occur in more collaborative interviews.

**Mapping Ambiguity**

In all studies we used pretested questions from ongoing surveys whose words and grammar have been shown to be understandable but whose interpretation for some respondents on some occasions might be unclear. Consider a question such as “During the past year, have you purchased or had expenses for household furniture?” A respondent who has bought an end table should not have much trouble answering “yes,” but a respondent who has bought a floor lamp may be less sure. Or consider a question such as “Last week, how many hours did you work?” This should be clear for a respondent who has a 9-to-5 job that includes overtime but less clear for a respondent who does business over lunch or solves work-related problems while jogging. We presume that the second respondent would be more likely than the first to request clarification such as, “What do you mean by work?”, or “Should I consider business lunches to be work?” We refer to this as a mapping ambiguity; the mapping between the question term and the respondent’s circumstances is unclear. Interviewers following strictly standardized procedures cannot meaningfully resolve a mapping ambiguity; they would be obliged to use neutral probing techniques, including “whatever it means to you.” More conversationally flexible interviewers could provide information to the respondents that would help clarify what the author of the question had in mind.

In our laboratory studies, we designed the fictional scenarios as half of our respondents received the questions in a straightforward way (straightforward mappings) and the other half corresponded to the questions in a more complicated way (complicated mappings). In the household study, we had no control over the frequency of complicated mappings. Our prediction was that response accuracy would be high for both standardized and more conversational interviewing when the mappings were straightforward, when the mappings were complicated, accuracy should suffer for strictly standardized interviewing but not as much—maybe not at all—for conversational interviewing. This pattern of results in the household study, in which mappings were not under our control, would indicate that complicated mappings are frequent enough in the real world to warrant further exploration of more collaboratively interview techniques.

In order to ensure that conversational interviewers could answer respondents’ substantive questions, we needed to teach them the official definitions of key concepts in the questions. Providing standardized interviewers with this knowledge might seem to violate the principles of standardization. But the only role for definitions in the standardized interview is to be read in their entirety to all respondents on all occasions or not read at all. But the logic of our experiments required us to train all interviewers together on the concepts so that any nonstandard responses could not somehow be attributed to different levels of knowledge between standardized and conversational interviewers. Standardized interviewers were told that the concept training was necessary so that interviewers would be able to judge when respondents had answered a question completely (see Beatty, 1995). In all studies the training lasted about 10 minutes; interviewers first studied the official definitions and then actively carried out exercises to ensure that they had grasped the concepts in detail.
viewing techniques. The standardized instructions were based on guidelines that appeared in an interviewing manual for a survey on which many of these interviewers regularly worked and were consistent with Fowler and Massey's (1985) approach. Using this material, we reviewed standardized question-asking and neutral probing techniques and supplemented this with role-playing exercises.

The interviewers who were trained to use more conversational techniques were instructed to initially read the question as worded. Then (depending on the study) they could subconsciously answer respondents' request for clarification, either following a script or in their own words; they could also provide unsolicited clarification (scripted or in their own words) when respondents seemed to need it, even if respondents hadn't asked for help. In another study (Schroeder et al., 2004, Experiment 2), interviewers were not trained in a particular technique but were told to do whatever they ordinarily do.

Procedure

In the laboratory studies, respondents were given a packet of scenarios to study, and then they were questioned over the telephone about the scenarios. The experimenters in conversational interviews were instructed to work with the interviewers to make sure they had interpreted the questions as the survey designers intended; they were encouraged to ask for clarification if they needed it. Response accuracy was measured as the percentage of questions for which responses matched what the official definition indicated was correct.

Although the interviewers knew that respondents were answering on the basis of fictional scenarios, they were not familiar with the contents of the individual scenarios, and so knowledge was allocated much as it is in real surveys: Interviewers knew the questions, and respondents knew about their own circumstances. We counterbalanced the issue so that the respondents who were assigned to a particular interview received different versions of the scenario. This way the interviewers could not become familiar with the scenarios based on anything the earlier respondents might have said.

