

A CONVERSATIONAL APPROACH TO COMPUTER-ADMINISTERED QUESTIONNAIRES¹

Frederick G. Conrad, Bureau of Labor Statistics
Michael F. Schober, New School for Social Research
Frederick G. Conrad, Bureau of Labor Statistics Room 4915,
2 Massachusetts Ave., N.E., Washington, DC 20009

Key Words: CASI, interface design,
conversational interviewing, data quality

INTRODUCTION

What is the future of the interviewer in collecting survey data? Will interviewers be necessary in the era of computer assisted self interviews (CASI)? These questions have been brought into sharper focus by the arrival of the World Wide Web which, in principle, makes it easy to deliver a computer administered questionnaire to anyone in a sample. And there may be other advantages to this method: for certain types of questions, CASI seems to improve the quality of data relative to interviewer administered questionnaires.

In particular, CASI seems to give respondents a sense of privacy relative to interviewer administered questionnaires: CASI respondents seem more willing to accurately report sensitive behaviors like drug use and certain sexual activities than their counterparts answering questions posed by an interviewer (O'Reilly, Hubbard, Lessler, Biemer, & Turner, 1994; Tourangeau & Smith, 1996). This advantage is increased when the CASI system presents spoken questions over headphones rather than as text (O'Reilly et al., 1994; Tourangeau & Smith, 1996).

Another reason that CASI may be good for data quality is that it promises to eliminate interviewer-related error from the total error for a particular survey. Under self-administration there is no interviewer to potentially mislead or bias respondents. And with computer assisted self interviews (as opposed to self administered paper questionnaires), it is possible to assure that all respondents answer the questions in the intended order. In many respects, CASI makes it possible to truly standardize the data collection process. As is the case in the ideal standardized interview (e.g. Fowler & Mangione, 1990), CASI respondents do not obtain information beyond what is included in the question itself.

But removing interviewers from the data collection process also eliminates the conversational skills that interviewers bring to the job. We have found that allowing interviewers to use their conversational skills to clarify concepts can substantially improve response accuracy (Schober & Conrad, 1997; Conrad & Schober, under review). In these studies, "conversational interviewers" worked with respondents to make sure they understood concepts as intended. They defined terms when respondents asked for clarification and when they judged that respondents did not understand particular terms. In addition they could ask the respondents exploratory questions. Thus they could resolve respondent confusion about situations such as: Does buying a lamp count as a furniture purchase? Does baby-sitting for multiple employers count as more than one job? Does a person away at college live at home?

We found that this type of conversational intervention was effective primarily when the concepts in survey questions did not clearly correspond to respondents' circumstances. In such situations, conversational interviewers produced nearly 60% more accurate responses than standardized interviewers who could not exercise such flexibility (Schober & Conrad, 1997). But this improvement came at a cost: conversational interviews took more than three times as long as standardized interviews.

Is it possible to implement this kind of conversational flexibility in CASI instruments, or does it rely on uniquely human skills? If it *is* possible, would the costs be so high? In the current paper we explore possible techniques for clarifying concepts in CASI instruments. Although one can imagine doing this with advanced software techniques that involve natural language comprehension and plan recognition, here we implemented clarification facilities using simple, conventional programming techniques. In addition, we examined how different instructions to respondents affect their use of such clarification facilities.

¹We thank Susan Brennan, Cathy Dippe, Scott Fricker, Susan Schnipke and Clyde Tucker for their advice and assistance. This material is based upon work supported the National Science Foundation under grant No. SBR-97-30140 and by the Bureau of Labor Statistics. The opinions are those of the authors and not the Bureau of Labor Statistics.

LABORATORY EXPERIMENT

In this experiment, respondents (users) interacted with a CASI system that presented 12 questions from ongoing government surveys. So that we could measure response accuracy directly, respondents answered based on fictional scenarios (described below). We measured response accuracy for (1) a conventional interface, where the interpretation of questions is left entirely to respondents; (2) an interface that provided respondents with information (official definitions) to clarify survey concepts when (and only when) they requested it by clicking highlighted text; and (3) an interface that both provided respondents with information when they requested it and also offered information when respondents indicated they were uncertain about how to answer – when they were inactive for too long.

