VISUAL AND VERBAL CUES OF SURVEY RESPONDENTS' NEED FOR CLARIFICATION [1]

Frederick G. Conrad, University of Michigan
Michael F. Schober, New School for Social Research
Wil Dijkstra, Free University of Amsterdam
Frederick G. Conrad, University of Michigan
426 Thompson Street, Ann Arbor, MI 48104
fconrad@isr.umich.edu

Key Words: Comprehension of survey questions, visual and verbal cues of comprehension problems, mode effects, face-to-face and telephone interviews, conversational interviewing, clarification.

INTRODUCTION

Is there any reason that the quality of survey responses should differ between telephone and face-to-face (FTF) interviews? Large scale comparisons between these modes (e.g. DeLeeuw & van der Zouwen, 1988) have found few if any differences. Historically, the lack of difference between the modes has been celebrated as evidence that telephone interviewing has come of age.

We propose that if mode effects on data quality exist they are likely to depend on what interviewers can and cannot do in response to respondents' visual cues of comprehension difficulty. Different interviewing techniques provide interviewers with different tools. Standardized interviewers (see e.g. Fowler & Mangione's [1990] prescriptions) are required to administer non-directive probes like "let me repeat the question," when respondents ask for clarification. They are expressly forbidden from clarifying words in question because to do so for some respondents would mean that not all respondents would receive the same stimulus. Although Fowler & Mangione (1990) do not explicitly mention respondents' visual cues of confusion, the spirit of their technique would prohibit interviewers from responding to visual cues of confusion in FTF situations. Thus it is may not be surprising that mode effects have not been observed in standardized interviews.

However, interviewers do not always adhere to strictly standardized practice. In some cases organizations that purport to conduct standardized interviews actually train or allow interviewers to individualize their practice (Viterna & Maynard, 2002). Contrary to what is assumed by advocates of strict standardization, when interviewers spontaneously depart from scripted data collection, it can improve the quality of responses (Dykema, Lepkowski & Blixt, 1997; Schober, Conrad & Fricker, 2004). If the non-standardized interviewer behavior that improves response accuracy could be codified, it could lead to an alternative approach to interviewing.

Schober, Conrad and Fricker (2004) found that interviewers' departures from the script consisted largely of clarifying concepts mentioned in the questions by providing all or part of the official definitions. Imagine a respondent who has recently purchased a floor lamp and is asked whether they have recently purchased any household furniture. Without a definition, it is hard to know how to answer this question, but if interviewers can define household furniture – e.g. tell the respondent that in this survey lamps do not count as furniture – the potential misconception can be resolved. We have experimented with this approach (Conrad & Schober, 2000; Schober & Conrad, 1997; Schober, Conrad & Fricker, 2004) in telephone interviews and found that it can improve response accuracy dramatically. We have dubbed it "Conversational interviewing" because it exploits the ordinary conversational technique of talking about what has been said until both speaker and listener are sure they understand each other. Schober, Conrad and Fricker (2004) found the technique to be most effective when interviewers can provide clarification both in response to explicit requests (e.g. "Does a lamp count as furniture?") as well as when respondents seem to need it (e.g. they pause before speaking or their speech is disfluent). Conversational interviewing could potentially be even more effective in FTF interviews, if interviewers take advantage of respondents' visual cues that they need clarification.

Given the range of clarification behaviors that interviewers currently do and potentially might engage in, it is important to understand more about how interviewers react to visual cues of respondents' need for clarification. One could imagine that if interviewers react to visual information about the respondents' understanding this could lead to better comprehension, and thus response accuracy, in FTF than telephone interviews. This would produce a mode effect. An alternative possibility is that the mode of conversational interviews will not affect accuracy

[1]We thank Jan Smit, Gerhard Van de Bunt, Yfke Ongena and the U.S. National Science Foundation, Grant No. IIS-0081550.

because one type of cue compensates for the other. In particular, on the telephone, verbal cues of confusion fill in for missing visual cues. We know from other domains that visual signals – e.g. physically placing an object – can take the place of words (Brennan, 1990; Brennan, 2004; Clark & Krych, 2004). Perhaps when interviewers and respondents cannot see each other respondents compensate for the lack of visual information by displaying more verbal cues of comprehension difficulty (cf. Whittaker, 2003). If this is so, then there should be no mode effects under either conversational or standardized interviewing technique, even if response accuracy is better in conversational interviews.

