



City Research Online

City, University of London Institutional Repository

Citation: Fitzgerald, R., Widdop, S., Gray, M. & Collins, D. (2011). Identifying sources of error in cross-national questionnaires: Application of an error source typology to cognitive interview data. *Journal of Official Statistics*, 27(4), pp. 569-599.

This is the unspecified version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/1160/>

Link to published version:

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Identifying Sources of Error in Cross-national Questionnaires: Application of an Error Source Typology to Cognitive Interview Data

Rory Fitzgerald¹, Sally Widdop¹, Michelle Gray¹, and Debbie Collins²

This article evaluates a Cross National Error Source Typology that was developed as a tool for making cross-national questionnaire design more effective. Cross-national questionnaire design has a number of potential error sources that are either not present or are less common in single nation studies. Tools that help to identify these error sources better inform the survey researcher when improving a source questionnaire that serves as the basis for translation. This article outlines the theoretical and practical development of the typology and evaluates an attempt to apply it to cross-national cognitive interviewing findings from the European Social Survey.

Key words: Cross-national research; questionnaire design; cognitive interviewing; Cross National Error Source Typology (CNEST); translation; European Social Survey.

1. The Cross National Error Source Typology

The Cross National Error Source Typology (CNEST) was developed as part of the European Social Survey (ESS) questionnaire design process. The ESS is a large scale cross-national survey that has included over 34 European countries to date. It insists on transparency during all stages of design, execution and archiving. This includes publishing known deviations on its website, for example highlighting a translation problem with a specific question. Reviewing such deviations, issues arising from previous rounds of ESS questionnaire design and pretesting and evaluations of the process and data (see Harkness et al. 2003; Saris and Gallhofer 2007), a pattern of problems began to emerge. This information was used to develop the CNEST. It was anticipated that this typology would provide cross-national questionnaire designers with a clear basis on which to try to reduce and avoid measurement error by basing remedial action on the underlying cause of the error found or produced during the design phase. The CNEST was then applied during

¹ Centre for Comparative Social Surveys, New Social Sciences Building, City University London, Northampton Square, London EC1V 0HB, UK. Email: ess@city.ac.uk

² Questionnaire Development and Testing Hub, National Centre for Social Research, 35 Northampton Square, London EC1V 0AX, UK. Email: Debbie.Collins@natcen.ac.uk

Acknowledgments: We would like to acknowledge the contribution made by the UK Economic and Social Research Council who kindly co-funded the British fieldwork referred to in this project. This study was part of a wider initiative organised by the CSDI workgroup to develop best practice guidelines for undertaking cognitive interviewing in cross-national research. We are particularly grateful for the input of Kristen Miller (NCHS), Rachel Caspar (RTI), Martin Dimov (CSD), Janet Harkness (University of Lincoln, Nebraska), Cátia Nunes (Institute of Social Sciences – University of Lisbon), José-Luis Padilla (University of Granada), Peter Pruefer (GESIS), Nicole Schöbi (SIDOS), Alisú Schoua-Glusberg (RSS) and Stephanie Willson (NCHS).

Table 1. The Cross National Error Source Typology (CNEST)

Error classification	Description	Error found in:	
		Source language testing	Non source language testing
1) Poor source question design	All or part of the source question has been poorly designed, resulting in measurement error	Always	1 or more countries
2) Translation problems. . .	Errors occur in translation, resulting in a loss of functional equivalence		
(a) resulting from translator error	Errors stem from the translation process (i.e., a translator making a mistake or selecting an inappropriate word or phrase) rather than from features of the source question that make translation difficult	Never	1 or more countries
(b) resulting from source question design	Features of the source question, such as use of vague quantifiers to describe answer scale points, are difficult/impossible to translate in a way that preserves functional equivalence	Occasionally	1 or more countries
3) Cultural portability	The concept being measured does not exist in all countries. Or the concept exists but in a form that prevents the proposed measurement approach from being used (i.e., you can't simply write a better question or improve the translation). For example, to measure religiosity a different question might be needed in a Christian country compared to a Muslim one.	Less likely*	1 or more countries

Note: *Cultural portability problems should be less likely in the source country (language). This is because the question designers should have a greater familiarity with this culture. However, this is not always the case and is complicated further by within-country diversity in cultural practices.

