SCREEN DESIGN AND QUESTION ORDER IN A CAI INSTRUMENT
EFFECTS ON INTERVIEWERS AND RESPONDENTS

Marek Fuchs, Catholic University of Eichstätt
Catholic University of Eichstätt, 85071 Eichstätt, Germany

Key Words: Computer Assisted Interviewing; Usability Research; Screen Design; Question Order

1 Introduction and Hypothesis

Computer assisted interviewing (CAI) is on its way to become a standard survey technique (Couper et al., 1998). Many of the large government surveys in the US are in the transition to CAI or have completed it already. Even in Europe, we observe a shift towards computer assisted interviewing (Schneid, 1991; Fuchs, 1994, 1995; Laurie/Moon, 1997; Projektgruppe SOEP, 1998).

In general the movement towards CAI is evaluated positively. Researchers and field directors benefit from it (Nicholls/deLeeuw, 1996) and interviewers (Couper/ Burt, 1994) as well as respondents (Baker, 1992) reveal a great deal of acceptance. On the other hand, computer assisted interviewing has introduced some additional problems into the interview situation, too: in the early years methodological research was mainly concerned with hardware and software problems (see Couper/ Groves, 1989; Weeks, 1992 for overviews). Instead, recent studies dealt with interview and respondent acceptance, interview duration, and usability issues (Couper et al., 1998 for an overview). The present paper contributes to this discussion of "technology effects" (Fuchs et al., 2000). For the purpose of the following analysis the focus is mainly on two usability issues: (1) segmentation of the interview flow and (2) lack of interviewer flexibility in a CAI environment.

(1) Segmentation: in a CAPI environment the interviewer got an additional burden: the process of keying takes place in the interview situation. Usually, an interviewer reads a question, gets an answer, enters the data, presses [enter] and then the next screen with the following question comes up. As a result of this procedure the interviewer respondent interaction is segmented. Compared to PAPI interviewers cannot look ahead and anticipate the next upcoming question while recording the answers to the previous one and they cannot start reading the next question before pressing [enter]. So far we do not have quantitative evidence that this segmentation harms the data or the interview situation. But it is argued that the interviewer loses the "big picture", and the relevance of questions may be unclear (House, 1985; Groves/Mathiowetz, 1984).

In 1997, two series of tests concerning the screen layout of a household roster were conducted (Couper et al., 1997). Two different versions of a series of questions were tested under laboratory conditions in terms of the time necessary in order to complete the questions and ease of use. We compared a so-called item based design with a grid based design.

The item version tested in our experiment matches the characteristics specified by House and Nicholls (1988) for an item based approach: one question and one input field are displayed at a time, and logic operations are performed in the transition from one item based screen to the next. In contrast, the grid based design is best described as a form based instrument: many items are presented at the same time in a table or grid and the interviewer may use the cursor keys move from field to field and to complete them in any order. It allows interviewers to record the information in the order chosen by the respondent, it provides the interviewer with a better overview of the instrument and it more easily allows updates and backups (for details see Couper et al., 1997).

We found evidence that the grid based design reduces the segmentation: interviewers could start reading the next upcoming question while still entering the data to the previous question. Even backing up seems to be easier within a grid design. On the other hand, we found only modest support for a grid based design in terms of time used to complete the task (Couper et al., 1997). This leads to the question: what can we do to decrease segmentation and to further improve the efficiency of a household roster in terms of duration?

(2) Lack of flexibility: The second feature that might cause problems in a computer assisted interview is the lack of flexibility. One of the advantages of a CAI instrument is the fact that an interviewer can hardly skip any questions. Although CAI instruments can make extensive use of skip patterns and filters, they apply a predefined question order. Usually, each question needs an [enter] key before the system goes on to the next screen. It is seen as an advantage that this rigid question order avoids any trouble the interviewer might have with the routing through the instrument, questions for specific respondents, filters and skip patterns and so on. He or she can abandon this task and focus on the administration

of the actual items. On the other hand, this causes a very strict question order and provides the interviewer with little flexibility in terms of question order.

