Tangible evidence: using modern technology for recording and analysis of interviews

Beukenhorst, Dirkjan

Veröffentlichungsversion / Published Version
Konferenzbeitrag / conference paper

TANGIBLE EVIDENCE: USING MODERN TECHNOLOGY FOR RECORDING AND ANALYSIS OF INTERVIEWS

DIRKJAN BEUKENHORST[1]

1. Introduction

Two broadly defined methods exist to pre-test or evaluate a questionnaire. On the one hand we can try to detect problems by observing the interaction between interviewer and respondent during this 'conversation with a purpose'. On the other hand we can try to reconstruct the cognitive processes taking place inside the respondent in order to discover how respondents answer the questions. Both methods have their advantages and their drawbacks, and both uncover certain types of problems that the other can not detect. These methods can both be adapted to evaluating self-completion questionnaires.

Both benefit from collecting much detail in a systematic way because the cues pointing at problems often are rather subtle and can be easily overlooked. Problems not only risk being overlooked, sometimes they are 'found' without real evidence. This happens, for example, when respondents are overly stimulated to mention problematic aspects. In these cases it is as important to have detailed material to check this possibility.

Both methods, too, have a common problem: it is very difficult to strike a balance between the richness of detail giving insight in the existence and nature of a problem and on the other hand the ease and efficiency of analysis in qualitative or (semi-) quantitative ways. In this paper it will be argued that deployment of modern recording techniques in

[1] The views expressed in this paper are those of the author and do not necessarily reflect the policies of Statistics Netherlands.

combination with the use of dedicated software can contribute enormously to the quality and ease of analysis of pre-tests based on cognitive interviews and observational data.

Problems in a questionnaire can arise in the interaction process between the interviewer and the respondent. An interview is a conversation and by consequence should meet (sub-) cultural expectations concerning a conversation (Maynard a.o. 2002). It is, for example, rather unusual if someone asks "what does that mean?" to answer "whatever it means to you". This is, however, prescribed practice for interviewers. Just reading the question a second time if someone does not take his turn, is another rule interviewers follow that can annoy or confuse respondents. An interview thus violates many cultural expectations and presupposes furthermore knowledge about 'rules' peculiar to interviews (Houtkoop 2000, Stanley 1996). Not all respondents will have that knowledge: does everyone know what "tick all that apply" means or know how to state opinions on a five-point scale or how to express quantities in that way? (Miller, Integrating Socio-Cultural Factors into the Question Response Model, this volume). More trivial problems that can arise during an interview based on a questionnaire are difficulties an interviewer can have reading the question, or formulations that encourage premature answers (putting a question mark in the middle of the question followed by some illustration, for example). Such interaction problems are most of the time studied by some kind of observation of the interview process.

Many observational methods exist to test or evaluate these interact ional aspects of an interview. A simple and cheap method is having CATI-interviewers code the behaviour of respondents. Up till five different codes can be assigned in such a 'respondent behaviour coding' without distracting the interviewer too much from her main task (Burgess and Paton 1993). Codes are usually only given to a limited number of questions because of the extra burden for interviewers. Prüfer and Rexroth (1996) asked field-interviewers to indicate on the paper form if a question is problematic or not ('problem scoring'). Obvious limitation of both these methods is that the behaviour of the interviewer remains out of sight. More elaborate behaviour coding, including the behaviour of interviewers, is done by supervisors or researchers. This is applied mainly in CATI-settings where conditions to listen in are relatively easy to realise. Frequently used codes are the following: coded is whether interviewers read the question exactly as written, whether they make a slight change or whether they make a major one. For respondents is coded whether they ask for repetition of the question, whether they ask for clarification, whether they interrupt the interviewer, whether they give a qualified answer or whether they give an inadequate answer. Presser and Blair (1994) conclude this method of behaviour coding gives good results compared to cognitive interviews. The Australian Bureau of Statistics (2001) uses this method for uncovering interviewer errors and respondent fatigue.

Behaviour coding is often done ‘real-time’, that is during the interview itself, because of efficiency considerations or technical difficulties recording the interviews. This calls for simple codes, especially if the interviewer himself has to do the coding. Even with simple codes one cannot have very much confidence in the validity of the assigned codes (for an example of pre-tests with well-trained cognitive interviewers see Schechter a.o.1996). What is even worse, without a recording one lacks the possibility to check this validity.

Another reason for recording the interview, even if coding is done real-time, is that the necessarily simple codes do not give much information about the causes of problems. If we want to know more about the nature of a problem or want to remedy it, we will have to go back to the original interview sequences. In the second paragraph we will describe what level of detail we would like to record and what level is possible.