In the household study respondents were telephoned at home and asked about their own lives; no scenarios were involved. Because we could not directly determine the accuracy of the respondents' answers, we designed the experiment to provide two indirect measures. One measure was response change between interviews. All respondents participated in two interviews. The first was strictly standardized for all respondents, and the second was strictly standardized for half of the respondents and conversational for the other half. If respondents' circumstances mapped in a complicated way to the question concepts, they should be more likely to change their answers between an initial standardized interview and a subsequent conversational interview than between two standardized interviews. The reason is that in the conversational interviews the interviewers were instructed to clarify question meaning and correct respondent misconceptions, which could lead to different answers than provided in the initial interview. In contrast, in the second standardized interview, the interviewers were not permitted to clarify meaning or to correct initial respondent misconceptions; therefore, we hypothesized responses would likely remain unchanged—reliable but incorrect.

The other measure in the household survey was the "legality" of respondents' explanations for their responses. If respondents answered "yes" when asked if they had made certain types of purchases, they were asked to briefly describe the purchase(s). These descriptions were then coded for their consistency with official definitions—legality. For example, the definition of moving expenses explicitly excludes payments for do-it-yourself moving; a respondent who answered "yes" when asked if he or she incurred moving expenses and based this response on having rented a moving van would have provided an illegal explanation. When interviewers were licensed to clarify question meaning (in a conversational interview), we expected respondents to be more likely to base their responses on legal purchases than when interviewers were not able to clarify question meaning (in any of the standardized interviews).

Results

We first analyzed the transcripts of the interviews to verify that interviewers had followed our instructions and implemented the appropriate technique for each specific study. One way we demonstrated that interviewers followed instructions was by coding the various deviations from standardized questioning and comparing their frequency between interview techniques. These deviations included rephrasing all or part of the question, providing all or part of a definition (either verbally or paraphrased), converting the respondents' descriptions of their circumstances (the fictional scenario in the laboratory studies) into an answer, offering to provide clarification, confirming or disconfirming the respondent's interpretation of the question, and requesting particular information about the respondent's circumstances. For example, in the following exchange (from Schroeder & Conrad, 1997), the conversational interviewer paraphrased the long government definition of "household furniture" to answer the respondent's question:

1: How Kelly purchased or had expenses for household furniture?
2: Um, . . . , a lamp furniture,
3: No sir, we do not include lamps and lighting fixtures.
4: Okay, no.
5: (goes on to next question)

In strictly standardized interviewing, the interviewer should not have answered the respondent's request for clarification, because by doing so she helped interpret the survey question for this respondent but may not have done so for another. Across the various studies, our coding gave us confidence that our interviewer training led to fundamentally different types of interaction. In the Schroeder and Conrad (1997) study, for example, 80% of the question-answer sequences in conversational interviews contained deviations from standardization, compared with only 2% in strictly standardized interviews.

We can now turn to response accuracy. Again, an accurate response in our experiments is one that is consistent with the official definition of the relevant concept. In our earlier lamp example, the correct answer is "no," because a lamp purchase does not qualify as a furniture purchase for the purposes of the survey from which this question was drawn; therefore, the respondent's answer was accurate.

Across all our lab studies (Schroeder & Conrad, 1997; Schroeder et al., 2004, Experiments 1 and 2), a general pattern emerges. When mappings between question concepts and people's circumstances (scenarios) are straightforward, all interviewing techniques lead to nearly perfect accuracy; virtually all respondents interpret question concepts in the ways that surveys designates intended. But when mappings between question concepts and people's circumstances are complicated, strictly standardized interviewing leads to quite poor response accuracy (89%). Response accuracy improves when interviewers can provide clarification on request (77%), and it improves substantially more when interviewers can both offer clarification that they believe respondents need as well as provide it when respondents request it (77%). Response accuracy is highest when the interviewers who can both volunteer clarification and provide it on request can also use their own words to clarify question concepts, even if this departs from what is scripted (87%). This last group of interviewers can exercise much of the discretion and flexibility that is typical of everyday conversation.
These results are mirrored in the household study (Conrad & Schober, 2000), in which we used indirect measures of response accuracy. Respondents changed their answers more often when their second interview was conversational (22%) than when it was strictly standardized (11%). In addition, more respondents were based on legal purchases when the second interview was conversational (95%) than when it was standardized (57%). For example, in conversational interviews respondents who answered “yes” to the question about moving expenses were more likely to do so because they had hired commercial movers or had had other expenses included in the definition than in standardized interviews, in which they were more likely to incorrectly respond “yes” because of excluded expenses such as do-it-yourself moving. This result was primarily due to interviewers’ explaining what to count and what not to count, that is, grounding the meaning of the term. The improved response accuracy in conversational interviews suggests that respondents’ actual circumstances (as opposed to the fictional scenarios presented in the lab studies) are complicated sufficiently, often at least for these questions; to justify employing the technique further.