Table 1 displays how respondents can express comprehension difficulty in telephone and FTF interviewers. Respondents in both modes can convey comprehension problems directly through their words or indirectly with paralinguistic cues such as *um*s and *uh*s. The main difference between the two modes is that visual information is not available on the telephone but is available in face to face interaction. If respondents compensate on the telephone for the lack of visual information we would expect more verbal cues of difficulty, either through their words or paralinguistic utterances.

	FTF	Telephone
Words e.g. "I'm not sure." "What do you mean?"	X	X
Paralinguistic Cues e.g. "*um*s," restarts	X	X
Visual Cues e.g. looks of confusion, nods, direction of gaze	X	

Table 1. Information available in different modes for communicating comprehension difficulty.

We conducted an experiment to examine whether respondents' cues of comprehension difficulty differ by mode (telephone and FTF) and type of interview (standardized and conversational). We also examined whether the presence of visual cues that clarification is needed and the opportunity to exploit those cues leads to more accurate answers.

LABORATORY EXPERIMENT

Design

Interviewers conducted either strictly standardized or conversational interviews, either on the telephone or FTF. This led to four experimental groups (see Table 2). A total of 8 professional Dutch interviewers participated, 2 interviewers in each group. Each interviewer conducted 5 or 6 interviews for a total of 42. Respondents were Dutch students (15 males, 27 females, mean age 22.3 ranging from 19 -28 years) who were paid roughly the equivalent of $25 US to participate. All interviews were conducted in Dutch and carried out at the Free University of Amsterdam.

	Stand.	Conv.
FTF	11	10
Telephone	11	10

Number of Interviewers

Table 2. Experimental design and number of interviews in each experimental condition.

Interviewer Training.

Concepts. All interviewers were trained on the survey concepts being measured in each question. This primarily involved a supervisor assessing their competence with the definition for each concept through mock interviews.

Interviewing Technique. Interviewers were then trained in one interviewing technique or the other. Standardized interviewers were trained to strictly follow the prescriptions of Fowler and Mangione (1990). Interviewers were required to read the question as worded; if the respondent did not provide an adequate answer, the interviewer was instructed to administer a non-directive probe such as "Let me repeat the question" or "Is that a 'Yes' or a 'No?'" If the respondent requested clarification, the interviewers could only respond with non-directive probes such as "Whatever it means to you." The rationale for prohibiting standardized interviewers from defining concepts is that the stimulus would not be standardized if only those respondents who request clarification receive it.

The instruction for conversational interviewers followed the approach of Schober and Conrad (1997) and Conrad and Schober (2000). In this technique interviewers also were to read the question as worded, but they could subsequently provide clarification if respondents explicitly asked for it or if they seemed to need it. Interviewers were instructed to say whatever seemed necessary for the respondent to understand as intended, all or part of the definition, verbatim or in the interviewer's own words. FTF interviewers were not given any special instructions about using visual versus verbal cues.

Experimental Setting

In all of the interviews, the questionnaire was displayed on a laptop computer in front of the interviewer. She read the questions from the computer and entered answers into the computer. The definitions of the survey concepts were printed on a sheet of paper available to the interviewer during the session. For the telephone interviews the interviewer and respondent were situated in separate buildings (see Figure 1).

Figure 1. Interviewer (left) and respondent (right) in telephone interviews.

For the FTF interviews, the interviewer and respondent were seated at a table in the same room (see Figure 2). All interview sessions, whether conducted on the telephone or in person, were video recorded. The images in Figures 1 and 2 are extracted from these videos. The key thing to note is that in the FTF interviews, there is a view of both respondent and interviewer that allows us to determine where the interviewer and respondent are looking, in particular whether they are looking at each other.

Figure 2. Interviewer (left), respondent (right) and view of interviewer and respondent (lower left).

Survey Questions

The questionnaire consisted of 18 questions, 7 of which concerned facts or behaviors (e.g. student status, employment status, and membership in clubs) and 11 of which explored respondents' opinions (6 questions about asylum seekers and 5 about illegal aliens). For the factual items, it was possible to assess data quality using a post-interview questionnaire. After the interview, we administered a paper questionnaire to respondents in which we asked them to answer the same questions they had just answered in the interview. The questions in the paper questionnaire were accompanied by the definition for the relevant concept. Change between respondents' answers in the interview and in the post-interview questionnaire was interpreted as improved comprehension. The logic is that if clarification was given during the interview, respondents' understanding should be more closely aligned with the subsequently presented definition than if no clarification were given. Thus response change reflects the initial misalignment being "corrected" by the subsequently presented definition.