Table 2. Illustration of the application of the CNEST

Error type	Example Q from ESS	Problems
1) Poor source question design	“Using this card, if you add up the income from all sources, which letter describes your household’s total net income? If you don’t know the exact figure, please give an estimate. Use the part of the card that you know best: weekly, monthly or annual income.” ESS Round 1	Problems with source question were replicated in translation. High item non-response (Widdop 2007) probably reflecting topic sensitivity and poor question design. Anecdotal evidence provided by interviewers: respondents don’t understand term “net income”; question is hard to answer for large households.
2) Translation problems		
a) Resulting from translator error	Simple error “Please tell me how important you think each of these things should be in deciding whether someone born, brought up and living outside [country] should be able to come and live here. Please use this card. How important should it be for them to . . . be wealthy?” ESS Round 1 Functionally equivalent translation not realised “Using this card, please say how much you agree or disagree with each of the following statements. . . If people who have come to live here commit <i>any</i> crime, they should be made to leave” ESS Round 1	Simple error: “be wealthy” was inadvertently translated as “être en bonne santé” (be healthy) in the French questionnaire. Functionally equivalent translation not realised: In Denmark “commit any crime” was translated as “begår nogen som helst form for lovovertrædelse”, which roughly back translates as “breach of the law”. In a preceding question the word “forbrydelse” was used for “crime” which could have been appropriate here too. The statement used was interpreted as being less serious than committing a crime and resulted in very few Danes agreeing with the statement compared to citizens of other countries.

Table 2. Continued

Error type	Example Q from ESS	Problems
2) Translation problems		
(b) Resulting from source question design	“Which option on this card best describes whether or not you can discuss personal issues such as feelings, beliefs or experiences with <i>any</i> of these friends?” I can discuss all personal issues I can discuss almost all personal issues I can discuss most personal issues I can discuss some personal issues I can discuss a few personal issues I can discuss no personal issues ESS Round 4	Translation of the vague quantifiers (all, almost all, most, some, a few, no) used in the answer scale was difficult in some languages. This meant that in some countries the answer scale was not functionally equivalent with the British English version (Behr et al. 2008).
3) Cultural portability	“In choosing your regular GP, do you feel that you have. . . READ OUT. . . enough choice . . . or, not enough choice?” ESS Round 2	Asking questions about a “GP” caused problems since the concept of a “general practitioner family doctor” did not exist in the same way in all European countries.

cross-national cognitive interviewing and its usefulness evaluated. The CNEST is summarised in Table 1.

Table 2 provides examples of questions included in earlier rounds of the ESS that were subsequently found to be problematic and highlights where they fit into the typology. The discovery of such problems often came after the questions had been fielded in the survey. These examples helped to shape the CNEST.

1.1. Comparison With Other Error Source Typologies

Similar typologies have been developed independently by U.S. researchers when analysing cognitive interviews to test questionnaire translations in cross-cultural settings. Levin et al. (2009, p.14) report that “Willis and his colleagues identified three categories of questionnaire problems: *translation* problems [. . .] *culture-specific* problems [. . .] and *general* problems” across a range of cognitive interviewing projects (see Forsyth et al. 2007; Kudela et al. 2006; Willis et al. 2005a and Willis et al. 2005b). Levin et al. (2009) successfully applied the three categories identified by Willis et al. (2005b) in their evaluation of a Spanish-language translation of a dietary questionnaire. A similar four-category system was designed by Willis et al. in 2007 specifically for an investigation examining the translation and subsequent understanding of a tobacco use survey in Spanish and Asian languages. The intention

in using the scheme “was to provide a heuristic device for organising detailed examples [from behaviour coding] into a set of general categories of results” (Willis et al. 2007, p.1081). Other schemes, which included categories for “linguistic problems” and “questionnaire design problems” related to culture-specific language use have been applied by Carrasco (2003) and Schoua-Glusberg (2006). In addition, Goerman and Caspar (2007) identified a range of categories when classifying the problems found in designing a bilingual census form.

The main difference between these typologies and the CNEST is the development background. The other typologies all originated from analysis of findings from behaviour coding or cognitive interviewing. In contrast, the CNEST was developed on the basis of experience of cross-national questionnaire design, translation assessment, feedback from data users and quantitative assessments of question quality, and without prior knowledge of the other typologies. However, despite these different development backgrounds there are significant similarities between the CNEST and the other schemes, especially that developed by Willis et al. (2007). Table 3 compares these two typologies.

Category 1 is the same in both typologies. A key difference is how translation problems (Category 2) are identified. The CNEST differentiates between mistakes during the translation process (with a further distinction between simple mistakes and lack of functional equivalence) (2a) and difficulties with regard to finding a functionally equivalent translation because of the source question itself (2b). These distinctions are important since they point to whether remedial action in the source questionnaire is necessary. If one country simply makes a mistake, there is little that can be done to improve the source question. However if pretesting reveals that translators are struggling to find a good translation then further guidance can be provided. Category 2b identifies instances where one or more features of the source question makes achieving a functionally equivalent translation difficult or even impossible. In this case, the source questionnaire itself would need amendment.

