

## Testing questionnaires with cognitive methods: part 2, Thinkalouds

In the 1980s, methods from cognitive psychology were increasingly used for improving questionnaires. The common goal of these approaches is to arrive at a better understanding of processes governing the comprehension of questions, and of how responses and judgements are generated. (Tourangeau et al. 2000; Sudman et al. 1996).

Part 1 described probings (Geyer 2003), a method that provides information on specific aspects of survey questions. Part 2 presents the second major cognitive approach to questionnaire development, thinkalouds, which evaluates the overall process of generating responses. If respondents are requested to report the number of visits at the doctor within a defined time period, say the last 12 months, the results are not the same if subjects start searching their memory from the end of this period or from its beginning, they may even not have an idea at all about a certain time frame. This can be illustrated with a question from a quality of life-questionnaire. Respondents are requested to estimate the frequency of 12 complaints that have to be assessed for the last seven days before interview. Asking subjects to think aloud revealed that already at the second item the instructions were ignored and subjects reported whether they remembered having had the symptoms at any time.

Two types of thinkalouds have been developed:

In *retrospective thinkalouds* subjects are requested to reconstruct how they came to a conclusion after having produced an answer. This approach has the advantage that responses can be generated without secondary tasks interfering with respondents' ongoing thought processes. As a disadvantage ex-post rationalizations may occur (Wilson et al. 1996) if the process of reconstruction is difficult, if socially undesirable thoughts are involved, if responses are generated automatically, or if responses are strongly determined by emotional states.

In *concurrent thinkalouds* respondents are instructed to speak loudly while generating an answer. A potential disadvantage

of this approach is that respondents are expected to perform too many complex tasks simultaneously: listening to a question, recollecting relevant information, forming a judgement, giving a response on a rating scale and speaking simultaneously. Finally, upcoming thoughts may interfere with each other, and compared to the normal procedure of completing a questionnaire different responses may result (Sudman et al. 1996).

Up to now, there is no clear consensus which one is the most effective way of getting access to respondents' thoughts.

Another crucial point of thinkalouds is the relatively high demands on verbalisation aptitudes, thus becoming increasingly difficult as the educational level of the respondent decreases (Bickart & Felcher 1996). The best results should be obtained with well-educated subjects, but this is not helpful if questionnaires for population-based surveys have to be constructed.

The following example shows how instructions for thinkaloud sessions can be formulated. It was adapted from Bolton and Bronkhorst (1996); omitted sections are designated by [...]:

*"I will be reading questions to you from the survey. ... I want you to constantly think aloud while you are deciding about your answers. ... tell me everything you are thinking from the time you hear the question until you have given your final answer to the question. Thinking aloud will help us understand your thoughts about the question.*

*When I say think aloud, I mean say aloud everything that goes through your mind. You shouldn't worry if you sometimes feel that what you are thinking is not relevant to the question. I am interested in all your thoughts. ...*

*Just act as if you are alone in the room talking to yourself. If you are silent for any length of time, I will remind you to keep talking. ..."*

Responses should be tape-recorded and analysed after having completed a session. The analyses have to be

standardized by laying down rules before performing thinkalouds. This includes writing down in advance how survey questions ought to be understood and how responses have to be interpreted, i.e., when a response is within a pre-defined range and when it is beyond.

Apart from topic-specific analyses, responses can be classified by means of general coding schemes (Sudman et al. 1996). This may include the speed responses are generated with, the type of knowledge used to come to conclusions, whether responses are directly referring to a question or not, what cognitive anchors are used etc. (Bickart & Felcher 1996). In such cases, reliability measures can be obtained by calculating the consistencies between two or more raters. The average coding reliabilities reported in a previous study ranged from 0.66 to 0.87 (Bolton & Bronkhorst 1996). In order to handle thinkalouds correctly, interviewer trainings are essential. In their own study, Bolton and Bronkhorst carried out one day training sessions. This turned out as too short for ensuring sufficient reliability, thus more practise is necessary. The length of interviewer trainings is also dependent on the complexity of the tasks and on the category sys-

tem used for rating responses. The same holds for the number of trials, i.e., it should be higher if the topics to be dealt with are complicated. In the recent literature it varies between 17 and 100; with 30 to 40 sessions it should be possible to obtain sound assessments of the applicability and the problems of survey questions. It should have become clear that thinkalouds could hardly be applied by interviewers with a short training. Instead they should be performed by investigators that are acquainted with cognitive concepts and methods. Moreover, they should not be applied in the field, but rather in laboratory-like settings until a higher degree of standardisation has been achieved.

To summarise, in spite of the experimental character the usefulness of thinkalouds for improving questionnaire design has been widely acknowledged (Sudman et al. 1996; Mittag et al. 2003). In contrast to probing questions, thinkalouds are less standardised, take place in laboratory-like settings, and require highly experienced interviewers. Measures for assessing validity and reliability as well as standards for training interviewers have to be established, and more developmental work has to be done.

---

## References

- Bickart B, Felcher EM* (1996). Expanding and enhancing the use of verbal protocols in survey research. In: Schwarz N, Sudman S, eds. Answering questions. San Francisco: Jossey-Bass: 115–42.
- Bolton RM, Bronkhorst TM* (1996). Questionnaire pretesting: computer-assisted coding of concurrent protocols. In: Schwarz N, Sudman S, eds. Answering questions. San Francisco: Jossey-Bass: 37–64.
- Geyer S* (2003). Testing questionnaires with cognitive methods: part I, Probing questions. *Soz Präventiv Med* 48: 133–4.
- Mittag O, Böhmer S, Deck R, et al.* (2003). Fragen über Fragen: das „cognitive survey“ in der Fragebogenentwicklung. *Soz Präventiv Med* 48: 55–64.
- Sudman S, Bradburn NM, Schwarz N* (1996). Thinking about answers. San Francisco: Jossey-Bass.
- Tourangeau R, Rips LJ, Rasinski K* (2000). The psychology of survey response. Cambridge: Cambridge University Press.
- Wilson TD, LaFleur S, Anderson DE* (1996). The validity and consequences of verbal reports about attitudes. In: Schwarz N, Sudman S, eds. Answering questions. San Francisco: Jossey-Bass: 91–114.

---

## Address for correspondence

**Prof. Dr. Siegfried Geyer**  
**Medizinische Soziologie OE 5443**  
**Medizinische Hochschule Hannover**  
**D-30625 Hannover**  
**e-mail: geyer.siegfried@mh-hannover.de**



To access this journal online:  
<http://www.birkhauser.ch>

---