This interpretation of response change can be further tested by comparing the content of responses to what the definitions actually require. For example the response to a question on student status can be checked against to University records; for a question about membership in clubs, the clubs that respondents list can be assessed for whether they fit the definition. Similar logic was used in Conrad and Schober (2000) and Suessbrick, Schober and Conrad (2000).

Analysis Strategy

The current analyses focus on one question about membership in *verenigingen*, which translates roughly as "registered clubs." In the Netherlands, the equivalent of the US Chamber of Commerce maintains lists of registered clubs and membership. Accurately answering the question (which is presented here in Dutch followed by its English translation) requires listing clubs that are legal *verenigingen* and not listing clubs that are not.

> Ik wil je nu enkele vragen stellen over je lidmaatschap van verenigingen. Kun je alle verenigingen opnoemen waarvan je persoonlijk lid bent?
>
> I would now like to ask you some questions about your membership in clubs. Can you list all the clubs in which you are personally a member?

The official definition of *verenigingen* is long and complex. This makes it a good candidate for testing

differences between standardized and conversational interviewing because only in the latter can the interviewer provide information from the definition. Here is an approximate English translation.

> *A 'vereniging' is a legal entity (local authorities and natural persons are legal entities as well). A vereniging has members and has a particular non-profit purpose. The purpose of a vereniging does not have to be idealistic. Any profit cannot be divided among its members but must be spent on the goal of the vereniging.*
>
> *A vereniging may not do either of the following:*
> 1. *Fulfill its members' material needs; an association that does this would be a co-operative society.*
> 2. *Provide insurance for its members.*
>
> *A vereniging is normally established by a notarial deed, containing the articles of association stating the name, the registered seat, the goal and the commitments towards its members. The articles also include the procedure for calling the general meeting and for nominating and discharging members of the board. Members of the board as well as the vereniging bear full responsibility for debts and other obligations.*

One can imagine that respondents' personal notions of what counts as a *vereninging* could differ from this definition; for example, they might consider that they belong to a *vereninging* if their group is profit seeking, offers insurance or has no board. Respondents who are presented with a definition during the interview should be more likely conceive of the *vereniging* as the Chamber of Commerce does.

Results

We turn first to the effect of interviewing technique (strictly standardized or conversational) on data quality. The primary question is whether the current data replicate the advantage of conversational over standardized interviewing for comprehension and response accuracy that we have observed elsewhere (Conrad & Schober, 2000; Schober & Conrad, 1997; Schober, Conrad & Fricker, 2004). If so, we can then examine whether the techniques differentially exploit visual and verbal evidence of comprehension difficulty.

Conversational interviewing can only improve comprehension if interviewers clarify question meaning when, in their judgment, respondents need this. In the current data set, conversational interviewers provided the definition of *verenigingen* in 18 out of the 20 interviews. In contrast, standardized interviewers provided the definition just once in 22 interviewers, consistent with their instructions, χ^2 (1) = 26.64, $p < .001$.

How does the increased provision of definitions in conversational interviewing affect comprehension? One measure is response change between the interview and post-interview questionnaire. Recall that the post-interview questionnaire included definitions of the key survey concepts. Thus, if a respondent in a standardized interview misunderstood the concept in the question – where interviewers could not correct the misconception – the misunderstanding would likely be corrected by the definition. If this happened relatively frequently in the standardized interviews then we would expect notable response change in the post-interview questionnaires. In contrast, we would expect little response change for conversational interviews because conversational interviewers can clarify concepts and correct misconceptions. By this logic, when the definition appears in the self-administered questionnaire, the conversational respondent may well have already have been exposed to the definition, leading to relatively little response change.

We observed more response change in standardized than conversational interviews, as the preceding logic would predict. Responses changed in 7 out of the 22 standardized interviews but only 1 out of the 20 conversational interviews, χ^2 (1) = 6.30, p = .012. When the listed associations were coded for the extent to which they fit the definitions, a greater proportion of those listed in conversational interviews were valid (.93) than those listed in standardized interviews (.78), although this difference was not reliable with the small sample, $F(1,31) = 2.22$, p = 0.14. Conversational interviewers were apparently able to help respondents improve their understanding by providing definitions. But what cues did they use to determine that respondents might benefit from a definition? More specifically, did they rely strictly on verbal cues, which of course are available in both modes, or did they also exploit visual cues in FTF interviews?

