

Appendix

Manual for the Three-Step Test-Interview (TSTI)

Respondent selection

The main aim of the TSTI is to produce data on how respondents (would) complete a questionnaire ‘in real life’. Respondents, therefore, should be members of the intended respondent population, i.e., patients of hospital X if the questionnaire will be used in a survey of patients of hospital X, elderly women in state Y if the questionnaire will be used in a survey of elderly women in state Y, etc. Because every study has its own objectives and design, which determine the sampling principles and procedures that should be used, sampling as such will not be discussed in this manual. As with all pretest research, it is important to explain to potential respondents that the object of study is the quality of the questionnaire, not their skill in dealing with it. (This will also be stressed below, in the discussion of the introduction to the interview.)

Setting

Because the main aim of the TSTI is to produce data on how respondents (would) complete a questionnaire ‘in real life’, respondents should complete the questionnaire in a setting that is as similar as possible to ‘real life’ unobserved completion. There are two main options:

- (a) *‘Field interview’*: The TSTI is conducted in the setting in which, according to the respondent, ‘normal’ (unobserved) completion by the respondent would take place. This will be in the respondent’s home (at the dining table, at a desk, or on the couch) if a questionnaire is tested that will be mailed to the respondent’s home address. Or, this will be on a hospital ward if a questionnaire is tested that patients must complete during their stay in the hospital, etc. The main advantage of the ‘field interview’ is its ‘ecological validity’. Its main disadvantages are that the interview is disturbed by contingencies (telephone calls, etc.), that video recording of the interview is difficult or impossible, and that the technical quality of recorded data (e.g., the quality of the audio recording) can be compromised.

(b) *'Laboratory interview'*. The TSTI is conducted in a 'cognitive lab' or a 'pretest lab'. The obvious advantages of the lab setting, both in technical and in organizational terms, have to be weighed against the loss of 'ecological validity'.

Informed consent

Before the interview begins, informed consent should be obtained for its recording. The usual safeguards should be given regarding issues such as confidentiality of everything that is said or seen in the interview, the way quotes from transcripts will be used in publications of the research, and the security of the collected data. These safeguards apply to the responses to answers in the questionnaire that is tested as well as to the data that are specifically produced in the TSTI procedure.

Introduction to the main aim of the interview: observation

The aims and procedures of the interview must be explained to the respondent. It must be emphasized that the interview's aim is to test the questionnaire, not the respondent. It must be made very clear to the respondent that the researcher (interviewer) wants to *see for her/himself* how good or problematic the questionnaire is by *observing* how actual respondents proceed. In this context it might be mentioned that there will be two distinct phases in the interview, first, the completion of the questionnaire by the respondent and, secondly, an evaluation of what happened. Respondents can be advised that there will be ample time for them to express their opinions, ideas, and proposals in the second stage of the interview. But *"for the sake of the test, could you please abstain in the first part of the interview from any commenting on what you do?"*

Introduction to think aloud

It is useful that the respondent knows why the think aloud technique is used. Most respondents understand that their response to the questionnaire mainly occurs 'in their head' and that, therefore, an observer needs additional information. It can, however, be more difficult to instruct the respondents in performing the think aloud technique. It is essential for the TSTI to be successful (as a procedure) as well as productive (for analysis) that the instruction to the think aloud task is done well. Minor deviations in

instruction and execution might invalidate the entire interview. First, the conditions should be right. Has a good relationship between interviewer and respondent been established? Should anything been done to make the respondent feel more comfortable? The instruction itself is rather simple: respondents are asked to *say aloud what they think as they think it*. Note that the instruction is not to ‘think aloud’ because this is usually interpreted by respondents as a request to *think* in a specific way (which is confusing), whereas the request is (only) to *say aloud* what one is thinking anyway. Note also that this implies that no explanations of these thoughts are required, but just the verbalization of the thoughts themselves. Interviewers must explain this to respondents and they must also instruct them that they should not invent thoughts just to avoid silences at all cost. Respondents must be instructed to say aloud only those thoughts that come ‘naturally’ as part of the task of completing the questionnaire. Respondents differ considerably in the degree to which they are able to perform such a think aloud (or say aloud) task. Some find it very difficult. Usually such difficulties are not the result of a lack of understanding of what is requested (just saying aloud thoughts) but rather of an inability to do it. Therefore, it is usually useful to do some exercises in think aloud before starting with testing the questionnaire.

Think aloud exercises

The following are some exercises from the literature (see, e.g., Willis, 1999, and Van Someren et al., 1994).

1.

Try to visualize the place where you live, and think about how many windows there are in that place. As you count up the windows, tell me what you are seeing and thinking about. (from Willis, 1999: 4)

2.