A second issue we examined was how willing respondents would be to use such help. Research in other testing situations has shown that people rarely ask for help when they need it (e.g., Graesser, Swamer, Baggett, & Sell, 1996). In our own research on interviewer-respondent interaction (Schober & Conrad, 1997; Conrad & Schober, under review), respondents in survey interviews using similar materials also asked for help far less often than their answers showed they needed it.

To examine this second issue, we varied instructions to respondents (Rs) before they used the CASI instrument. Some respondents were told that getting official definitions from the computer was essential for responding accurately (“definitions essential”) because their everyday definitions might differ from those of the researchers. Other respondents were merely informed that definitions were available from the computer if they wanted them (“definitions available”), but they were not explicitly encouraged to use them. When no definitions were available, respondents were not given any instructions about definitions.

So there were five experimental conditions:

	<u>Type of help</u>	<u>R instructed that...</u>
1	no help	
2	at R's request	definitions essential
3	at R's request	definitions available
4	when R takes too long <u>or</u> at R's request	definitions essential
5	when R takes too long <u>or</u> at R's request	definitions available

Questions. All respondents answered the same 12 questions from three surveys: four employment questions from the Current Population Survey (e.g., “Does anyone in this household have a business or a farm?”), four housing questions from the Consumer Price Index-Housing survey (e.g., “How many people live in this house?”), and four purchase questions from the Current Point of Purchase Survey (e.g., “Last week, did Carla have any purchases or expenses for car tires?”). The three question domains (employment, housing, purchases) were randomly ordered for different respondents, although the questions within a domain were always presented in the same order as they appeared in the original surveys. Official definitions existed for key concepts in all questions.

Because these same questions were asked by interviewers in Schober and Conrad (1997), we could compare how often respondents asked the CASI system for help to how often they asked interviewers for help.

Stimuli. Respondents answered on the basis of fictional scenarios: textual vignettes, floor plans, or receipts from purchases. For each question there were two scenarios. With one scenario, the survey question was designed to be easy for respondents to interpret—to map onto the fictional circumstances in a *straightforward* way. For example, for the question “Last year, did Kelley purchase or have expenses for household furniture?”, the scenario leading to a straightforward mapping was a receipt for an end table, which is clearly a piece of furniture. With the other scenario, it was less clear how the survey question should be answered; the scenario led to a *complicated mapping* between the question and the respondent’s circumstances. For example, for the household furniture question the scenario leading to a complicated mapping was a receipt for a floor lamp, which is harder to classify without knowing the official definition of “household furniture.”

For any one question, a respondent would see either the scenario leading to a straightforward mapping or the scenario leading to a complicated mapping. In each interview, half the scenarios led to straightforward mappings and half to complicated mappings. The official definitions always clarified what the correct answers should be.

CASI interface. Respondents used the keyboard and mouse to enter responses and navigate the questionnaire, which was implemented as a Windows application on a desktop computer. When definitions were available, a word or phrase was highlighted, and respondents requested definitions by clicking the mouse on the highlighted text (conditions 2-5). When

they clicked, the official definition would appear as text in another window. In conditions 4 and 5, respondents were also offered help when they took too long to respond, that is, when their response time was longer than the median response time for complicated mappings for that question in the No Help condition (ranging from 12 to 46 seconds). A dialogue box stating "Do you want help?" would appear, and respondents could accept help by clicking "yes" or reject help by clicking "no."

Participants. 54 paid respondents were recruited from an advertisement in *The Washington Post*. 22 were women and 32 were men; 13 were Black, 38 White, and 3 Asian; and they came from a range of educational backgrounds (24 high school only, 21 college degrees, 9 with postgraduate education). Most participants were fairly experienced computer users; 44 reported using a computer every day, 5 once a week, 2 once a month, and 3 once a year.