If we look at how often interviewers provide definitions in all of the experimental conditions (Table 3, top panel) the frequencies are the same for the two modes for each type of interview (i.e. within each row interviewers provided definitions about as often). And responses during interviews changed – presumably due to definition giving – equally often in each mode (the frequencies are similar within rows in Table 3, bottom panel). This suggests that interviewers – particularly conversational interviewers – were reacting to roughly the same number of cues in both

modes. What's not clear is whether these cues are strictly verbal regardless of mode or whether interviewers rely on a combination of verbal and visual cues in FTF interviews

.

		Telephone	FTF
Definition Giving	Conversational	9	9
	Standardized	0	1
Response Change	Conversational	1	0
	Standardized	4	3

Table 3. Frequency of interviews in which definitions provided.

Verbal indications of comprehension difficulty. As indicated earlier, telephone respondents may produce more verbal cues of comprehension difficulty than their FTF counterparts because telephone respondents cannot convey comprehension difficulty visually. Such verbal cues might include disfluencies like "*um*" and "*uh*." In several studies (Brennan & Williams, 1995), listeners accurately interpret increased rates of *um* and *uh* as evidence that the speaker lacks confidence in her answer to knowledge questions. Schober and Bloom (2004) observed that telephone survey respondents produced *um*s and *uh*s at a higher rate in their first speaking turn after a question that required subsequent clarification than after one that was clear by itself. In the current study, telephone respondents produced *um*s and *uh*s at a higher rate (8.0 per 100 words) than their FTF counterparts (6.1 per hundred words), F (1,41) = 5.26, p = .026 (see Figure 3).

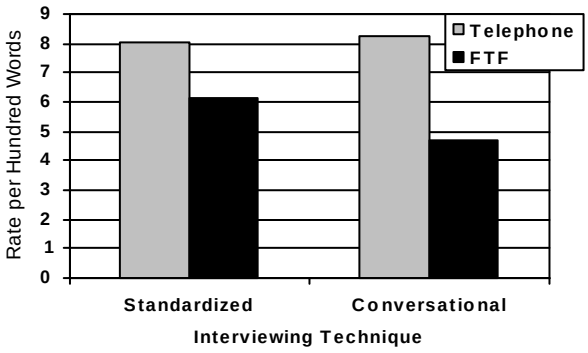

Figure 3. Rate of *um*s and *uh*s uttered by respondents.

Although increased rates of respondent disfluency on the telephone did not lead interviewers to provide more definitions in one mode than the other, it did seem to lead them to take more immediate action on the telephone. The number of conversational "moves" before interviewers provided a definition – where moves correspond roughly to speaking turns – was smaller on the telephone (4.22) than FTF (11.44), F (1, 16) = 13.42, p = .002. Conversational interviewers seemed to get down to business quicker the more respondents provided verbal evidence that they were having comprehension difficulty. This is at least part of the reason why conversational interviews were shorter on the telephone (31.9 moves) than FTF (42.3 moves), although the pattern reverses for standardized interviews, where telephone interviews were longer (21.5 moves) than FTF (12.9), interaction of mode and interviewing technique F(1,38) = 5.12, p = .029 (see Figure 4).

Overall, the conversational interviews involved more moves than did the standardized interviews, F(1,38) = 22.63, p < .001, reflecting the additional talk (moves) required to clarify concepts. The pattern in Figure 4 is almost exactly replicated when interview length is measured in seconds instead of moves. The longer time for conversational than standardized interviews is consistent with our earlier comparisons of these techniques (Conrad & Schober, 2000; Schober & Conrad, 1997; Schober, Conrad & Fricker, 2004).

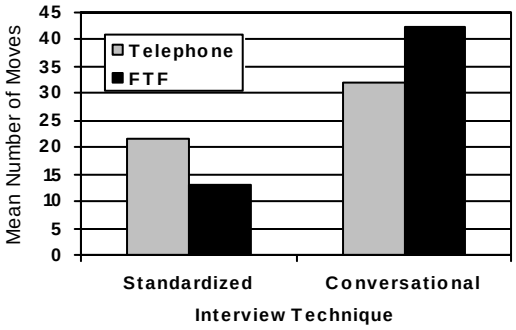

Figure 4. Mean number of moves per interview.