An interview is a very complex interaction process between an interviewer and a respondent. The interviewer reads the questions, the respondent answers, asks for some explanation, gives spontaneous comments etcetera. The interviewer reacts (adequately or not) by repeating the question or explaining a concept, the respondent takes his/her turn again and so on. During this response process the respondent has to accomplish different cognitive tasks which can lead to cognitive problems that can be the cause of invalid answers. The most common classification of different types of cognitive problems arising during the response process are problems with the comprehension of the question, those that arise during the retrieval of the information from memory, and problems concerning the formulation of the answer or choosing the right answer category. Because we cannot look directly inside the head of the respondent to study the ways in which he accomplishes the cognitive tasks, we have to rely on information the respondent can give about his cognitive activities. We do this by using techniques borrowed from cognitive psychology. The respondent often is asked to ‘think aloud’, that is to tell us what he is thinking while looking for an answer, or we ask him to describe the ways in which he searches his memory to find the required information, or we ask to repeat the question in his own words (‘paraphrasing’). These tasks are for many respondents far more trying than answering the questions of the questionnaire themselves. Because these tasks are so difficult measurement errors in cognitive interviews may occur more frequently than in normal interviews.

Although cognitive interviews are called ‘cognitive’, this does not mean that we can rely solely on ‘verbal information from the inside’ elicited by cognitive techniques to diagnose questionnaire problems arising in the ‘black box’. These verbal reports need to be supported by other evidence to establish the validity of problems. The complexity and subtlety of the content of a *pre-test* interview (cognitive, in-depth) is so great that based

on an audio recording conversation analysis as a technique to analyze all text, turn taking and para-linguistic behaviour of the (pre-test) interview is not sufficient because visual data are not recorded or analyzed. This omission is unfortunate because non-verbal behaviour, facial expressions and the like can tell a lot about comprehension problems, affective reluctance to answer and other response problems. For a thorough and rigorous analysis all relevant data have to be recorded in as much detail as possible. This means that cognitive techniques rely as well on observational data as the so-called observational studies proper.

Errors may arise in the reporting of cognitive process information by the respondent but also in the interpretation of this information by the researcher. The relative freedom of the interviewer during the test-interview to probe in case of suspected problems or to let pass unusual incidents as negligible, reinforces the subjectivity and complicates the analysis afterwards (Conrad and Blair 1996). This is another reason to record all data, in order to create the opportunity to invoke the judgement of other researchers. Another very important and not to be underestimated reason to record everything in detail is the fact that in a cognitive interview the normal course of the interview is often disrupted and the interview may therefore produce artificial results. If an interviewer is too persistent in probing for problems respondents may become inclined to mention more difficulties than they really encounter (Garas, Blair and Conrad 2003). In self-completion interviews interruptions by the researcher may distract the respondent so much that this in fact becomes the cause of problems. To diminish this risk of finding artificial problems careful analysis afterwards of interview sequences is necessary. This can only be done if a complete recording is at hand.

In conclusion, we need detailed observations not only for pre-tests using different types of behavioural coding but also for pre-tests using primarily cognitive or in-depth interviewing techniques. In the next section we will describe what kind of data we ideally would like to collect, depending on the data collection mode.

2. What evidence to record

In the case of computer assisted self-completion questionnaires (CASI) a pre-test can be executed by an interviewer or the pre-test can be done within the self-completion context: the respondent has to answer meta-questions or write comments on the questionnaire itself while answering the questions themselves (Beukenhorst, Giesen & de Vree 2002). In both cases we have to record everything that happens on the computer screen, the movements of the cursor, the entering of answers, the correction of answers etcetera. This can be done with a screen capture device that records all screens and what happens on

those screens. While recording it is possible, for example, to highlight the cursor or make mouse clicks audible. With the help of some hardware the recording can be done on another computer. This is convenient if test interviews are done on location where the respondent uses his own computer and does not want software installed on his pc. Another possibility is to make with a fixed camera a video of the computer screen. The video can be digitized and saved as a computer file. So-called audit trails make a file of all or a selection of keystrokes, including very precise timekeeping. This way one can reconstruct what the respondent did before and while entering answers. If an interviewer or researcher asks questions or helps the respondent with the questionnaire we should have a video and audio recording of all those events. If the respondent is asked to think aloud, an audio recording of all speech is of course indispensable. An extra camera focused on the face of the respondent can be very useful when interpreting the audio recording. If the test is especially aimed at the lay-out of the electronic questionnaire an eye-tracking device which records all eye-movements can be handy (Redline and Lankford 2001). A test of a self-completion paper questionnaire can be recorded more or less in the same way except that a camera has to be directed at the paper questionnaire and on the hands of the respondent in stead of making a screen capture and an audit trail.