A small example demonstrates this effect: most CAI instruments apply a question order to their household roster, where all items for one person are asked before the interviewer works through the same items for the next person ("person based design" see Couper et al., 1997; Fuchs et al., 2000 or "grouped questions" see Moyer, 1996). When completing the questions of a household roster it might happen (and in fact it happens quite often, see below) that the respondent provides not only the answer to the current question (e.g. "I'm 34 years old") but also to a related question: "I'm 34 years old and my wife is 32 years old" or the respondent might answer "We are all Black" (Oksenberg et al., 1992).

While working with a paper instrument it is an easy task for an interviewer to make immediate use of the additional information provided by the respondent. In case he or she answers, for instance, "We are all black" the interviewer can easily mark the appropriate check boxes for all household members at once. For someone interested in questionnaire design this leads to the following question: given the lack of flexibility in a computer assisted environment, what is the best question order for collecting information about all household members?

Moore and Moyer report results from an experiment on two different question orders designed for collecting information about all eligible persons in a household (Moore/Moyer 1998a, 1998b). The first question order asks all questions for the first eligible person in the household and moves on to the next person, when all questions are completed. This question order is called a person based approach. In the second version, the topic based approach, the first question is asked for all eligible persons, then the second question for all persons and so on. Moore's and Moyer's results show strong support for a topic based design: this version leads to less item non-response, less break offs and refusals and is substantially shorter. Besides interviewers show significant preference for this version.

In the experiment presented in this paper we combined the two screen designs (item based design vs. grid based design) with the two question orders (person based order vs. topic based order) and tested all four resulting versions in a field experiment. In doing this, we had the following assumption in mind: the usability of a CAI instrument is not only a programming issue, but it is also connected to the questionnaire design and also to the

interview as a social situation. Both aspects of a computer assisted instrument, its screen design and its question order, support or hinder a smoothness of the interview flow. Based on the results of the previous research we had the following hypothesis: The combination of a grid based screen design and a topic based question order allows the most efficient interviewer respondent interaction.

2 Methods

The experiment took place in Germany in March 1998. Immigrants of German origin from Poland, Rumania and the former Soviet Union were surveyed. Starting February 28, 1998 and ending March 20, 1998 15 interviewers completed n = 501 interviews. All respondents received an advanced letter and were called by phone up to 15 times. Item non-response was considerably low and the response rate reached 84 %. The interviews were conducted using the CATI program CI3. About 95 questions on various topics were asked. The average interview lasted 23 minutes.

Figure 1. Four Versions Tested in the Experiment (Each Box Represents One Screen)
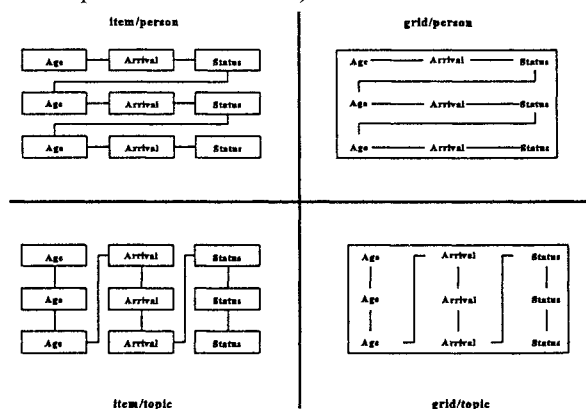
Four versions of a small household roster with just three items per person were included in the instrument: an item/person version, a grid/person version, an item/topic version and a grid/topic version (figure 1). The item based person version is considered to be the standard version – it represents the questionnaire design usually applied to socio-demographic portions in CAI surveys. One of the four versions was randomly assigned to each interview – and thus to interviewers and respondents. We measured the total time needed for the household roster and in addition the time spent on each single item in that section of all 501 interviews. In addition, 234 interviews were selected at random and the interviewer working through the household roster section was video-taped. The video segments were coded in terms of interviewer

behavior and respondent behavior and the resulting data was combined with the time measurements.

3 Results

The durations of the four versions differ significantly from each other: interviewers needed 6.6 seconds per item in the item based person version (which is considered to be the standard). In contrast each item took 5.5 seconds in the grid based topic version. This is a reduction of about 17 % for the grid based topic version. The two other versions are in between.

