

Cognitive Pre-tests in Spanish National Statistics Institute

Miguel Ángel Martínez Vidal

Spanish National Statistics Institute

José Luis Padilla Garcia

University of Granada, Spain

1. Introduction

Until recently, questionnaire evaluation procedures applied by the Spanish National Statistics Institute (NSI) were based on traditional pilot studies. In most of these studies interest focused on frame problems, non contact or refusals. Not structured reports of interviewers involved in these pilots inform about problems with questions.

Interest in incorporating new evaluation procedures to the traditional practices of the NSI has increased in recent years. Among the main reasons for this are: 1) the greater sensitivity of professionals and users to so-called ‘non-sample errors’; 2) the growing involvement of the NSI in European projects promoted by EUROSTAT; 3) the introduction of new survey application methods based on new technology, 4) the growing relationship between NSI professionals and researchers in various academic centres and 5) the increasing number of demographic surveys with, sometimes, sensitive topics join with the need of using different data collection methods. An example of this relationship is the collaboration between a team of professionals from the NSI and researchers from the University of Granada (Spain). In recent years, this collaboration has led to work on the evaluation of various surveys using cognitive procedures. The long-term aim of this work is to compile guides of procedures which will enable us to incorporate cognitive procedure evaluation in NSI projects.

In order to describe the process of designing and evaluation questionnaires in the NSI, we need to say a little about the structure and function of the organization. Different departments of the NSI are responsible for different surveys, depending on their objective and content. The so-called ‘promoters’ in the various departments are responsible for establishing survey objectives and once the objectives have been fixed, the questionnaire project is sent to the Data Collection Unit, which normally takes responsibility for designing and evaluating the questionnaire and for the whole application process. When data has been collected the “promoters” are in charge of analysis and dissemination.

In this presentation we will try to illustrate the work which the NSI has carried out over the last two years to apply cognitive procedures in questionnaire evaluation. We will describe three studies in order to show the application of three different cognitive procedures, namely the cognitive interview, behaviour coding and focus groups. We should stress, however, that various procedures were involved in the evaluation of the questionnaires in each study. The studies in question were carried out by professionals of the NSI Data Collection Unit working in collaboration with researchers from the University of Granada (Spain).

2. Study 1: Improving the measurement of actual hours worked and usual hours worked by cognitive interviews.

In 2005, the Spanish Labour Force Survey (LFS) data collection procedure has been completely renewed because of the next reasons:

- New variables demanded by Eurostat and the use of the wave approach (distinction between structural and not structural variables)
- New data collection method. CATI for the second to sixth interview. CAPI for the first interview.
- The Spanish LFS questionnaire came from the earliest 90's and it has been partially modified many times. So a complete revision was needed to assure that variables are better measure and to make it more respondent and interviewer friendly.

In order to fill these objectives various departments of the NSI worked together to make proposals of questions. The process involved evaluating the established question formulation, proposing ways in which the formulation should be modified and then evaluating the modification proposals. The results of the process were incorporated into the new version of the questionnaire which was applied in the first term of this year. Users' opinions were taken into account before and after the cognitive procedures were carried out.

Traditional pilot survey and focused groups were undertaken but in this paper we are going to present only about the cognitive interview procedure in order to improve measurement of 'usual hours worked' and 'actual hours worked' in the Spanish LFS. The objective was to evaluate the proposed questions for measuring these variables, using the 'question-answering process' model as the basis of the evaluation (Tourangeau & Rasinksi, 1988). The study design included population groups of particular interest to the Spanish LFS: 'immigrants', 'young people aged between 16 and 30 who have just entered the labour market', 'housewives', 'unemployed people under 44', 'employed people between 30 and 65' and 'people over 65'. In addition, participants were divided into two subgroups, consisting of direct informants and proxies.