Another key difference in the typologies is reflected in the interpretation of the category labelled as “problems of cultural adaptation” by Willis et al. (2007) and as “cultural portability” in the CNEST. The scheme developed by Willis et al. (2007) reflects cross-cultural work mostly within a single country where adaptation may face fewer barriers than on a cross-national survey being implemented in a large number of countries and languages. Willis’ “cultural” category focuses on adapting elements of the source questionnaire for use in other cultures. So for example, it might be necessary to add information to clarify the response task required or to modify the question wording for a specific cultural group, depending on the problems experienced by respondents in it (see Willis et al. 2007). In contrast, the “cultural portability” category in the CNEST takes into account that different concepts may not exist in all target countries and that even if they do exist, the source question formulation may need to be altered to accommodate cross-national differences.

In order to assess whether the CNEST was comprehensive in terms of classifying cross-national questionnaire problems, it was applied to findings from cognitive interviews conducted on a set of ESS questions. Since the range of questionnaire problems found in cognitive interviewing is likely, in part, to reflect earlier questionnaire development phases, it is worth summarising the question design and pretesting procedures that were in

Table 3. Comparison of two error source typologies

Cross National Error Source Typology	Willis et al. (2007)
1) <i>Source question issues</i> – all or part of the source question has been poorly designed.	1) <i>Generic design problems</i> – problems with the original survey question.
2a) <i>Translation problems resulting from translator error</i> . Translated questions are not functionally equivalent to the source question, resulting from avoidable errors and/or non-realisation of functionally equivalent translation (even though there was no cultural or linguistic reason to suggest it was not possible). Distinction between simple error and nonrealisation.	2) <i>Translation problems</i> – “items that failed to express the meaning as intended to participants due to defects in the translation process” (Willis et al. 2007, p. 1081).
2b) <i>Translation problems resulting from source question design</i> . The question appears to work reasonably well in the source questionnaire but has features in its design which make translation difficult, leading to measurement problems.	No distinct category.
3) <i>Cultural portability</i> – the concept being measured does not exist in all countries or does exist but in a form that prevents the proposed measurement approach from being used.	3) <i>Problems of cultural adaptation</i> – it is necessary to adapt the question or instructions to ensure suitability across different cultures.

use on the ESS at the time and the extent of their application prior to this cognitive interviewing.

The ESS questionnaire is developed in the source language (British English) and subsequently translated into every target language with the aim of achieving equivalent meanings (Harkness 2007). This is a common method used in large-scale cross-national surveys (e.g., The Survey of Health, Ageing and Retirement in Europe (SHARE) and The International Social Survey Programme (ISSP) also use this approach). Some projects develop questionnaires in multiple languages throughout the questionnaire design process (see Goerman and Caspar 2007). However, on a study like the ESS with at least 25 language versions in each round, this would be impractical on cost grounds.

Almost all ESS questions are closed and administered in the same basic format in all countries, essentially an “Ask the Same Question” (ASQ) approach (Harkness 2007). The success of this approach depends on the suitability of the content of the source questionnaire, the formulation of its questions and the quality of the eventual translations (ibid). A small number of concepts require country-specific questions that are later coded to a standard classification, e.g., education. (In the ESS, countries develop one or more country-specific questions to capture the highest level of education respondents have

Stage	Process	Description
1	Proposals for new question modules, identifying key concepts, definitions and measurement aims	Question design template
2	Proposals reviewed by multi-disciplinary specialist panel	Expert review
3	Survey quality predictor program (SQP)*	Program used to predict reliability and validity of new items
4	Cognitive interviewing	Cognitive interviewing in 6 test countries
5	Revised proposals submitted in the light of Stages 2 and 3	Revised question design template submitted
6	ESS National Coordinators consulted on substantive and translation issues	Comments fed into process via email and face-to-face meeting
7	Split ballot MTMM experiments developed	Tests of alternative questions wording
8	Large-scale, two-nation quantitative pilot run containing MTMM experiments	In the UK and Bulgaria (in Round 4)
9	Analysis of pilot data – including examination of item nonresponse, scalability, factor structure, correlations, analysis of the MTMM experiments and assessment of translation	Conducted by Question designers and CCT members
10	Further specialist review of the proposed questions in the light of Stage 9	Expert review
11	Further consultation with the National Coordinators	Comments fed into process via email and face-to-face meeting
12	Final source questionnaire is produced and translated according to a committee approach following the ESS TRAPD† procedures	Source questionnaire finalised in British English then translation takes place

Fig. 1. ESS Questionnaire Development Process (in 2008). *SQP is usually only used once. †Translation, Review, Adjudication, Pre-testing and Documentation (TRAPD) translation procedures are used on the ESS for translation and assessment (Harkness 2007). “Pretesting” in this context refers only to pretesting translations in a specific country to ensure that optimal translations are used.

successfully completed. Data from all countries are then recoded into a standard cross-national ISCED coding scheme.) All concepts and dimensions in the ESS are ultimately represented in an integrated dataset in an identical format for all countries, facilitating ease of comparison. The ESS uses a range of techniques to develop and test new questions (Figure 1), striving to achieve equivalence (Jowell and Eva 2009). The ESS has had some success developing questionnaires that measure attitudinal constructs equivalently across countries (Harkness et al. 2003; Saris and Gallhofer 2007, p. 71). However, findings from a program of Multitrait-Multimethod (MTMM) evaluations also suggest large differences in measurement quality between countries (ibid), with a recommendation that these differences be corrected prior to commencing comparative analysis (ibid). Consequently, in 2008 cognitive interviewing was used (for the first time) as part of the ESS questionnaire development process to try to reduce such differences in advance of fieldwork.