Fowler and Mangione (1990) have raised the concern that interviewers whose wording is not strictly standardized will damage the quality of the responses. They are particularly concerned that in questions about opinions, nonstandardized interviewers may bias respondents by presenting their own opinions or by reacting to the respondents’ answers. Regarding interviewers’ explaining the intent of questions, Fowler and Mangione are concerned that nonstandardized interviewers may mislead respondents by providing inaccurate information. In our experiments this has not been the case. For example, in the Schober and Conrad (1997) study, the information provided by conversational interviewers to clarify the question was accurate (complied with the official definitions) 99% of the cases in which it was given. This 99% consisted of 87% in which respondents received accurate information from interviewers and provided accurate answers, and only 6% in which respondents received accurate information but provided inaccurate answers. For the 7% of cases in which interviewers provided inaccurate information, respondents were still accurate about half the time. The 7% consisted of 6% in which respondents received inaccurate information from interviewers yet provided accurate answers, and only 1% in which they received inaccurate information and answered the question incorrectly. Conversational interviewers provided highly accurate information. When they did provide inaccurate information, it did not necessarily lead respondents to produce incorrect answers; in fact, respondents produced incorrect answers resulting from inaccurate information only 5% of the time.

Close analysis of the interviewer–respondent interaction (see Conrad & Schober, 2000; Schober & Conrad, 1997) shows that it is really not very different from standardization that led to the increases in response accuracy. It would seem that interviewers’ correction improved response accuracy because respondents had requested clarification or not (that is, even when interviewers provided the information, either to the respondents having asked for it. For example, in the Schober and Conrad (1997) study, for the 96 complicated mapping cases in which interviewers provided additional help, respondents produced 75 accurate answers, an accuracy rate of 80%. In contrast, for the 116 uncomplicated mapping cases in which interviewers did not provide any help, respondents produced 55 accurate answers, an accuracy rate of 63%. This is close to the 20% accuracy rate for complicated mapping in standardized interviews and suggests that when conversational interviewers do not provide clarification but behave like their standardized counterparts, response accuracy will suffer.

Across our studies, an improvement in response accuracy came at a significant cost.

Conversational interviews took much longer than standardized interviews (from 80 to 200% longer, in the various studies), this was true for both straightforward and complicated mappings. Apparently, clarification just takes time. In Schober and colleagues (2004), experiment 1, the correlation between interview time and response accuracy across the five types of interviewing was .98. The more flexibility given to interviewers, the more accurately respondents answered, and the longer the interviews last. Practitioners will need to know how reliable their results are; the case that respondents understand all questions as intended. If they are willing to live with some uncertainty, then it may be possible to shorten interviews while maintaining levels of response accuracy above what we observed for strictly standardized interviews.