2.1. Method

2.1.1. Participants

The group of participants in the cognitive interviews for questions relating to hours worked consisted of 55 people, of whom 21 were men and 34 were women. Age range of male participants was between 18 and 57, with an average age of 34.48; age range of female participants was between 21 and 72, with an average age of 38.47. Table 1 shows distribution of participants in each subgroup according to sex and informant role (direct or indirect).

Table 1. Participants according to sex and informant type

		Sex		Overall
		Men	Women	Total
Direct	Immigrants	6	8	14
	Young people	3	7	10
	Employed	6	5	11
Indirect	Housewives		5	5
	Older people		5	5
	Employed	5	5	10

2.1.2. Target questions and sources of error

‘Target questions’ refers to the proposed questions, whose functional performance was to be evaluated by means of the cognitive procedures. The two proposed questions for measuring hours worked were as follows: Q1. ‘*How many hours a week do you/does (s)he usually work in this job?*’ and Q2. ‘*In the course of last week, how many hours did you/(s)he actually work in this job?*’

The study team analysed the proposed questions with a view to identifying possible sources of error, taking into account the categories of the ‘question-answering process.’ Subsequently, the most suitable type of probe was determined and protocol questions were formulated for the cognitive interview (Willis, DeMaio & Harris-Kojetin, 1999). In Table 2 we can see the anticipated problem, probe and protocol question.

Table 2. Anticipated problems and questions for interview protocols

Target question	Problem	Probe Technique	Protocol questions
Q.1	Differences in the interpretation of the words ‘a week’	Elaboration	How many days did you have in mind when calculating the number of hours you work a week?
	Differences in the meaning of the ‘key’ concept: usually	Meaning oriented probes	What does the expression ‘ <i>usually work</i> ’ mean for you?
	Cognitive difficulty in estimating the number of hours	Elaboration	How did you calculate the number of ‘hours a week’?
Q.2	Differences in the meaning of the ‘key’ concept: ‘actually’	Meaning oriented probes	What do you understand by hours actually worked?
	Cognitive difficulty in estimating the number of hours	Elaboration	How did you calculate the hours you actually worked last week?

The protocol included other questions designed to determine if participants’ calculations of work hours included time spent travelling to work and time spent eating. In addition, formulation of some of the questions in the protocol was adapted to suit the indirect informants.

2.1.3. Procedure

The cognitive interviews were carried out according to a retrospective design. In a first step, participants answered a shortened version of the LFS target questions and questions concerning other complementary variables included in the analysis. Subsequently, they took part in the cognitive interview. These were conducted by interviewers with long experience in carrying out in-depth interviews. At the beginning of the session, participants were informed of the objectives of the study. Each participant received 20 Euros for his/her collaboration.

2.2. Results

Analysis of the cognitive interview was carried out on the basis of session transcriptions as well as audio and video recordings. Two coders analysed the transcriptions using a specially designed Coding Sheet.

Following, we will to summarise the analysis of the direct informants in the cognitive interview, with regard to the target question for the variable ‘hours usually worked’.

- Comprehension of ‘hours usually worked’. Principal Findings: 1) Most participants interpret the expression ‘hours usually worked’ to mean ‘timetable at work’, and use ‘daily’ as a time reference; 2) the group of immigrants bring concepts such as ‘stipulated work’ or ‘agreed work’ to their interpretation of the expression.

- Calculation of ‘hours usually worked’. Principal Findings: 1) Most interviewees acknowledge they performed an arithmetical operation multiplying the number of daily hours by the number of working days. Again, this arithmetic supports the idea that respondents take ‘daily’ as a time reference to calculate the total number of hours a week; 2) Most participants did not include time spent travelling to work, but on the other hand did include time spent eating, referred to as the lunch break.
- Comprehension of ‘hours actually worked’. Principal Findings: 1) Interviewees share the idea of ‘duration of time at work’; only a few participants interpret the expression to mean ‘effective time worked’ or ‘time spent carrying out’ the tasks or functions of the job concerned.
- Calculation of hours actually worked: Principal Findings: 1) Most interviewees describe a calculation process based on multiplying the daily timetable by the total of days worked; 2) There are also references to memory processes concerning terms stipulated in work agreements or contracts.