RESULTS

Overall response accuracy. Accuracy was defined as the extent to which answers matched what the official definitions required. For straightforward mappings, accuracy was nearly perfect in all five conditions. For complicated mappings (see Figure 1), accuracy was poor for respondents in the "no help" condition. Response accuracy was no better in both conditions where respondents were merely informed that definitions were available, whether they received help

only when they asked for it (condition 3), $F(1,49) = 1.71, n.s.$, or whether the CASI system also offered help when they took too long (condition 5), $F(1,49) = 0.01, n.s.$ Response accuracy for complicated mappings was much better when respondents had been instructed that definitions were essential, whether respondents received help only when they asked for it (condition 2), $F(1,49) = 9.82, p < .01$, or whether the system also offered help when they took too long (condition 4), $F(1,49) = 14.38, p < .01$.

This suggests that it is not sufficient to merely make help available in a conversational CASI system. Respondents must also believe that the help is necessary.

When was help obtained? As Figure 2 shows, when respondents had been told that definitions were essential they asked for help most of the time, both for straightforward and complicated mappings. They asked the computer for help far more frequently than respondents asked telephone interviewers for help in an otherwise comparable study using the same materials (Schober and Conrad, 1997). This makes sense because asking for help with CASI is low cost (see Clark & Brennan, 1991; Schwarz et al. 1991), in at least two ways. First, respondents needed only to click the mouse to obtain help, which is far simpler than planning and uttering a question to ask an interviewer. Second, while respondents with human interviewers may be reluctant to ask for help understanding everyday terms, there is little social cost to obtaining help from the computer.