When was the last time you had dinner in a restaurant? Please say aloud everything you think in the process of finding that date.

3.

- (a) Talk aloud while answering the following question: How often have you been to the grocery store during the last week?
- (b) Please describe to me your first visit of last week to the grocery store. Tell me in a strictly chronological order what you did during that visit after having entered the store until you left it with the things you had bought.

Feedback by the interviewer

In order to be able to learn, in these exercises, what adequate ‘think aloud’ (or rather ‘say aloud’) is, respondents must receive adequate feedback. This should reinforce good performance as well as discourage misinterpretations of what is expected from them. A distinction can be made between the following forms of feedback. (These three forms do not only apply to feedback in the exercises but also to interviewer feedback during the ‘real’ think aloud task.)

- (a) Many respondents find it difficult to verbalize all thoughts they have, because they experience this verbalizing as an interference with the thinking itself. Therefore, one important aim of interviewer feedback is to support the respondent in saying aloud as many thoughts as possible. This can be done through fairly simple remarks such as “Please continue talking” and “Please say aloud what you are thinking”. Note that these are requests to (only) say aloud what one is already thinking anyway. These are not requests to do and, thus, think something else.
- (b) Another aim of interviewer feedback is to positively reinforce good performance. Interviewers might, therefore, use expressions such as “Good, you are doing this (very) well. Please continue in this way” when respondents have, either partially or completely, successfully verbalized a thought process.
- (c) Obviously, a third aim of feedback is to alert respondents when they seem to deviate from what they are expected to do and, if possible, to correct them. The interviewer must ascertain continuously that the respondent only says aloud those thoughts that (apparently) belong to the response process. Respondents should not explain or justify these thoughts to the interviewer or comment on them. If necessary, the respondent should be reminded of the fact that ‘think (or saying) aloud’ is a

technique for making observable ‘behavior’ (thinking) that is going on, not an ‘interview’ in which accounts, opinions or other reports about self are sought. The interviewer might, thus, make comments such as “Please only say aloud what you think in order to respond to the question. Please do not comment on these thoughts just because I am here listening. Just ignore me, do as if I am not here. In the next phase of this interview you will have ample time to comment on what you have done and why you have done it the way you did”.

Interviewers might demonstrate explicitly to respondents that they are mere observers, at least during the think aloud task, by positioning themselves off the respondent’s view (e.g., by sitting behind the back of the respondent) or by adjusting their posture accordingly. Interestingly, this is a recommendation from the literature on think aloud which contradicts ‘normal’ interviewer behavior in the regular survey interview as well as in cognitive pretesting.

Step 1. Concurrent think aloud

When both respondent and interviewer agree that enough exercises have been done and that the interviewer’s feedback is understood, the TSTI proper can begin with its first step, which is the ‘think aloud’ task concurrent with the completion of the questionnaire that is tested. The object of the TSTI is the way a respondent completes the task of responding to a questionnaire as a whole, not just to isolated parts of it. Therefore, in principle, step 1 (think aloud) should cover the entire response process, i.e. from its very beginning (e.g., opening the envelope in which the questionnaire will be mailed to respondents) until the end (e.g., putting the questionnaire in the pre-paid stamped mail-back envelope, if provided). Consequently, step 2 cannot begin until this entire process is completed. This requirement follows directly from the explicit aim of the TSTI to keep the procedure as close as possible to what ‘normally’, i.e. without the test situation, would happen (i.e., to achieve a high level of ecological validity). Note that this constitutes a difference with the common practice in pretesting in which usually problems with one question are explored as widely and deeply as possible (e.g., by means of both concurrent and retrospective probes) before moving on to a following question. In

contrast, no probing is allowed in step 1 of the TSTI. Talk by the interviewer, who really is an *observer*, should be confined to giving feedback, as discussed above.

Apart from supporting the respondent in performing the think aloud task, the main other task of the interviewer is to make notes (in ‘real time’) of remarks and behaviors of the respondent. Specific notes might be made for different reasons and uses. Some might indicate specific problems with questions or, for that matter, other aspects of the response process that might be interesting for further exploration (in step 3). More importantly, notes must be made of instances in which the observer feels that a part of the thought process (i.e., of the response process) has not been properly verbalized. Notes on such ‘missing’ data are important as ‘input’ to step 2, because the main function of step 2 in the TSTI is to provide for observational data that, for whatever reason, were missed in step 1. (Later analysis of the response process and of the quality of the questionnaire does not solely depend on these notes made in real time by the interviewer. The interview will be recorded as well. But these real time notes are important because steps 2 and 3 of the interview will, to a large extent, depend on these notes.)