It is important to mention that both factors seem to contribute to the decrease in time used to complete the task. If we distinguish between the two factors, we end up with the following results: the two topic based versions are significantly shorter than the two person based versions and the two grid based versions take significantly less time than the two person based versions. The combined effect applies to the grid based topic version and leads to the value of 5.5 seconds per item. An analysis of variance reveals that both factors – the screen design as well as the question order – contribute independently to the decrease in time (screen design: $p < 0.01$, one third of total effect; question order: $p < 0.001$, two thirds of total effect; no significant interaction).

Analyzing the video tapes we can provide reasons for these differences at least in part: given the grid based topic condition, both, interviewers and respondents, adapt differently to the interview situation compared to the item based versions. When asking the questions for all persons in the household, the respondent recognizes the logic of the procedure very quickly. In quite a high proportion of all cases (about 30 %) their reaction to this is "We all arrived in the same year" (meaning: "Don't ask me this question again and again").

If the instrument follows a person based design, the interviewer has to memorize this piece of information, and if it comes to the next person, he or she needs to remember: "Do not ask this question again, the respondent gave you the appropriate answer already!" Only in a few cases they really do, most of the time they just ask the question again. This is especially true when using an item based screen layout that gives no clues in terms of the answers to the same question for the other household members. In a topic based design instead, the interviewer can easily adapt to that situation. Thus he or she just enters the same code for all persons in the household without asking the question repeatedly. Both interviewer and respondent can anticipate the next question and the

interview runs more smoothly. As a result the time used per item is substantially shorter and the interviewer can provide respondent oriented interviewer behavior similar to Schober and Conrad's (1997) findings.

Providing feedback by the interviewer sometimes works as a signal that he or she has recorded the answer to the previous question in order to stimulate the respondent, so that the latter guesses about the next question and reveals the appropriate answer even without an additional stimulus. In extreme this might lead to a respondent behavior, where he or she provides the information about all persons in the household at once: "We all came in the same year". The different versions tested in this experiment impel and support such behaviors to different degrees. From our results we can conclude that the grid based topic version stimulates interviewers to deviate from the scripted interview to a higher degree than the other versions. As far as duration is concerned this version allows the interviewer to make efficient use of information provided for all household members at once. Evidence from the video coding support our interpretation of version-specific occurrences of time saving interviewer behaviors (1) and respondent behaviors (2):

(1) By means of analyzing the video tapes we observe quite a lot of interviewer behaviors that do not follow standard interviewer procedures: besides the fact that about 78 percent of all items are read as worded, interviewers do not administer 9.3 percent of all items to the respondent. In another 5 percent of instances, the interviewer does not read the question but instead provides a different stimulus containing the relationship of the next person to the respondent (e. g. "... and your wife?"). In 5.5 percent of all cases the interviewer does not read the question but rather confirms the answer ("... and your wife is 32 years old?"). Some incomplete questions and wrong fills are observed, too. In total we have about 22 percent of all items affected by at least one interviewer behavior that does not follow a standardized interview script.

We draw the following conclusion from these particular findings: most of these behaviors indicate kind of a short-cut. E. g., the interviewer does not read the question text as worded, he or she tries to make the conversation smoother and more suitable in terms of conversational rules. From our point of view this indicates that interviewers do not want to ask for information the respondent provided already. They do not want to behave unresponsively toward the verbal contributions of the respondent and follow conversational rules. As a side effect these behaviors are less time consuming than standard interviewer behaviors. In our

perspective, the priority therefore lies not with saving time, but with customize the question answer process to respondent behaviors not anticipated and not absorbable by the computer assisted instrument.