With regard to the analysis of interviews with indirect informants, this revealed greater variability in the comprehension of key concepts in the target questions. In particular, the group of people over 65 gave unsuitable answers and performed erroneous calculation processes. For example, in this group it was common to interpret the ‘hours actually worked’ in terms of the content of the work or of the fixed or temporary nature of the contract.

3. Study 2: Improving questions on salary incomes by means of behaviour coding

A team of professionals from the NSI and of University of Granada carried out this study in response to a call from EUROSTAT. The general objective was to study the most appropriate formats for questions concerning salaries into the LFS questionnaire and to determine the effects of a wide range of variables, such as interview method and direct versus indirect informant. Evaluation of the proposals was carried out by traditional pilot studies and various cognitive procedures. The present summary is limited to information obtained by means of behaviour coding.

3.1. Method

3.1.1. Participants

The behaviour of 40 people during the interview was coded. Participants included 16 men and 24 women, with an age range between 18 and 65. 60% of participants had at least completed primary studies. In addition to these demographic features, participants were selected according to criteria related to their employment activity. Many were salaried workers employed who had been working at least in the month before the probes were carried out. An alternative condition was that respondents cohabited in their usual place of residence with at least one salaried person who had been working at least in the month preceding the probes to research the question-answer process in proxy respondents. Participation was voluntary and interviewees received 20 Euros for their collaboration.

3.1.2. Target questions and sources of error

The target questions posed by interviewers as they appeared in the questionnaire were as follows (1) ‘*Last month*, what was your/his/her MONTHLY gross income, i.e. before deducting tax, social security payments, pension fund instalments etc?’; and (2) ‘Now I am going to ask you about your/his/her *net income*. *Last month*, what was your/his/her MONTHLY net income, i.e. after deducting tax, social security payments, pension fund instalments etc?’

In addition, the questions had different answer formats. Half of the participants answered according to an open format; in other words, they were required to state the exact salary. By contrast, the other half answered with a closed format, by which they were shown a range of salary bands and were asked to select the band which most accurately reflected the income. The Figure 1 shows possible sources of error which could affect answers to these questions.

TARGET QUESTION PROBLEM
Differences in the meaning of the concept ‘salary income’
Differences in the meaning of the concepts ‘net’ and ‘gross’
Differences in the interpretation of the temporal period (monthly or weekly)
Estimation process in the open question
Estimation process in the closed question

Figure 1. Possible sources of error

3.1.3. Procedure

A retrospective design was used to carry out the cognitive pre-test. In a first step, participants responded to a questionnaire with the target questions and other question referring to different variables. Subsequently they took part in the cognitive interview, which was conducted by interviewers with long experience of carrying out cognitive interviews. Application of the questionnaire was video-recorded. At the end of the session, each interview was behaviour-coded by two previously trained observers.

Coding was based on an adaptation of the scheme present by Snijkers (2002). Table 3 shows the categories employed and gives a brief explanation of each one.

Table 3: Behaviour coding scheme

INDICATORS	EXPLANATION
Behaviour of interviewer	
1. Reading of question	Exact/Slight changes/Significant changes
Behaviour of interviewee	
1. Interruption	The interviewee interrupts the reading of the question to give his/her answer
2. Clarification	The interviewee asks for repetition or clarification or makes comments indicating doubt
3. Appropriate answer	The interviewee gives answers appropriate to the question objective
4. Qualified answer	The interviewee gives answers appropriate to the question objective but makes comments indicating doubt
5. Inappropriate answer	The interviewee gives answers which are inappropriate to the question objective
6. Don’t know	The interview answers ‘Don’t know’ or equivalent

After the interviews, the coding of the observers was compared to determine the level of coincidence and clarify possible discrepancies. An agreement level of approximately 90% was achieved regarding the identification of problems.