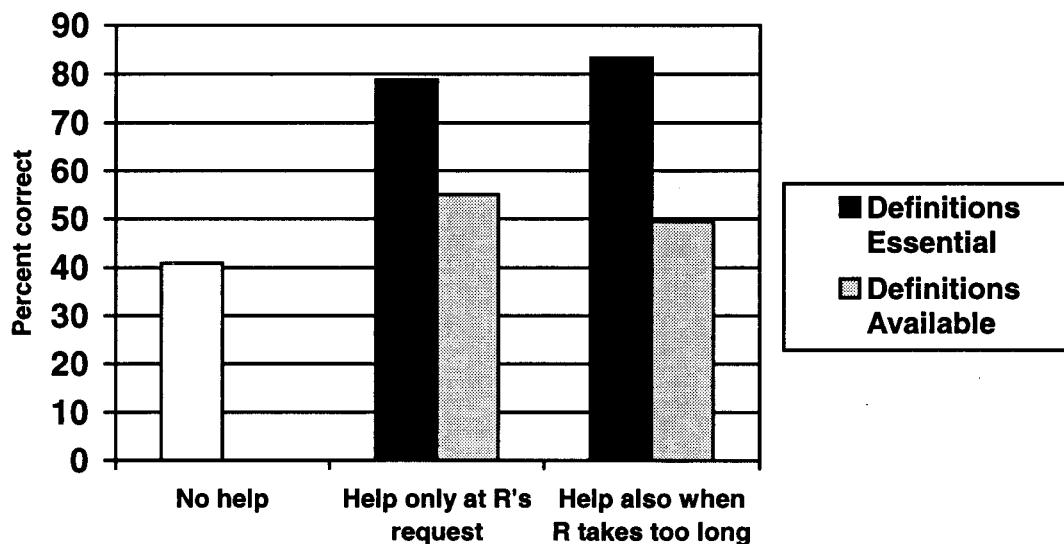


Figure 1. Response accuracy for complicated mappings

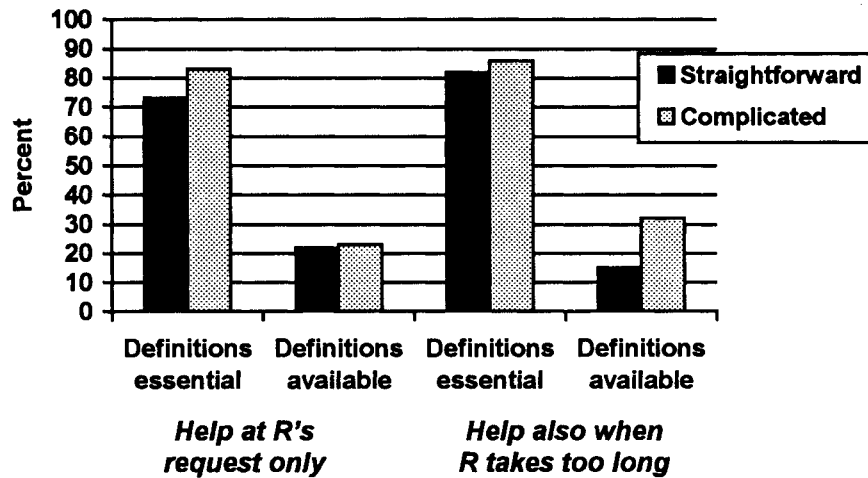


Figure 2. Percentage of questions on which help was obtained

In contrast, when respondents were told merely that definitions were available, they rarely asked for help. In fact they asked for help less often than in the Schober and Conrad (1997) telephone interviews. Perhaps respondents using computers don't feel compelled to be as conscientious as they do with human interviewers

In the Schober and Conrad (1997) interviews, respondents asked for help almost exclusively for complicated mappings, when, presumably, they knew they needed help. In this study, respondents asked the computer for help just about as often for both straightforward and complicated mappings, which means that they asked for help in many cases where they probably didn't need it. We suspect this is because asking the computer for help was low cost.

How help affected accuracy. Even if respondents were told that definitions were essential, this didn't mean they would always ask for help for complicated mappings. For those cases where respondents didn't ask for help, or where they answered quickly enough that the CASI system didn't provide help (Condition 5), they were quite inaccurate (see Figure 3). Accuracy for complicated mappings was high only when respondents actually received the help, whether they requested it or the computer offered it, and whether or not they were told that definitions were essential.

Duration of surveys. Not surprisingly, respondents took longer to complete the surveys in the conditions where they received the most help, $F(4,49) = 3.38, p < .02$. Respondents in the No Help condition (Condition 1) averaged 25 seconds per question. Respondents who

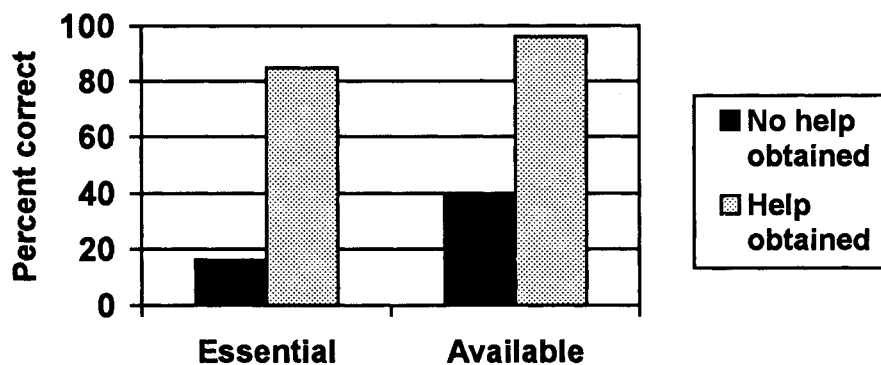


Figure 3. Accuracy for complicated mappings when any help was obtained

had been instructed that help was merely available (Conditions 3 and 5) didn't take reliably longer (34 and 27 seconds, respectively). But respondents who had been instructed that definitions were essential (Conditions 2 and 4) took about twice as long to complete each question, 53 and 47 seconds.