Step 2. Focused interview

Step 1 (think aloud) can be concluded with a thank you for having completed this difficult and onerous task of completing the questionnaire while ‘think aloud’. If the respondent shows signs of fatigue, it is important to show understanding for this. However – although tempting – this is not a good time for a break. Step 2 is aimed at ‘filling in’ bits of the thought process that were not sufficiently captured in step 1, although the process itself cannot be observed for a second time. Because step 2 can only produce respondents’ self-reports about what has happened, rather strict criteria should apply to how the interview in step 2 is conducted in order to get ‘factual’ data about how the process occurred. An important condition is that step 2 immediately follows step 1, and that no time is lost. The most important element of the role of the interviewer in this step is that he – similar to his role in step 1, but now retrospectively – must help the respondent to focus on the reporting of actual thoughts rather than of interpretations of them. This restricts the repertoire of possible questions in this step quite rigorously to only one format: “What happened next?” Note that this ‘probe’ is much more restricted

than ‘retrospective probes’ in cognitive interviews, despite its similar location in the interview process. Recall and reporting of these missing bits of data on the actual thought process of the respondent could be invited and supported by quotes from the notes that the interviewer has made. For instance, he could say “You said/did ... [quote from the notes from step 1] ... and then you did not verbalize your thoughts for a while. What were you thinking at that moment?” However, in some exceptional cases, in order to reconstruct a complicated step in a thought process (e.g., a decision about whether one agrees or disagrees with a statement which involves a series of different and incomparable criteria) it might be necessary to explore the respondent’s ideas and views about it.

The result of step 2 is a more complete reconstruction of the thought process that resulted in the choice of a response category in this specific case. The guidelines in this manual for how to proceed in the steps 1 and 2 of the TSTI are meant to ascertain that exactly this kind of descriptions of actual response processes are produced. Questionnaire designers can inspect these results in order to get insight in how their questions function in actual practice.

Step 3. Semi-structured interview

Because the completion of steps 1 and 2 might have tested the endurance of the participants, this might be the right time for a break.

Steps 1 and 2 of the TSTI are uniquely designed to produce data about actual behavior (thought processes). Both the (concurrent) think aloud procedure in step 1 and the focused interview in step 2 are described fairly extensively in this manual because they are not frequently found in this form in current practices of pretesting. The following step 3 of the TSTI, a semi-structured interview, is much less specific than the other two steps but due to its placement after the two other steps it is productive in a different way than it is in regular pretesting. Depending on the kind of questionnaire that is tested and on the specific aims of the pretesting, this part might entail different forms of qualitative (semi-structured) interviewing. Examples are:

- (a) Respondents might (be requested to) ‘explain’ their response behavior. Particularly when specific problems were encountered in responding to the questionnaire,

respondents could comment on what they think the exact nature of the problem was and why they behaved the way they did – as recorded in steps 1 and 2 of the ‘interview’. Also, respondents might suggest improvements in terms of wording of questions, layout of the questionnaire, instructions, etc. The aims and form of this interview will be similar to those of ‘respondent debriefing’. It is important to acknowledge that this kind of comments constitute ‘opinions’ and informal ‘hypotheses’, not facts, regarding the causes of problems detected in steps 1 and 2. Researchers must make their own analysis of problems associated with the questionnaire (based on observations in all interviews of the pretest).

- (b) Respondents might be asked to paraphrase questions and to comment on their definitions of terms or, in other words, some form of ‘cognitive interviewing’ might be done in this stage of the TSTI.
- (c) Respondents might be probed about the substantive issues that are covered by the questionnaire that is tested. For instance, if an alcohol consumption questionnaire is tested, respondents might be invited to describe their alcohol consumption in their own words. Or, if a scale for the measurement of attitudes towards ‘illegal aliens’ is tested, respondents might be asked to explain these attitudes to the interviewer in their own words.

Conclusion

As in any kind of interview, respondents should be asked whether they have anything to add to what has been said and whether they have liked the experience. Respondents should be encouraged to bring forward any recommendation (about the test interview or about the questionnaire that has been tested) that they find relevant. They should be thanked for their cooperation. If respondents request this, a promise should be made (and kept) that they receive a summary of the study’s findings.

Box 1. Example from alcohol consumption questionnaire

Question:

How often did you drink six or more glasses on one day, during the last six months?

Response categories ranging from and (1) 'every day' until '(8) 'never' and (9) 'don't know' (one answer permitted).

The expert review did not predict problems with the response categories, but some appeared during the TSTI.

Step 1: Concurrent think aloud

R9 marks two response categories: 3 (*3 or 4 times a week*) and 9 (*don't know*)

Step 2: Focused interview

I: So it is about three or four times a week you drink six glasses or more?

R9: [There] May also [be] a week that I don't drink ... [you] can take also a week that six [times]...