3.2. Results

Behaviour coding was undertaken separately for direct informants and indirect informants. Similarly, separate coding was carried out in accordance with the format of questions Q1 and Q2 (opened or closed questions). In the interests of brevity, I shall restrict myself to the results of interviews with direct informants, including both the direct and indirect formats. Table 4 shows coding frequency for questions 1 and 2 for direct informants responding in open format (‘exact salary’):

Table 4: Coding frequency for Q1 and Q2 with direct informants and open format

Questions	Indicator frequency					
	Interruption	Clarification	Appropriate answer	Qualified answer	Inappropriate answer	Don't know
Q.1	0	0	4/7	0	2/7	1/7
Q. 2	0	0	5/7	0	1/7	1/7

Although the frequency of appropriate answers is high, there are also inadequate answers and ‘Don’t Know’ answers for both Q1 and Q 2.

Table 5 shows coding frequency for questions 1 and 2 for direct informants responding in closed format (‘salary bands’):

Table 5: Coding frequency for Q1 and Q2 with direct informants and closed format

Questions	Indicator frequency					
	Interruption	Clarification	Appropriate answer	Qualified answer	Inappropriate answer	Don't know
Q.1	0	2/9	5/9	1/9	1/9	0
Q. 2	0	0	7/9	2/9	0	0

The frequencies reveal qualified answers for question Q2 (‘net income’) and the interviewees’ request for clarification in Q1.

4. Study 3: Improving a questionnaire on health and sexual habits by means of focus groups.

This study demonstrates how focus groups may be used in the process of compiling and evaluating NSI questionnaires. The questionnaire concerned was compiled in order to carry out a survey of health and sexual habits related to HIV. This originated as an initiative of the Ministry of Health and the NSI was given responsibility for carrying out the survey.

The decision to evaluate the questionnaire through cognitive procedures was based on two considerations: 1) the delicate nature of many of the questions (initiation of sexual relations,

infidelity, consumption of alcohol and drugs etc.) and 2) the application of the questionnaire by means of the CASI system (Computer Assisted-Self Interview). Various cognitive pre-tests as well as traditional pilot studies were undertaken. The focus groups were viewed as a particularly appropriate way of determining the frame of reference or perspective from which respondents would answer the questionnaire (Snijkers (2002)). It was hoped to reveal their perceptions of aspects such as the objective of the survey, their 'role' as respondents, the credibility of the organizations carrying out the survey, etc. In addition, it was hoped that the narrative discourse of the participants in the focus groups would shed light on interpretations of 'key' concepts in the survey and indicate the actual experience of participants.

4.1. Method

4.1.1. Participants

Five focus groups were held in May and June 2003 with a total of 49 participants. Criteria for selection were established in accordance with the questionnaire filters and variables relevant to the survey objectives, i.e. sex and sexual orientation (heterosexuals and homosexuals). The groups were homogenous with regard to these variables. One of the groups consisted of 'homosexuals', whose participation was contacted through an association for gay rights.

4.1.2. Procedure

The discussion sessions took place after respondents had answered the questionnaire, which was done using the CASI system. The sessions were held in different rooms from those in which the questionnaire had been completed, and were led by a moderator with long experience in conducting focus groups. The moderator initiated the session by explaining the objectives of the study, the proposed session plan and the fact that it was to be recorded for analysis. Subsequently the participants were presented and the first discussion topic was introduced. The moderator had access to a script in which the topics to be discussed were listed. Approximate duration of the sessions was 50 minutes. Transcriptions of the recording were used for analysis of the discussion, which was carried out by technicians with experience in narrative discourse following guidelines established by the project leaders.

4.2. Results

Presentation of results are limited to the most significant aspects in relation to the questionnaire content and survey method.

4.2.1. Questionnaire content

In general, the participants evaluated the questionnaire positively, and usually justified this positive evaluation in terms of the importance of the topic as a public health problem. However, one of the most significant results is the appearance of two 'respondent typologies', which tended to approach the questionnaire from different perspectives. On the one hand, 'less informed' respondents (older women) had a very positive attitude to the questionnaire and to the survey in general. On the other hand, 'highly informed' respondents (young women and homosexuals) adopted a more critical attitude, indicating that the treatment of the topic showed lack of depth and that the objectives were not enough clear.