The results of our studies should not be taken as the final word on the issue, nor should they be taken as showing definitively that conversational interviewing is always a good idea. Our studies have examined nonsensitive fact-based questions, and the results may not generalize to questions about sensitive topics or opinions in a straightforward way. Just like fact-based questions, opinion questions contain phrases with alternative possible interpretations—consideration abortion or adoption, and thus opinion surveys may benefit from more collaborative approaches to interviewing. Whether this can be done without influencing the opinions is unclear, especially because response accuracy for opinions cannot be validated as directly as it can for fact-based questions in our studies. Nonetheless, O’Hara and Schober (2004) present evidence that differences in attitudes toward euthanasia are related to how respondents define the concept. One implication of this result is that presenting uniform wording does not guarantee uniform interpretation of the “attitude object,” and so attitude researchers may want to standardize the attitude object by defining it as part of the question.
ists turned out to be the wrong question-naire path. The Saxe et al. (2009) study was conducted in the laboratory with a convenience sample; nonetheless, the findings are consistent with other evidence (e.g., Belbin, 1986) that suggests that ordinary words in survey questions are interpreted in numerous ways. Overall, our findings suggest that if complicated mappings are known to be rare, then strictly standardized techniques could lead to accurate responses at lower costs than collaborative techniques. If the complicated mappings are known to be frequent, or if (more realistically) their frequency is unknown, more collaborative techniques might be worth the increased costs they use would, no doubt, entail.

Dialogue Features in Automated Interviewing Systems

When respondents self-administer a paper-and-pencil survey or type and click answers on a Web survey, they are usually concerned of doing something quite different than they are when they answer questions asked by a human interviewer. With self-administered questionnaires, respondents typically read rather than hear questions; they control the pacing of the interaction, and they write rather than speak their answers. Also, there is no interviewer present who might react to the respondent’s answers (and potentially judge the respondent on the basis of those answers). The combination of such mode differences is no doubt what leads respondents to willingly report more socially undesirable behaviors (e.g., drug use and taboo sexual practices) in self-administered computer-based “interviews” than in actual interviews conducted by human interviewers (e.g., Tourangeau & Smith, 1996).

We propose, in contrast to the usual view, that new technologies used for presenting self-administered survey interviews fall on a continuum of anthropomorphism, with different kinds of self-administration incorporating different features of human interviewing (see Conrad & Schober, 2006). For example, sexual behavior steps close to human administration when it is supplemented with audio files of computer-generated voices asking questions; computerized data collection becomes more like a human-administered interview when the text is replaced entirely with recorded human voices (see Cooper, Singer, & Tourangeau, 2004). It becomes even more human-like when the interviewing system provides human-like prompts, feedback, and clarification. A self-administered questionnaire with a drawing of a human interviewer is less anthropomorphic than one with an animated, virtual interviewer whose lips move in synchrony with its speech and whose eyes blink; if the agent nods and smiles in response to answers, the system becomes yet more human-like.

One conceives of interviewing as involving interaction between the respondent and an interviewing agent that is either a human or a computer program, ranging from fully conversational to robotically standardized. We can begin testing which features of interview interaction lead to high-quality data and respondent satisfaction (Conrad, Schober, & Coerr, 2007). We can also begin to better understand the nature of interviewing by decomposing interviewing behaviors into their separable parts and beginning to test what happens when we add features of human interviewing to self-administered interviews. For example, interviewers, like most conversationalists, probably make certain assumptions about the respondent’s abilities as a conversationalist. They probably judge whether a particular respondent is more or less likely to need help in understanding questions and may interpret a respondent’s behavior for a particular question in light of such judgments. We refer to such judgments as “respondent models.” Such models may not be useful to all interviewers, but to those who are empowered to provide clarification, said in order to create the perception of automation. We developed versions of both systems that reacted differently depending on the respondent model.

When the interviewing system (whether text or speech-based) could voluntarily clarify, it interpreted respondent behavior (e.g., inactivity or silence) based either on a generic or a group-based (or "stereotypical") respondent model. Under the generic model, the system treated the behavior of all respondents identically. Under the group-based model, a particular behavior (e.g., a pause of 2 seconds) was interpreted differently for respondents in different groups. The respondent attribute that we intended to model was added quickness; a particular interval of no respondent activity or speech may signal difficulty for a quick respondent but ordinary thinking for a slower respondent. Thus help should be most useful after shorter lags for quick respondents and longer lags for slower respondents. Rather than assigning respondents to groups on the basis of their quickness, we used their age as a surrogate for quickness based on the well-known impact of cognitive aging on response time (e.g., Salthouse, 1976, 1982).