I: You also marked "don't know"

R9: Well, the one time three and the other time nothing

Step 3: Semi-structured interview

It appears that this respondent is a shift worker at Heineken brewery (!). He only drinks alcohol in weeks (one in four) in which he does not work. In such weeks he often drinks more than 6 glasses of beer a day. But that varies a lot too. In some weeks he might drink alcohol on 3 or 4 days, in other weeks 5 or more.

Conclusion

The respondent wants to express the variability of his drinking behavior in his response.

Box 2. Example from Illegal Alien scale

Item:

Illegal aliens provide the Netherlands with a valuable human resource

Step 1: Concurrent think aloud

R: *illegal aliens provide the Netherlands with a valuable human resource.....*I immediately think, in my opinion it doesn't make a difference whether you are legal or illegal, to be a valuable human resource, so, well, I have not, a straightforward opinion ... so, it is uncertain....because one can be valuable also if one is legal.

Step 2. Focused interview

R: well ... I think, I lack knowledge a bit ... I have *uncertain a valuable human resource*, well, what can I say about it?

I: well there might also be a kind of logical reasoning behind it ... that the item suggests a difference between illegal and legal people

R: oh right yes that's what I said

I: you said if this applies to everyone, why should I confirm it here for illegal people only? that kind of reasoning

R: yes yes could be yes

I: but now ... you say, you say also the facts ... you don't know the facts

R: no, you're right ... indeed I thought that the distinction between illegal or normal ... that is not clear to me ... I don't know either whether they are valuable or damaging ... those are two things

Step 3. Semi-structured interview

[In an extensive discussion of how he had answered this item, the respondent confirmed that he had clearly recognized that the authors would expect him to demonstrate his friendly attitude to illegal immigrants wherever possible, i.e. by reading this and other items as invitations to position himself politically or ideologically rather than as questions about economic or social facts. He described himself as someone who tends to "interpret everything always very literally". This discussion then moved to the following fragment.]

Item:

Illegal aliens cost the Netherlands millions of guilders each year

I: take the item *illegal aliens cost the Netherlands millions of guilders each year* .. such an item ... you take it literally ... you said that a couple of millions is not much so it's very likely that illegal aliens cost us millions

R: yes

I: do you think ... so you interprets the item as a statement about facts but is it right that you assume that the designers of the questionnaire have something else in mind?

R: yes

I: now if this questionnaire is a test in logic ... you have performed very well on the test

R: yes (laughs)

I: but if it is true that ... if this questionnaire aims at measuring ... say your benevolence regarding illegal aliens ... in that case you have been almost deliberately, deliberately

R: yes on the wrong side

I: so you mislead the researchers ... so one can say that for that reason alone the item does not measure your ... what your real opinion is about illegal aliens

R: no ... not at all

I: you say ... well it is their responsibility to ... how they interpret the responses ... now it is possible that ... if they want to draw political conclusions ... that they as you say will interpret your response incorrectly

R: yes.. but I try to attend to what ... as much as possible ... to what is printed here ... that's in principle the only thing I have.

Box 3. Example from Quality of Life measurement

Questions:

- *Do you have any trouble taking a **long** walk?*
- *Do you have any trouble taking a **short** walk outside of the house?*

Response categories: (1) 'not at all'; (2) 'a little'; (3) 'quite a bit; and (4) 'very much'.

These two questions are items from the EORTC QLQ-C30, a questionnaire that we tested at three different points in time (T1, T2, and T3) using a TSTI format.

T1. Step 1: Concurrent think aloud

R: *Do you have any trouble taking a long walk? Yeah, that must be very much ... yeah at this moment ... the shopping center ... it's 450 meter ... I cannot make it.*

T2. Step 1: Concurrent think aloud

R: *Do you have any trouble taking a long walk? Yes, with a long one ... I cannot walk kilometers.*
[data omitted]

R: *Do you have any trouble taking a short walk outside of the house? No, a short walk ... I mean ... 500 meter that way ... that's nothing ... I do it without any problem.*

T3. Step 1: Concurrent think aloud

R: *Do you have any trouble taking a long walk? Yes, with a long one ... I cannot do it ... but I walk too fast, it's my own fault ... I haven't tried it yet but I think quite a bit... I can make it to the shopping center though*
[data omitted]

R: *Do you have any trouble taking a short walk outside of the house? Well, a little.*

Conclusion

The respondent's definition of a 'long' walk is a walk that is difficult to make. The response to the question about the long walk is, therefore, always *quite a bit* or *very much*. But sometimes this answer refers to a walk to the shopping center (450 meter) that on another occasion might be considered a 'short' walk (because, at that moment, it can be walked without much trouble).