The appearance of these typologies may also explain the differences in comprehension of the key concepts 'sexual relations' and 'partner'. For example, the discourse of the 'highly informed' participants revealed a wider and more varied concept of 'partner' than that of the 'less informed'.

A similar pattern appeared with respect to the meaning of the expression 'sexual relations', with the discourse of the more highly informed respondents including a greater variety of practices which might involve risk of HIV infection.

4.2.2. Survey Method

The great majority of the participants were highly satisfied with the application method. All respondents highlighted the 'intimate' atmosphere in which such sensitive questions were posed. Several respondents acknowledged that they would not have taken part or would have lied if the questionnaire had been applied by means of personal interview. Moreover, the portable computer itself was considered an attractive novelty by most participants.

5. Conclusions

We hope this summarised paper of the three studies gives some idea of how cognitive procedures have been applied to the evaluation of questionnaires in the NSI over the last three years. The information obtained by means of these procedures has improved the measures obtained through questionnaires.

The results of cognitive interviews have led to important changes in the Spanish LFS questionnaire. For example, for the first time a question has been included to determine the hours agreed or reflected in the work agreement ('*How many hours have you agreed to work or are stipulated in your contract or work agreement?*'). In addition, the evidence that interviewees did not differentiate between the concepts of 'hours *usually* worked' and 'hours *actually* worked' led to a simplification of the question concerning this last variable, and the word 'actually' was eliminated. Similarly, as a result of data from the interviews, the question now includes the condition: 'Please do not take time spent eating into account.'

The formulation of the questions concerning salary incomes has likewise been improved thanks to results obtained from the different cognitive pre-tests. Although it cannot be said that the behaviour coding results have led to a proposal for question modification, the results revealed problems which needed to be clarified by means of cognitive interviews. For example, the question concerning gross salary in the indirect informant group produced the highest frequency of possible problem indicators.

Evidence obtained in the focus group analysis indicated the need for substantial modification proposals and resolved doubts on important aspects of the survey of health and sexual habits. Notable among the modification proposals was the inclusion of an introduction to one of the questionnaire sections clarifying the sense in which the expression 'sexual relations' is intended. The introduction extended the concept of 'sexual relations', attempting to establish it as inclusive of all sexual practices which might pose a risk of HIV infection, independently of the types of partner involved. Among the various doubts which were clarified by the focus groups, I would emphasize the following: the acceptability of the interview method (CASI), the proposal to reinforce the role of the public organizations carrying out the survey, and the need to highlight the importance of participants' collaboration, given the magnitude of the issue.

Finally, from the methodological point of view, the collaboration between the NSI and researchers from the University of Granada is making it possible to carry out research into many different aspects of cognitive procedure application. Examples include the most suitable conditions for use of verbal reports on answer processes, examination of predictions derived from 'optimization-satisfaction' answer-process models, analysis of determinants in partial non-

answers and so on. In addition, we are also working on the compilation of procedural guides with recommendations on how to analyse and evaluate the results of the cognitive procedures described.

References

Snijkers, G., 2002, Cognitive Laboratory Experience: On Pre-testing Computerised Questionnaires and Data Quality. *Ph.D thesis. Utrecht University, Utrecht, and Statistics Netherlands, Heerlen.*

Tourangeau, R., Rasinksi, K.A., 1988: Cognitive Processes underlying Context Effects in Attitude Measurement. *Psychological Bulletin, Vol. 103, No. 3*, pp. 299-314.

Willis, G. B., DeMaio, Th. J. and Harris-Kojetin, B., 1999: Is the Bandwagon headed to the Methodological Promised Land? Evaluating the Validity of Cognitive interviewing Techniques. In: Sirken et al. (eds.), *Cognition and Survey Research* pp. 133-153. Wiley, New York.