

Problems with cognitive testing in multicultural surveys

Gunilla Davidsson, Birgit Henningson

Statistics Sweden

1. Cognitive testing

We would like to continue the discussion from the last QUEST meeting about problems in large surveys conducted in many countries (most often developing countries) by international organizations (most often). Since our last QUEST-meeting in Mannheim Birgit and Gunilla have continued to think about and to be reminded of problems with using either a literally translated or a harmonized questionnaire in many countries. Last autumn the Measurement Lab at Statistics Sweden was contacted by WHO for an expert review and a cognitive testing of their new health survey among elderly (55+) people (SAGE). The cognitive testing were conducted in South Africa and in Vietnam, kindly assisted by Margie Schneider and Duong Huy Luong. Gunilla is the Swedish delegate in Washington Group where the work with presenting a few “global” questions on capturing disability now have come to a point where cognitive testing ought to be an important task. (See Kristen Miller’s paper). The work with harmonized surveys within Eurostat is also continuing. Birgit has had a new course on ‘How to conduct a Statistical Survey’ in Kosovo.

1.1. SAGE - The cognitive follow-up interviews to assess the respondents’ impression of the interview and cognitive reactions to specific questions

Our prime task was the expert review, but we were also asked to conduct a cognitive test as cheap as possible. Since we have old contacts in Vietnam and South Africa we suggested to WHO to ask them if they were interested in taking part – and they were. We had very little time and, as already said, almost no money to spend. That’s why we conducted a “budget version” using our standard cognitive testing as a model. A very short written introduction (see Annex) and the draft SAGE-questionnaires were sent to the participating parties.

Below is a short summary of the results:

Most of the 10 interviews - household + personal questionnaires - lasted 2 to 2,5 hour, but the time consumption ranged from 1 hour 40 min to 6 hours. Two additional parts were not included in the testing: one part based on vignettes and finally a number of biomarkers. Even with these two parts left out the general impression of the questionnaires were unanimous in both South Africa and in Vietnam – the interview was too long! Respondents added sentences like “I’m too old to cope” or “the concentration goes away” or “I was even feeling sleepy and exhausted”. A 74 years old lady said “the problem is that the questions ask about things we cannot remember well, the memory is not too good when you grow older”. The interviewers in South Africa wrote that they found the questionnaires too long for old people. Even if the respondents were cooperative the length of the questionnaires put a lot of strain on them. The respondents asked for shorter or longer breaks during the interviews. On the whole the questionnaires would be more interesting if made shorter. One of the interviewers also pointed out that elderly people do not feel comfortable being asked about sexual matters, particularly when asked by someone younger.

Problematic questions were discovered, which needed to be discussed more. We presented suggestions based on our experience on what tends to cause poor data quality or be a heavy part of the response burden. Our suggestions did not take into consideration the subject matter motives

since we had no knowledge of how the results were going to be used.

Examples of questions found to be difficult to answer by the respondents:

Income - four of the five respondents in South Africa had problems differentiating between household expenditure and sources of income (household questionnaire) and own expenditure and sources of income (individual questionnaire). If they knew something, it was on how much money was spent in the household, but they could not account for sources of income - “I don’t know money, I did not go to school”. The layout also “invited” to a mix-up of figures per month and per year, which cannot be discovered afterwards by logical controls. The interviewers have noted down in the margins that they had to explain the questions, which is “dangerous” – interviewers tend to give different explanations.

Household expenditure: The interviewers found some examples of contradictions between the answers to questions about household income and questions about household expenditure. In Vietnam the respondents could give an answer for the total amount (last 4 weeks), but could not split the spending.

Health status: Three sections all dealt with the respondent’s health in using different existing batteries of questions with almost similar questions, which caused a lot of trouble and confusion. The phrasing of the existing battery Physical Activity (GPAQ) entails a risk of poor quality among the very old and those who have mobility problems or are bedridden.

Delicate questions: Some respondents were very hesitant to answer questions they found very personal, like going to the toilet, sex interest, cognition etc. and the question about visiting people of other race.

Cultural questions: One question asked about the opinion on health care services in the country. In South Africa respondents compared the services between private and public (“white” and “black”) hospitals when answering and gave opinions about public hospitals.

Exact numbers/times: Only the “youngest” of the respondents had no problems with answering the questions asking for exact numbers/times. The other respondents found it hard to remember all these things. To questions like ‘How many years have you lived in this dwelling’ the respondent gave answers like “more than 20 years”, “about 10 years”, “not sure”, “many”, etc. Questions were even left unanswered because the respondent could not count, “many times”. Sometimes the problem to find the right number made the respondents change their answers over and over.