Respondents took longer to complete questions with complicated mappings (41 seconds) than questions with straightforward mappings (33 seconds), $F(1,49) = 7.91$, $p < .007$. This was true in every condition, whether or not they received help, interaction of mapping x condition $F(4,49) = 0.13$, *n.s.*

So, consistent with the accuracy results, if respondents didn't get help or didn't believe help was important, they were relatively quick. Getting help took time, just as it did in the Schober and Conrad (1997) conversational interviews. But the twofold increase in time for getting help here contrasts favorably with the threefold increase in Schober and Conrad (1997). And longer CASI sessions, which don't use interviewer resources, may be less costly than longer interviews. However, longer CASI sessions might reduce completion rates if respondents find the length burdensome.

User satisfaction. After the respondents had finished the experimental session, they were asked to complete a paper questionnaire about the CASI system. Overall, they viewed their interaction with the CASI system favorably. Respondents in all conditions used phrases like "easy" and "clear" to describe their overall experience, although a very few (three) indicated that they found the questions "tricky." The overall approval did not seem to be affected by the version of the CASI system or the instructions about using definitions.

When asked "How would you feel if surveys were actually administered this way in the future?" 94% of responded favorably, regardless of condition. The fact that respondents who were instructed to use the definitions did not evaluate the interaction less favorably than did their counterparts suggests that obtaining definitions was not particularly onerous.

This reaction was consistent with the respondents' preference for the CASI system over interviewer administered sessions (85% preference for CASI) or self administered paper questionnaires (85% preference for CASI). Again, this was not affected by the type of help or instructions.

Not surprisingly, respondents' perception of the value of the definitions was affected by the instructions. Respondents who could not obtain definitions (condition 1) were asked if they would have liked to be able to do so; the three respondents who

said "no" justified this response on the grounds that definitions did not seem necessary. Respondents who could get definitions (conditions 2-5) were asked if they obtained them; if they said no, they were asked why. Of those who were not told that they needed definitions, three of the eleven in condition 3 and five of the ten in condition 5 reported that they did not use them because definitions did not seem necessary. In contrast, all of the respondents who were told that definitions were necessary reported using them. Eight of the 11 respondents who could not obtain definitions (condition 1) reported that they would have used definitions if available and indicated that definitions would have helped them for 11 out of the 12 concepts involved in complicated mappings. This suggests that these respondents were aware of their own uncertainty and that they believed that they would have been more accurate if they could have obtained definitions.

Six of the respondents who were offered unsolicited help by the CASI system accepted it. They seemed to react differently to this type of help depending on their pre-survey instructions. The two respondents who received this help and who had been instructed to use help (condition 4) rated it as quite useful (6.0 on a 7-point scale), but the four respondents who accepted definitions but had not been instructed to use them (condition 5) did not rate the help as so useful (3.9 on the 7-point scale). When asked "Did you find this annoying?" condition 4 respondents' mean rating was 1.0, but for condition 5 it was 4.25.

SUMMARY

Our results show, first, that conversational techniques for resolving misconceptions can indeed be implemented in CASI instruments: respondents can request definitions and the computer can offer definitions to clarify concepts. And, at least in this setting, user satisfaction doesn't seem to decline with the extra work of getting and reading definitions.

Second, consistent with the Schober and Conrad (1997) findings, people don't necessarily ask for help when they need it. Respondents rarely asked for help if they were only told that it was available. If they were told that definitions were essential, respondents were much more likely to ask for them. Apparently pre-survey instructions to respondents make an enormous difference in how likely respondents are to ask for help in this environment.

Third, at least for these materials, respondents seem to request help more often with the CASI instrument than with human interviewers. That is, respondents who had been instructed that definitions

were essential asked for help not only when their fictional circumstances led to a confusing mapping with the survey concepts but also when the mapping was straightforward. We expect that this is because asking for help takes less effort with a computer—it requires merely a mouse click—and involves few if any social risks. Also, respondents can control the pace and level of detail of the interaction, which in some circumstances can be desirable (e.g. Shneiderman, 1997).