Face-to-face interviews offer still more challenges for the technical equipment because the interaction between interviewer and respondent has to be recorded in detail, not only the spoken dialogue but also the facial expressions of both persons because they can point at reluctance to answer, feelings of uncertainty and other emotions which can hinder the flow of the interview. Sometimes it will be necessary to bring along a cameraman. Apart from a detailed video and audio recording an audit trail is helpful if it concerns a CAPI-interview.

In a CATI environment the only relevant data are the speech and some paralinguistic utterances, the interviewer and respondent can only hear each other after all. However, what the interviewer enters in the computerized questionnaire is important too and should be easily accessible during analysis. An audit trail of the interview should therefore also be made. A video of the faces of both respondent and interviewer can of course be useful too, although they do not influence the interaction. Such video's can only be made in a laboratory context.

A first, rather obvious, objection one could raise against such elaborate recording is the risk that the natural setting is so much changed that the interview becomes an unrealistic representation of a real life interview. This is mainly the case if one works on location and has to install all kinds of cameras and an extra pc. However, most respondents get quickly used to being observed and filmed and modern equipment is not so voluminous as in the

past. Pre-testing on location may in many cases even with cameraman and equipment be more realistic than a lab interview. Besides, as mentioned before, especially cognitive interviewing has in itself already a rather disturbing effect because the natural character and flow of the interview are much affected by the extra questions. So the cognitive interviewing research method is by itself not an unobtrusive method. Several mentioned recording techniques are indeed unobtrusive: the making of an audit trail and a screen capture remain unnoticed by respondent and interviewer alike. In practice one has to balance pragmatically on the one hand the (un)realistic situation of a test which can threaten the validity of the data and on the other hand the validity and rigor of the analysis which are highly dependent on detailed recording. Depending on personality characteristics of the test person and on the kind of interview and interviewing mode one can choose a not too disturbing technique that still records all relevant information for a valid analysis.

A second objection against such level of detail in the (recorded) data is that analysis of this rich material becomes far too complicated and time consuming for normal pre-testing practice. In the third section it will be explained that modern software makes such analysis and coding of data maybe not simple but at least feasible.

3. Coding and analysis

If we would describe in words all data gathered with the above mentioned techniques during ten interviews, we would end up with an at least one hundred pages long ‘thick description’ (Geertz 2000) that would constitute a really qualitative analysis. Doing that would take too much time and no client would read it. So we have to find ways to analyse the data in another way. We will have to reduce the amount of data without however loosing during this reduction process the possibility to easily access the complete dataset again. Sometimes data reduction is completed by writing out only the pieces of a test interview that at first hearing seem ‘relevant’. Often some other pieces can appear highly relevant during the ongoing analysis. This is common in a qualitative analysis method where description of the data and analysis intermingle. If one cannot easily retrace in the audio or video record the part of the interview that is found afterwards relevant, the analysis gets very tedious. So we need a method to mark all major parts of an interview in order to retrace them if one wants to study them in a later phase of the analysis.

Preparation for analysis and data reduction are executed by different forms of coding. Before we can code the data, however, we need to combine all data existing in separate

files into an easily accessible format. Software to do exactly that is offered on the market[2]. This software can be used for the coding and analysis of the behaviour as well. All audio and video recordings can be digitized and linked to audit trails. Synchronizing those different records facilitates finding tracks or parts of the interview. In this way one can analyze all audio, video and other records made of one test person. To compare results between different test persons marks have to be placed in the different records at relevant places in the questionnaires. Most of the time these marks will be placed at certain questions.

After having organized the material in this way, coding can start. This can be executed in two fundamental different ways. On the one hand, coding can be done like simple 'problem scoring'. If any kind of problem arises in one of the records one can check with the other data concerning that respondent if there really is a problem. If so, one enters a 'problem' in the behaviour coding analysis software. In the same way the records of other respondents are analyzed. After analyzing all records one makes a tabulation of how many problems with how many respondents arose at which points in the questionnaire. Besides this simple procedure one can try to differentiate between different types of problems, like comprehension or retrieval problems for the respondent, or reading problems for the interviewer. This more detailed analysis is facilitated because all relevant tracks can be easily accessed again. The use of rather simple codes is appropriate for the more practical testing of questionnaires where interest lies foremost in the detection and reparation of problems.

Another more 'scientific' method of qualitative coding and analysis starts from the other side, with the most detailed observation and description, and generalizes step by step from there. Behaviour is described first in basic categories: the subject who initiates the behaviour, what this subject does (a state or an event) and qualifiers that describe in more detail the behavioural act. An example may clarify this procedure.

> *At question x John (subject) looks (state) flabbergasted (qualifier) at the interviewer (qualifier)*
> *Interviewer Mary (subject) says 'uhuh' (event) looking neutral (qualifier) in another direction (qualifier)*

---

[2] Programs for organizing digital files and behaviour coding are the Observer from Noldus and Interact from Mangold. Programs designed only for behaviour coding are a.o. EthoLog, Jwatcher of the Animal Behavior Laboratory of Macquarie University and Focal32 developed by W.L. Roberts. Will Dijkstra from the Free University of Amsterdam developed Sequence for coding and analyzing interviews. It only runs on Macintosh.