Unfortunately we had neither time nor money to a proper testing the way we would have liked to perform it. But still it was an interesting and rewarding experience.

1.2. Cognitive testing in Kosovo

The mission to the Statistical Office of Kosovo (SOK) and the Department of Agricultural, Forestry and Rural Development (DAFRD) was carried out during 11 – 22 March 2002. The mission purpose was to teach how a statistical survey is done.

The purpose of the course is to present and carry out all phases of a survey in the proper order. In particular the course demonstrates how the different phases of a survey are connected. In order to present the full picture, we usually do a small scale survey during the course. The statistical theory and practical problems are discussed in the light of this actual survey. In the case of Kosovo we

already had a survey to work with, the Survey of Agricultural Households 2001. Consequently we had to do things the other way round and start from the questionnaire and make the variable/indicator listing and tabulation plan afterwards. We were lucky that one of the facilitators had worked in Kosovo and analysed the tables in 2001 (without any prior tabulation plan). Accordingly we knew a lot about which questions had been working and which had not. In the course we ended up with a new draft questionnaire and a draft tabulation plan for the next round of the Survey of Agricultural Households. The new draft questionnaire was tested on seven respondents belonging to the target population. Several of the participants had their doubts about this new draft questionnaire, but after the cognitive testing and the revisions they were all convinced about the advantages of the new questionnaire. Each interview with the draft questionnaire took about 20 - 30 minutes – compared to 2001 when the average was two hours.

1.3. Cognitive testing in Cambodia

Statistics Sweden provided technical assistance to the National Institute of Statistics (NIS) in Cambodia before the implementation of the Socio-Economic Survey (CSES) which took place in 2003-2004. A pilot survey was carried out in May and made use of the questionnaire and corresponding manual that had been designed and worked out by a consultant from Statistics Sweden in cooperation with the NIS and PMATU staff. In order to ensure that the methods to be used in the field work by the interviewers were following what is considered as best practices, it was decided that training directed to core staff and supervisors would be conducted with the assistance by experts from Statistics Sweden. To advise NIS and PMATU core staff in best practices in interviewing. About 25-30 persons attended during two weeks training. During this training of the future trainers we wanted to test the questionnaire cognitively. A trip to a village not far from Phnom Penh was organized. We were 33 persons in four minibuses. We worked in groups with one interviewer and one observer. One test interview in the morning and one in the afternoon were conducted.

The cognitive testing gave us a lot of experience about the questionnaire. After this practice in the field the participants agreed on the need for improvements. And they also got a better understanding of how it is to work as an interviewer in their country.

2. Cognitive problems in harmonized surveys in Europe

More and more surveys launched by Eurostat are conducted under a Commission Regulation , which means that all 25 member states implement a harmonized survey within the same time frame. Eurostat is responsible for a reference questionnaire in English and the countries for the harmonization according to certain rules. This isn't an easy process! The first and most important task for Eurostat and the Member states is of course to decide if suggested variables/indicators are able to harmonize or not. Translation into almost 25 languages is another big problem. Eurostat has spent a lot of time and money on a very ambitious project about how this will be done. The new translation procedure will be used for the first time in launching a European Health Survey 2007. As an assistance aid in the implementation the member states have a special web place for discussion etc. But the problem is that no one is 'responsible' for questions and answers – let us give you an example: One member state is having problem with a question about limitation according to health problems, especially with the expression “activities people usually do”. Another member state gave this answer:

Point 3 of the note expresses concern about the wording "activities people usually do". I agree that there are problems with this wording. From what I can ascertain, this form of wording is out of date (it was used, for example, in the xxxx Census of 1991) and has been replaced by wording along the lines of "activities you usually do". The latter form is used in both the latest xxxx Census and the Health Survey of xxxxxx. It is also adopted by many other countries [see p55 of Eurostat publication (2003) 'Health in Europe Results from 1997-2000 Surveys']. By asking the respondent to compare their activities carried out during the last six months with activities they usually do, the respondent needs only to consider personal experience and they do not need to conceptualise their experience within a framework of what other people can/cannot do. In short, the potential problems which xxxx alludes to can easily be avoided by a simple change to the wording.

To us the first person is getting the advise to change the wording into quite another question which has nothing to do with "what people usually do". If for example these two member states (and maybe some others after reading the conversation) use the definition above and the rest of them the question proposed by the reference questionnaire – then what about the comparison and harmonization?

The reference questionnaire is made up by (what is supposed to be) validated questions and well defined concepts (if possible). SF36-questions about psychological distress and well-being are often used, but they do not have any defined concepts. A lot of countries in Europe (even UK!) ran into big trouble when translating the true meaning of questions like: "Did you feel full of pep? Did you have a lot of energy?" and "Have you felt so down in the dump, nothing could cheer you up? Have you felt down-hearted and depressed?" and "Do you feel worn out? Do you feel tired?"