Fourth, as in the Schober and Conrad (1997) study, the benefits of obtaining help understanding the survey designers' definitions are independent of whether the respondent requests help or the computer (or interviewer) offers help unsolicited. Any help improves accuracy for complicated mappings. One might propose, therefore, that definitions should *always* be provided, perhaps even as part of the question. In actual interviews this would be impractical, given how long and complicated definitions are, and how unlikely it is that every respondent needs to hear all the details of every definition. But in CASI, this might be more feasible. Of course, respondents might not take the time to read complete definitions or might find them annoying.

We have implemented only a rudimentary interface for resolving potential misunderstandings. There are other formats for presenting definitions (hierarchical, searchable, diagrammatic). User satisfaction or response accuracy might increase with systems that rely on adaptive feedback (e.g., Brennan & Hulteen, 1995) or detailed explanations (Moore, 1995). But standard principles of human-computer interaction for systems in which users take the initiative may not apply directly to computer-based surveys; in surveys, the system solicits information from a (possibly unwilling) respondent, and so goal structure and control may differ from user-initiated interactions.

We recommend that designers of computer-administered questionnaires consider conversational techniques. Conversational CASI may provide more accurate data, and it may allow respondents to obtain the information they need in a systematic way at far less administrative cost than human conversational interviewing. But the data show that respondents don't have particularly good metacognitive skills and that leaving it up to respondents to decide when they need help may lead to poor response accuracy.

REFERENCES

Brennan, S.E., & E.A. Hulteen. 1995. "Interaction and Feedback in a Spoken Language System: A

Theoretical Framework." *Knowledge-Based Systems* 8:143-151.

Clark, H.H., & S.E. Brennan. 1991. "Grounding in Communication." In L.B. Resnick, J. M. Levine, & S.D. Teasley, *Perspectives on Socially Shared Cognition*, pp. 127-149. Washington, DC: APA.

Conrad, F.G., and M.F. Schober. "Conversational Interviewing Can Reduce Survey Measurement Error: Evidence from a Household Telephone Survey." Manuscript under review.

Fowler, F.J., and T.W. Mangione. 1990. *Standardized Survey Interviewing: Minimizing Interviewer-Related Error*. Newbury Park, CA: SAGE Publications, Inc.

Graesser, A.C., S.S. Swamer, W.B. Baggett, and M.A. Sell 1996. "New Models Of Deep Comprehension." In B.K. Britton and A.C. Graesser (eds.), *Models of Understanding Text*, pp. 1-32. Mahwah, NJ: Lawrence Erlbaum Associates

Moore, J.D. 1995. *Participating in Explanatory Dialogues: Interpreting and Responding to Questions in Context*. Cambridge, MA: MIT Press.

O'Reilly, J., M. Hubbard, J. Lessler, P. Biemer, and C. Turner. 1994. "Audio and Video Computer Assisted Self-Interviewing: Preliminary Tests of New Technology for Data Collection." *Journal of Official Statistics* 10:197-214.

Schober, M.F., and F.G. Conrad. 1997 "Does Conversational Interviewing Reduce Survey Measurement Error?" *Public Opinion Quarterly* 61:576-602.

Schwarz, N., F. Strack, H.J. Hippler, and G. Bishop. 1991. "The Impact of Administration Mode on Response Effects in Survey Measurement." *Applied Cognitive Psychology*, 5, 193-212.

Shneiderman, B. 1997. "Direct Manipulation For Comprehensible, Predictable, and Controllable User Interfaces." *Proceedings of International Conference on Intelligent User Interfaces*, 33-39.

Tourangeau, R., and T. Smith. 1996. "Asking Sensitive Questions: The Impact of Data Collection Mode, Question Format, and Question Context." *Public Opinion Quarterly* 60:275-304.