Visual indications of comprehension difficulty. Telephone respondents' higher disfluency rates may reflect their recognition – at some level – that they cannot visually convey information about their comprehension state. What then is the visual information that respondents might convey in FTF interviews to indicate trouble answering? A behavior that respondents could exhibit and interviewers could observe which has been associated with difficult mental processing is gaze aversion. In particular, several studies have demonstrated that people look

away from their conversational partner while answering difficult questions (e.g. Glenberg, Schroeder, & Robertson, 1998; Doherty-Sneddon, Bruce, Bonner, Longbotham & Doyle, 2002). These studies have been used to argue that people avert the gaze of the questioner in order to temporarily eliminate visual information about the questioner's face which might be distracting and would otherwise be hard to ignore.

In the current experiment, FTF respondents looked away from the interviewer while they answered questions for longer percentages of the move's duration in conversational (15.7%) than standardized (2.1%) interviews, $F(1,19) = 6.25$, $p = .018$. In Figure 5, a respondent looks away from the interviewer while answering. This is even more pronounced if we look just at unreliable answers, i.e. those answers which respondents later changed after reading a definition. Respondents looked away from the interviewers for 20.4% of the time taken to provide an unreliable answer in conversational interviews, but 0% (i.e. respondents never looked away) when providing unreliable answers in standardized interviews. Despite the small number of cases – we only observed 5 cases of gaze aversion in conversational interviews in this analysis – the duration and frequency of gaze aversion seems to differ in the two kinds of interviews. Respondents may avert gaze more often and for longer periods when interviewers have the freedom to react to such visual cues by providing clarification.

Figure 5. Respondent (right) looks away from interviewer while answering question.

In order for an interviewer to react to a respondents' averted gaze, she needs to look at the respondent often enough to see that he is looking away. The proportions of time that conversational interviewers look at the computer, at the paper containing the definitions, and at the respondent are displayed in Figure 6. While the interviewers certainly looked at targets other than the respondent, it is possible they looked at the respondent often enough (31% of the time) to perceive periods of looking away while answering. In fact, the interviewer looks at the respondent at some point during 80% of the moves in which respondents looked away while answering the question. However, we do not know how often they looked at respondents precisely when respondents looked away or whether they were attending to this cue when respondents looked away.

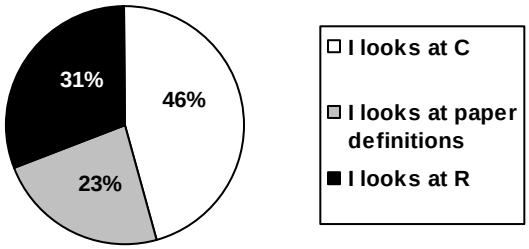

Figure 6. Percent of time interviewers look at computer (C), definitions and respondent (R).

Given that conversational interviewers could see respondents' averted gaze, do they provide clarification more often when it occurs than when it does not? The answer is no. Clarification was no more likely after respondents looked away than when they gazed at the interviewer.

Why didn't interviewers treat an apparent indication of respondent comprehension difficulty as an opportunity to improve understanding? One possibility is that interviewers simply did not look at respondents enough during periods of gaze aversion to see that it was occurring. Second, it is possible that even if they saw that respondents were looking away, they did not interpret this as an indication of need for clarification. Recall that we did not provide any instruction to interviewers on how interpret respondents' visual cues. Given their professional interviewing experience it is possible they concluded they could be penalized for engaging in this kind of behavior. A third possibility is that interviewers noticed the respondent's behavior but treated it as a signal that the respondent was not finished with his current speaking turn, and not as an indication of comprehension difficulty (see Goodwin, 2000). If this is the case, respondents might exhibit this kind of behavior in conversational and not in standardized interviews because only conversational interviewers are likely to interrupt respondents while they are thinking to offer unsolicited clarification.

SUMMARY AND CONCLUSIONS

Data quality, measured by response change and validity of listed clubs, was better under conversational than standardized interviews but was not affected by the mode of the interview. Mode did affect the way respondents indicated their need for clarification depending, in particular, on the type of information afforded by the mode and what actions, if any, interviewers could take to help resolve comprehension difficulties.