![Figure 8.1](image-url)
Therefore, under the group-based model, the system offered clarification sooner to younger respondents than to the older respondents.

With both the text-based and speech-based systems, respondents answered questions very similar to those used in the laboratory interview studies discussed earlier (Schuber & Coonrod, 1997; Schuber et al., 2004), using either complicated or straightforward scenarios. As in the interview studies, our focus was on response accuracy for complicated scenarios. When group-based respondent models were used in either system, they improved response accuracy relative to generic response models, presumably because the clarification was better tailored to respondents' needs (see Figures 8.2 and 8.3). Irrespective of the model, enabling the system to both volunteer clarification and respond to explicit respondent requests increased response accuracy relative to cases in which respondents could obtain clarification only by explicitly requesting it. Any clarification the system provided led to greater accuracy than no clarification, which approximates what happens in standardized interviews, as well as in most current Web surveys.

Just as longer-than-normal response times may signal difficulty of some kind, quicker-than-normal response times may also reflect suboptimal performance. If a respondent answers immediately, it is unlikely that he or she has given as much thought to the task as the researchers would like. If this is the case, it may be possible to create more accurate respondent models by designing them so that the system offers help when response times are outside an optimal range, either too slow or too fast. We have dubbed this the "Goldilocks range" and used it to model respondent speech (Edens et al., 2007). As can be seen in Figure 8.3 (generic and stereotyped respondent models), response accuracy was better when the system could volunteer clarification based on either respondent silence or overly quick answers than when it could not provide clarification.

These findings suggest that some of what conversational interviews do can be implemented in self-administered surveys. But there are a number of caveats and questions that need to be addressed. For example, despite the apparent benefits of group-based models for response accuracy, some respondents found the system intervention unpleasant. For example, in postexperimental satisfaction questionnaires, a substantial portion of the older respondents indicated that they would prefer a human interviewer to the system that embodied a group-based respondent model, whereas they registered a preference for the computer-based system when it was built around a generic model or provided clarification only when respondents requested it by clicking. It could be then that modeling in this study is not as accurate as it might be and that better estimates of when respondents are having trouble might improve their satisfaction with the system. But the dissatisfaction of some respondents with this kind of system could also reflect a belief that respondents, not computers, should be in control of the interaction.