To many of us we don't have any corresponding expressions to all of these American expressions in our languages and will partly get two questions more or less identical in wording since we don't get guidance by any definitions.

Our third example is the very often used question on self-perceived health (How is your health in general?), where the middle answering category (of 5) is 'fair'. But one member state wanted to have a better description or explanation of 'fair' since in that country the common perception of 'fair' was more like 'quite good' than 'rather bad'. The solution Eurostat offered was a change to 'neither bad nor good', which we find even worse –is it really possible to have a health that is neither bad nor good?

The last example is more about 'working manner' in Eurostat. Each country has the travel expenditures charged by Eurostat for one participant. (Additional participants are charged by the member states). This means that most often the participants have some more or less leading position at home but very seldom experience of question design, which our last example shows.

In one survey a question was needed on what possibilities people have to meet unforeseen economic problems. The first draft question was if the respondent could raise 1000 Euro in two weeks if needed. But the member states dismissed the question because of incomparability – 1000 Euro is not much money in Luxemburg but a fortune in Latvia! Next draft said: could the respondent raise enough money to repair or buy a new fridge if necessary. One delegate said that in her country it was very cheap to get a fridge repaired but very expensive to buy a new one. Another delegate told us that in his country it was just the opposite if you at all could find someone willing to repair your fridge. Eurostat's solution to the problem was to suggest voting between the two concepts – fixed sum of money or replacement of some valuable goods. In the Nordic countries we already since many years have questions about meeting unforeseen economic problems in our national surveys and tried to 'promote' those. Our idea is to ask if the respondent can come up with a sum which corresponds to a low monthly salary for a "blue collar work". This

sum will of course differ from country to country but still it will be comparable. It's been hard to try to convince Eurostat and all the delegates that this is the best way to get comparability.

These were just four of a number of problem we face daily when trying to harmonize surveys in Europe.

3. Questions about disability, Washington City Group

Washington City Group is now on the edge of presenting a common set of questions about disability to be adopted in censuses and surveys around the world. Hopefully as many countries as possible will be able to do some cognitive testing before promoting the questions.

At first we had in our minds also to talk a little about problems with cognitive testing in connection with the work of the Washington Group, but luckily Kristen Miller knows much more about that.

4. Discussion

Our ambition with this paper is once more to strike a blow for more lobbying about the importance of cognitive testing in multicultural surveys.

Appendix

Instructions for Testing SAGE Questionnaires

- A** Explain the purpose of the interview to the respondent and emphasis that it is the questionnaire and not the respondent that is to be tested and the answers will be treated confidentially.
- B** Introduce how you will work. First you will ask all questions in the original questionnaire. Afterwards there will be some more questions about the interview itself, about problems (if any) and about specific questions .
- C** While you are making the interview:
- Note the starting time of the interview
 - Note down in your own questionnaire the respondent's comments and behaviour (being annoyed by a question, asking for clarification, hesitating or changing an answer, pondering a very long time before answering a question, etc.)
 - Note the time when the interview is completed
- D** After the original interview you introduce **the follow-up interview**. Emphasise again that it is the questionnaire and not the respondent that is being tested and that any problems he/she reports about the questions or any comments he/she has are of the greatest value to us. "There are no right or wrong answers. What we want to know is how you **personally** experienced the questions and if you had any problems."

Throughout this follow-up interview you strive at noting **the respondent's answers as literally as possible** avoiding your own interpretations or vocabulary.

- Ask what she/he thought of the preparatory information and note the answer.
- Ask about the respondent's general impression of the questionnaire. Write down the answers.
- Ask if there were questions she/he had problems with or questions she/he wants to comment on. Any sensitive questions? Difficult questions or anything else?
- Ask follow-up questions based on your notes during the original interview (i.e. your observations).
- **Specific questions:** Ask the respondent if and how she/he managed to differentiate household expenditure and sources of household income (the household questionnaire) from the person's own expenditure and sources of income (the individual questionnaire).
Ask if she/he had any problems with answering with exact number/times in several questions. If yes, which questions and how did the respondent "reason" before coming up with an answer.
Ask if the respondent found some questions to be similar. Which ones? Did she/he give the same answer to those questions? If changed: Why?
Ask if the respondent had trouble understanding some of the words the questions about chronic conditions and diseases. Did she/he give an answer even if she/he wasn't quite sure about what was asked about?
- When the follow-up interview is finished note the time and thank the respondent for her/his kindness in helping us constructing a better questionnaire.

- E** Summarize your comments!