Figure 2. Proportion of Items Affected by Interviewer Behaviors not Following the Scripted Interview by Version

p < 0.0001
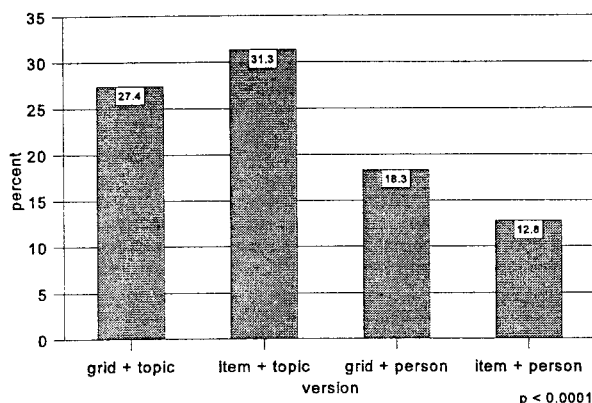
Large differences in interviewers not following the scripted interview between the four versions are to be noticed: Applying the grid based topic version to an interview results in more than twice as much such behaviors (27.4 % of the items administered are affected) than the item based person version (12.8 %) which is the standard for most studies so far. And this contributes to the time used for interviewing (p < 0.001): Items affected by a interviewer behavior not scripted in the interview take substantially less time (4.0 seconds) than the regularly administered items (6.8 seconds).

(2) Additionally an analysis of the respondents' behavior shows that the topic design leads to a higher proportion of cases where the respondent provides at least once in the household roster section the information for all persons or a group of persons at once (e. g. "We all came in the same year"; "We all have the same legal status"). The value is even higher for the item based topic version but does not differ significantly from the grid based topic version. The results displayed in the table suggests that a grid based screen design increases the proportion of respondents that provide the answers concerning all persons in the household or a group of persons at once, whereas a person based question order attracts less such shortcuts (figure 3).

It is surprising that results differ even for the two screen designs. The study was administered by telephone, the respondents not being aware of the screen design at all. The only possible explanation is based on the fact that the interviewers modify their behavior in concordance

with the screen design, stimulating the respondent differently. Accordingly, respondents, as well as interviewers, react to the screen design and the question order under the grid based topic design in a way that facilitates the interviewer respondent interaction and thus helps smoothen the interview flow.

Figure 3. Respondent Behavior (percent)

	grid + topic	item + topic	grid + person	item + person	total
respondent provides information for all persons in the household at once	38.2	45.3	29.5	10.8	29.8
no such behavior	61.8	54.7	70.5	89.2	70.1
total	100	100	100	100	100

p < 0.001

4 Discussion and Conclusion

Our results from a comparison of four versions for a small household roster indicate that interviewers as well as respondents perform more efficiently under the grid based topic condition than with the other three versions. Combining a grid based screen design and a topic based question order reduces the average duration by about 17 percent. Two thirds of this reduction can be attributed to the question order, approximately one third to the screen layout. Analyzing our time measurements as well as the data drawn from video coding we elicited two reasons for the better performance of the grid based topic version: (1) in the grid based topic version, the interviewer as well as the respondent adapt better to the logic of the question answer process, both anticipate the next question more easily and the question answer process runs more smoothly. (2) This version leads to more occurrences in which the respondent provides the information for the persons in the household faster and in about 40 % of all interviews respondents reveals the information for all household members or at least for one group at once. This particular version makes it easier for the interviewer to adapt to this situation, record the information and stimulate the respondent to give the next appropriate answer without repeating the full question text.

Our findings contribute to the discussion of how to design survey instruments for interviewer administered computer assisted data collection. Based on the results reported in this paper we can draw the conclusion that making use of grids facilitates the interviewer respondent interaction and helps speed up data collection. Our experiments on item design vs. grid design conducted in

the Survey Research Center's usability laboratory have shown that we can improve interviewer performance by providing grids (Couper et al., 1997). Moore and Moyer (1998a; 1998b) have demonstrated that we can improve interview efficiency by switching to a topic based question order, too. The present paper indicates that both parties benefit even more when combining both features.

Using grids and a topic based question order causes a greater amount of instances where the interviewer deviates from the scripted interview. From a rigid methodological point of view this might be seen as an important drawback. However, basic findings of behavior coding suggest that interviewers frequently deviate from specific interviewing procedures. "These changes often reflect adjustments made by the interviewers to meet the exigencies of the situation: to melt it more congenially with communications immediately preceding it, or to adjust to the respondent's particular situation" (Oksenberg et al., 1992: 3).