Despite the mixed impact on respondents' satisfaction, when interviewing systems offer clarification, they do so more effectively when they are based on respondent models. If interviewers and interviewing systems develop and use models of respondents, it seems likely that respondents similarly make certain assumptions about the abilities of the interviewer or interviewing system. Certainly an interface that provides clarification in conversationally savvy ways, such as adjusting its interpretation of respondent cues based on respondent characteristics, is likely to seem more human-like, or more animate. But what about embedding such animated characteristics in a virtual interviewing agent—that is, a computer-generated, graphical rendering of an interviewer—displayed in the user interface? This is sure to seem even more human. Due to the fact that animated agent technology is increasingly available, we believe it is a matter of time until survey researchers use it for collecting data. We have begun a series of experiments to explore situations in which animated agents as interviewers might help or
hurt research. In order to study the impact of standardized interviewing agents that are more conversationally skilled than today's most advanced agents, we developed our agents by capturing the motions of human interviewers and using them to control the movement of graphical computer models. This work is currently ongoing, but our experience to date suggests that the methodological concerns are quite different depending on the kinds of questions the interviewer asks. For example, when interviewing agents ask sensitive questions, do respondents assume that the agent can pass judgment on their answers as a human interviewer might? If so, they may well provide answers that are "adjusted" to be more socially desirable, as they do with human interviewers; if not, their answers will be as relatively free of social distortion as when they respond to questions that are unambiguously presented by a computer (e.g., Tourangeau & Smith, 1996). When interviewing agents ask nonsensitive, factual questions, the concerns are more focused on whether respondents believe that the interviewing agent can detect cues of response difficulty, for example, speech fluctuations and gaze aversion. If so, respondents may produce more of these cues as we have observed in human interviews (Conrad, Scherer, & Dijkstra, 2008; Scherer & Bloom, 2004). Agents that are more facially realistic may encourage respondents to provide more information. Our agents are not think how differently an interviewer might be with different respondents. One does not think that this kind of standardization or encouragement an interviewing agent might provide to different respondents is ambiguous. For example, does it improve response accuracy, particularly when the correspondence between respondents' circumstances and question wording is ambiguous. We held evidence that allowing interviewers to choose their wording in order to make sure that respondents interpret the questions the same way and as intended can dramatically improve response accuracy, particularly when the correspondence between respondents' circumstances and question wording is ambiguous. Our argument, in essence, is that standardizing interviewers' observable behavior may influence similarity to the interviewing agent for some respondents and dissimilarity for others. The notion of comparability in standardized interviewing, we argue, deserves rethinking even in human interviews. Some characteristics of interviewers, such as pitch range and race and gender, are immutable, but others are not: think how differently an interviewer might be with different respondents. One does not think that this kind of standardization or encouragement an interviewing agent might provide to different respondents is ambiguous. For example, does it improve response accuracy, particularly when the correspondence between respondents' circumstances and question wording is ambiguous. We held evidence that allowing interviewers to choose their wording in order to make sure that respondents interpret the questions the same way and as intended can dramatically improve response accuracy, particularly when the correspondence between respondents' circumstances and question wording is ambiguous. Our argument, in essence, is that standardizing interviewers' observable behavior may influence similarity to the interviewing agent for some respondents and dissimilarity for others. The notion of comparability in standardized interviewing, we argue, deserves rethinking even in human interviews. Some characteristics of interviewers, such as pitch range and race and gender, are immutable, but others are not: think how differently an interviewer might be with different respondents. One does not think that this kind of standardization or encouragement an interviewing agent might provide to different respondents is ambiguous. For example, does it improve response accuracy, particularly when the correspondence between respondents' circumstances and question wording is ambiguous. Our argument, in essence, is that standardizing interviewers' observable behavior may influence similarity to the interviewing agent for some respondents and dissimilarity for others. The notion of comparability in standardized interviewing, we argue, deserves rethinking even in human interviews.
CHAPTER 9
Emergent Approaches to Focus Group Research

David Morgan
Collin Follansbee
Heather Guevara

A recent international conference on qualitative research, one of the panels began with an apology for the amount of time she needed to explain her methods, because "they aren't something traditional, like ethnography or focus groups." Yet only 20 years ago, the rare presentation that used focus groups started with a similar apology for and explanation of what was then a little-known method. Now, in the two decades since focus groups (re)emerged in the social sciences (Krueger, 1988, Morgan, 1987; Morgan & Saunders, 1984), their popularity has soared to the point that a Web of Science database search for articles from 2006 that used the term “focus groups” found more than 1,000 articles in that year. Hence this chapter concentrates on newer or more innovative approaches to what is now a well-recognized method. It is also worth noting, however, that the history of focus groups illustrates how what was once a little-known, “emergent” method can become a widely used and even “traditional” research technique within a relatively short period of time.

Our overall approach in this chapter is to present these emergent alternatives to traditional ways of doing focus groups as sets of research designs. In particular, our goal is to make researchers aware not only of the range of options that are available for doing focus groups but also of how to evaluate those options for different purposes. This approach matches an overall position that there is no “right way to do focus groups”—instead, there are many alternatives, and it is up to the researcher to select a set of options that are appropriate for a given project (Morgan, 1990). This amounts to a “pragmatic approach” to research design (Morgan, 2007, 2008), in which a clear understanding of the goals and outcomes for the project should determine decisions about the research design. In other words, understanding why the research is being done is essential to answering questions about how to do the research.
Handbook of Emergent Methods

EDITED BY
Sharlene Nagy Hesse-Biber
Patricia Leavy

THE GUILFORD PRESS
New York  London