This way one gets as many codes as a full description of the act would need. So some categorization or generalization is needed to reduce the amount of codes. These categories fit more than one incident while leaving out some detail.

Reduction:

At question x
John > Respondent
Looks flabbergasted > does not understand anything
Interviewer Mary > Interviewer
Says 'uhuh' + two qualifiers > reacts in a paralinguistic way without steering

Further reduction

At question X
R. gives no immediate answer because of problem with understanding
I. gives neutral prompt.

Still further:

question X in some way problematic.
Qualification: I. makes maybe a mistake because neutral prompt does not solve the lack of understanding.

Depending on the kind of software used the program can aggregate the detailed codes into more encompassing ones by recode commands or the researcher has to do this manually in consecutive rounds of coding. This type of coding which starts from a thick description and gradually evolves into a data set suited for quantitative analysis is greatly facilitated by coding software. This kind of analysis is most appropriate for more fundamental research into the response processes.

4. Conclusion

The two most common types of pre-testing a questionnaire are some form of behaviour coding and 'cognitive' research. Both depend much on scrupulous observation of the behaviour of the respondent and interviewer. This observation is greatly facilitated by modern technology to record all that behaviour. There exists software which can organize those observational data in such a way that coding and analysis become more efficient. The scientific rigour and validity of both the 'quick and dirty' method of finding and solving problems and the more fundamental method of studying the response processes are considerably increased.

References

Australian Bureau of Statistics, Pre-testing in Survey Development: An Australian Bureau of Statistics Perspective., 2001: Population Survey Development. Australian Bureau of Statistics. www.sch.abs.gov.au/SCH

Beukenhorst, D./Giesen, D./de Vree, M., 2002: Computerized versus interviewer-guided evaluation of CASI questionnaires. Paper presented at QUEST workshop 2002 Washington DC

Burgess, M.J./Paton, D., 1993: Coding of Respondent Behaviour By Interviewers To Test Questionnaire Wording. www.amstat.org/sections/srms/Proceedings/y1993

Conrad, F./Blair,J., 1996: From Impressions to Data: Increasing the Objectivity of Cognitive Interviews. www.amstat.org/sections/srms/Proceedings/y1996

Geertz, C., 2000: Local Knowlegde: Further essays in interpretive Anthropology. Basic Books (originally published 1973).

Houtkoop, H., 2000: Interaction and the Standardized Survey Interview, Cambridge: Cambridge U.P.

Maynard, D. W./Houtkoop-Steenstra,H./Schaeffer, N.C./van der Zouwen, J. (eds.), 2002: Standardization and Tacit Knowledge. New York: Wiley & Sons

Miller, K., 2004: Implications of socio-cultural factors in the question response process. ZUMA special 9

Nadra,G./Blair, J./Conrad,F., 2003: Inside the Black Box: Analysis of Interviewer-Respondent Interactions in Cognitive Interviews. Paper presented at Federal Committee on Statistical Methodology Research Conference, nov 17-19 2003, Arlington

Presser, S./Blair, J., 1994: Survey Pretesting: Do Different Methods Produce Different Results? Sociological Methodology, vol 2 (no 12): 73-104

Prüfer, P./Rexroth, M., 1996: Verfahren zur Evaluation von Survey-Fragen: Ein Überblick. ZUMA-Arbeitsbericht Nr. 96/05

Redline, C. D./Lankford,C.P., 2001: Eye-Movement Analysis: A New Tool for Evaluating the Design of Visually Administered Instruments (Paper and Web). Paper presented at the American Association of Public Opinion Research, Montreal, Canada

Roth, E./Patterson, E., 2000: Using Observational Study As A Tool For Discovery: Uncovering Cognitive And Collaborative Demands And Adaptive Strategies. Paper presented at 5th Naturalistic Decision Making Conference. www.csel.Eng.Ohio-state.edu/emily/NDMrothfinal

Schechter, S./Blair, J./Hey, J.V., 1996: Conducting Cognitive Interviews to Test Self-Administered and Telephone Surveys: Which Methods Should We Use? www.amstat.org/sections/srms/Proceedings/y1996

Stanley, J. S., 1996: Standardizing Interviewer Behavior Based on the Results of Behavior Coding, www.amstat.org/sections/srms/Proceedings/y1996

Contact

Dirkjan Beukenhorst
Statistics Netherlands
CBS
Postbus 4481
NL – 6401 CZ Heerlen
Netherlands
email: dbkt@cbs.nl