This is especially necessary when respondents do not limit their answers to the information requested by the question, but elaborate it or provide additional information. "Avoiding the appearance of not paying attention to the respondent, interviewers in this situation frequently filled in the answer themselves without asking the question, or asking it only in part" (Oksenberg et al., 1992: 5). They thus try to switch to more respondent oriented procedures to avoid looking unresponsive. A grids based screen design and a topic oriented question order supports interviewers to interact according to conversational rules and with respect to the interview situation's needs. This might be acceptable or even preferable as long as we are talking about factoid questions and as long as these interviewer behaviors do not harm data quality (e. g. leading question or probes).

What needs to be done in order to improve the computer assisted instrument in its supporting function for the interviewer respondent interaction: The data suggests that the grid based topic version leads to a specific interview flow, so that interviewer and respondent can easily adapt to it. Moore (1996; Moore/Moyer 1998a, 1998b) has shown that interviewers prefer the topic based version. By contrast, we know little about the respondents' satisfaction with that question order. Assessing their opinion about the different version is consequently an important goal. Moreover, we do not know whether this version matches the way in which information is stored in the respondents' brains. It might be, that respondents can easily adapt to this version, but that in terms of cognitive and social burden or in terms of correctness of answers it is not the right method.

Additionally it would be nice to transfer our findings from a household roster to other segments of a questionnaire. The general question would be: what happens if we use grids or form based screens more extensively? Under what conditions and circumstances does it help to improve interview efficiency and what are the limitations to this approach?

These are some unanswered questions that need to be addressed in the future research. Personally I would like to suggest a specific approach to assess these questions assuming that computer assisted instrument design is of importance to different clients: researchers, interviewers and respondents. Of course, it is important that a CAI instrument meets the researcher's needs to obtain his or her measurements and also that the question answer process be well designed for each single item. However, in my view considering the social dimension of the interviewer respondent interaction and the behaviors in between single items is also a matter of importance. If the CAI instrument disturbs the social dimension of the measurement process it might harm even data quality. So far we do not know which approach allows the best compromise between validity and reliability of the measurement process, on the one hand and a smooth short and non-embarrassing interview flow on the other hand. In order to find out to what respect a specific CAI screen design might harm data quality and how it helps save time, money and interviewer effort we need to conduct more usability studies.

To assess the questions mentioned above we do need more field experiments. Due to the fact that we want to analyze the social dimension of the interview and its effects on interviewer behavior as well as on interview duration, laboratory experiments do not meet our needs completely. Of course laboratory experiments allow a more controlled setting, reveal more detailed information about both participants, and – as a result – need smaller numbers of cases. Still, without going into the field, we will never confront our prototypes and design solutions with real pressure to maintain and facilitate the interviewer respondent interaction and the question answer process at the same time. Usability testing should therefore be seen as a joint process of laboratory experiments and field tests.

References

Baker, R. P., 1992: New Technology in Survey Research: Computer-Assisted Personal Interviewing (CAPI). In: Social Science Computer Review 10: 145-157.

Couper, M. P.; Groves, R. M.; Kosary, C., 1989: Methodological Issues in CAPI. In: American Statistical Association (Ed), Proceedings of the Section on Survey Research: 349-354.

Couper, M. P.; Fuchs, M.; Hansen, S. E.; Sparks, P., 1997: CAPI Instrument Design for the Consumer Expenditure (CE) Quarterly Interview Survey. Final Report. University of Michigan.

Couper, M. P.; Baker, R. P.; Bethlehem, J.; Clark, C. Z. F.; Martin, J.; Nicholls, W. L.; O'Reilly, J. (Eds.), 1998: Computer Assisted Survey Information Collection. New York: Wiley.

Couper, M. P.; Burt, G., 1994: Interviewer Attitudes Toward Computer-Assisted Personal Interviewing (CAPI). In: Social Science Computer Review 12: 38-54.

Fuchs, M., 1994: Umfrageforschung mit Telefon und Computer. Einführung in die computergestützte telefonische Befragung. Weinheim: Psychologie Verlags Union.

Fuchs, M., 1995: Die computergestützte telefonische Befragung. Einige Antworten auf Probleme der Umfrageforschung. In: Zeitschrift für Soziologie 24: 284-299.