Telephone respondents produced more disfluencies (*um*s and *uh*s) than FTF respondents, possibly to compensate for the lack of visual information in telephone communication. FTF respondents seemed to exploit the availability of visual information (by averting the interviewers' gaze while answering questions) more in conversational than in standardized interviews; presumably respondents were sensitive to conversational interviewers' flexibility – and standardized interviewers' inflexibility – to respond to such cues. Conversational interviewers looked at respondents while they looked away from the interviewer but did not provide clarification more often during these periods than when respondents did answered without looking away.

The current results suggest that mode effects may be rare in the literature because mode comparison studies typically involve standardized interviews. The current results suggest that data quality – which is generally higher for conversational than standardized interviews – could be particularly high for conversational FTF interviews. We believe that respondents' gaze aversion holds promise as a visual cue that conversational interviewers might exploit in deciding when to provide clarification. Of course we have only looked at one question in this rich data set; clearly the results must be replicated and extended with other items and in other interviews.

Implications for practice

When practitioners choose a mode for a survey, the interviewing technique needs to be considered as well. If the study requires high confidence that respondents understand as intended, conversational interviewing may be appropriate and FTF interviews may provide richer cues to interviewers that respondents need clarification. What we don't yet know is if conversational interviewers can be trained to recognize respondent's visual behaviors such as gaze aversion as evidence they need help and to then intervene appropriately. Beyond this we will need to know what kinds of problems are indicated by different visual cues and what the best way is for interviewers to respond.

REFERENCES

Brennan, S.E. (1990). *Seeking and providing evidence for mutual understanding.* Unpublished doctoral dissertation, Stanford University.

Brennan, S.E. (2004). How conversation is shaped by visual and spoken evidence. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions* (pp. 95-130). Cambridge, MA: MIT Press.

Brennan, S.E., & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language, 34*, 383-398.

Clark, H.H. & Krych, M. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language, 50*, 62-81

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

DeLeeuw, E.D., & van der Zouwen, J. (1988). Data quality in telephone and face to face surveys: A comparative meta-analysis. In R.M. Groves, P.P Biemer, L.E. Lyberg,, J.T. Massey, W. L. Nicholls, & J. Waksberg, (Eds.), *Telephone survey methodology* (pp. 283-300). New York: John Wiley & Sons, Inc.

Doherty-Sneddon, G., Bruce, V., Bonner, L., Longbotham, S., & Doyle, C. (2002). Development of gaze aversion as disengagement from visual information. *Developmental Psychology, 38*, 438-445.

Dykema, J., Lepkowski, J.M., & Blixt, S. (1997). The impact of interviewer and respondent behavior on data quality: Analysis of interaction coding in a validation study. In L. Lyberg, P. Biemer, M. Collins, E. De Leeuw, C. Dippo, N. Schwarz, & D. Trewin, (Eds.) *Survey measurement and process quality* (pp. 287-310). New York: Wiley and Sons, Inc.

Glenberg, A. M., Schroeder, J. L., & Robinson, D. A. (1998). Averting the gaze disengages the environment and facilitates remembering. Memory & Cognition, 26, 651-658.

Goodwin, C. (2000). Practices of seeing: Visual analysis: An ethnomethodological approach. In van T. Leeuwen, & C. Jewitt, (Eds.), *Handbook of visual analysis* (pp. 157-182). London: Sage.

Schober, M.F., & Bloom, J.E. (2004). Discourse cues that respondents have misunderstood survey questions. *Discourse Processes, 38*, 287-308.

Schober, M.F., & Conrad, F.G. (1997). Does conversational interviewing reduce survey measurement error? *Public Opinion Quarterly, 61*, 576-602.

Schober, M.F., Conrad, F.G., & Fricker, S.S. (2004). Misunderstanding standardized language in research interviews. *Applied Cognitive Psychology, 18*, 169-188.

Viterna, J.S., & Maynard, D.W. (2002). How uniform is standardization? Variation within and across survey research centers regarding protocols for interviewing. In D.W. Maynard, H. Houtkoop-Steenstra, N.C. Schaeffer, & J. van der Zouwen, (Eds.) *Standardization and tacit knowledge: Interaction and practice in the survey interview* (pp. 365-397). New York: Wiley.

Whittaker, S. (2003). Mediated communication. In A.C. Graesser, M.A. Gernsbacher, & S.R. Goldman (Eds.), *Handbook of discourse processes* (pp. 243-286). Mahwah, NJ: Erlbaum.