Fuchs, M, 1997: Interviewer behavior and Design Issues in a CATI Survey. Keystroke Files from the Detroit Area Study. Paper presented at the Brown Bag Seminar, Survey Research Center, Ann Arbor, USA, May 1997.

Fuchs, M.; Couper, M. P.; Hansen, S. E., 2000: Technology Effects. Do CAPI Interviews Take Longer? In: Journal of Official Statistics (forthcoming).

Groves, R. M.; Mathiowetz, N. A., 1984: Computer Assisted Telephone Interviewing: Effect on Interviewers and Respondents. In: Public Opinion Quarterly 48: 356-369.

Hansen, S. E.; Fuchs, M.; Couper, M. P., 1997: CAI Instrument Usability Testing. Paper Presented at the Annual Meeting of the AAPOR, Norfolk, VA, May 1997.

Hansen, S. E.; Couper, M. P.; Fuchs, M., 1998: Usability Evaluation of the NHIS Instrument. Paper presented at the Annual Meeting of the AAPOR, St. Louis, MO, May 1998.

House, C. C., 1985: Questionnaire Design with Computer Assisted Telephone Interviewing. In: Journal of Official Statistics 1: 209-219.

House, C. C.; Nicholls, W. L., 1988: Questionnaire Design for CATI: Design Objectives and Methods. In: Groves, R. M.; Biemer, P. P.; Lyberg, L. E.; Massey, J. T.; Nicholls, W. L.; Waksberg, J. (Eds.), 1998, Telephone Survey Methodology. New York: Wiley: 421-436.

Laurie, H.; Moon, N., 1997: Converting to CAPI in a Longitudinal Panel Study. Working Papers of the ESRC Research Centre on Micro-Social Change, 97-11, Essex.

Moore, J. C., 1996: Person- vs. Topic-based Design for Computer-Assisted Household Survey Instruments. Paper presented at InterCASIC '96, International Conference on Computer-Assisted Survey Information Collection, San Antonio, TX.

Moore, J. C.; Moyer, H. L., 1998a: ACS/CATI Person-Based/Topic-Based Field Experiment – Final Report. Center for Survey Methods Research, Bureau of the Census.

Moore, J. C.; Moyer, H. L., 1998b: Questionnaire Design Effects on Interview Outcomes. Paper Presented at the Annual Meeting of the AAPOR, St. Louis, MO, May 1998.

Moyer, L. H., 1996: Which is better: Grid Listing or Grouped Questions Design for Data Collection in Establishment Surveys? In: Proceedings of the American Statistical Association, Section on Survey Research Methods, Alexandria, VA: 986-990.

Nicholls, W. L.; de Leeuw, E., 1996: Factors in Acceptance of Computer-Assisted Interviewing Methods: a Conceptual and Historic Review. Paper Presented at the Annual Meetings of the Section on Survey Research Methods of the American Statistical Association, 1996.

Nicholls, W. L.; House, C. C., 1987: Designing Questionnaires for Computer-Assisted Interviewing: a Focus on Program Correctness. In: Bureau of the Census (Ed.), 1987, Third Annual Research Conference, Proceedings, Washington, DC.

Oksenberg, L.; Beebe, T.; Blixt, S.; Cannell, C., 1992: Research on the Design and Conduct of the National Medical Expenditure Survey Interviews. Final Report. Survey Research Center, Ann Arbor, USA.

Oksenberg, L.; Cannell, C.; Blixt, S., 1996: Analysis of Interviewer and Respondent Behavior in the Household Survey. U.S. Department of Health and Human Services. AHCPR No. 96-N016.

Projektgruppe SOEP, 1998: Funktion und Design einer Ergänzungsstichprobe für das Sozio-oeconomische Panel. Diskussionspapiere des DIW, 163, Berlin.

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

Suchman, L.; Jordan, B., 1990: Interactional Troubles in Face-to-Face Survey Interviews. In: Journal of the American Statistical Association 85: 45-54.

Weeks, M. F., 1992: Computer-Assisted Survey Information Collection: A Review of CASIC Methods and Their Implications for Survey Operations. In: Journal of Official Statistics 8: 445-465.