The aim of cognitive interviewing is to provide evidence on whether the survey questions under scrutiny are meeting their measurement objectives in the sense that respondents are able to provide meaningful answers of the quality required by the question designer (Collins 2003; Beatty 2004). This evidence helps the question designer make

decisions about whether and how to revise questions and is a particularly useful tool for the designers of cross-cultural and cross-national survey questions, as it can also assist with the early detection of translation problems (Nápoles-Springer et al. 2006; Forsyth et al. 2007; Willis and Zahnd 2007; Willis et al. 2007; Levin et al. 2009).

The ESS test questions considered in this article had already been subject to a series of different drafting stages following expert review (Stages 1–3, Figure 1). However, they had not been the subject of review by the ESS National Coordinators from each country (Stage 6), nor had they been tested in a large-scale pilot (Stage 8).

2. Cognitive Interviewing Methodology

Six ESS participating countries – Bulgaria, Germany, Great Britain,³ Portugal, Spain and Switzerland – volunteered to undertake a minimum of ten cognitive interviews each, which was felt to be the minimum number necessary to allow for some within- as well as between-country analysis and was a realistic number to ask countries to undertake.

Protocols covering sampling and recruitment, translation of the test survey questions, interviewing procedures and analysis methods were developed to try to ensure consistency in the implementation of cognitive interviewing across participating countries (see Miller et al. 2008 for more information). The working language of the study was English.

2.1. Sample Design

Cognitive interviewing methods are qualitative in nature (see, for example Gerber 1999; Willson and Miller 2005) and this study adopted a qualitative approach to sampling. A stratified purposive sampling approach that reflected the target population (Patton 2002) for the ESS was designed and implemented to recruit respondents in each country. Details of the composition of the characteristics of those interviewed are shown in Table 4.

It is possible that larger samples, in the form of more interviews per country and/or more countries, would have increased the likelihood of errors being identified, improved the chances of identifying problems that may be more prevalent in certain population subgroups and reduced the chances of between-country differences that are due to the composition of a small sample occurring. However, limited resources restricted the number of participating countries and the number of interviews that could be conducted, a common constraint in many cross-national surveys.

2.2. Translation of Test Questions

This study attempted to adopt key stages from the committee approach to translation used on the ESS involving Translation, Review and Adjudication (Harkness 2007) when translating the test question from the source language (British English) into other languages. The full TRAPD process involves three types of people – translators, the reviewer and the adjudicator (see Harkness 2007). This team-based approach avoids the risk of problems experienced when it is just one person's responsibility to produce the translated questionnaire(s) (Harkness et al. 2003, p.40). Countries were encouraged to

³ In this project, cognitive interviewing took place in Great Britain. However the ESS is a UK-wide survey.

Table 4. Composition of Respondents Interviewed by Country

Country	Number of interviews achieved	Sex		Age (in years)			Education proxy (age left continuous full-time education)	
		Men	Women	18–29	30–69	70 +	Aged 16 or younger	Aged 17 or older
Bulgaria	10	5	5	2	4	4	4	6
Germany	10	5	5	2	4	4	4	6
Great Britain	29	15	14	8	9	12	9	20
Portugal	8	3	5	3	3	2	3	5
Spain	18	10	8	6	6	6	9	9
Switzerland	15	9	6	5	4	6	8	7
Total	90	47	43	26	30	34	37	53

adhere to the full procedures; however, in some countries (notably Portugal and Bulgaria) this was not fully possible due to limited resources. In these cases country representatives either translated the test questionnaires alone or with the help of one other person.

2.3. Interviewing Procedures

The two most commonly used cognitive interviewing techniques are “think aloud” (respondents verbalise their thoughts as they attempt to answer the survey question) and “probing” (respondents are asked scripted or nonscripted (spontaneous) questions, or a combination of the two, about how they attempted to answer the survey question). Both methods have their strengths and weaknesses (Willis 2004; Willis et al. 1999) and can be combined (DeMaio and Landreth 2004; Willis 2005). For this study the main focus was respondents’ understanding of the test questions, and for this reason it was felt that probing was a more appropriate technique than think aloud (Beatty 2004; Willis 2004; Beatty and Willis 2007).

A criticism of cognitive interviewing methods is that they lack standardisation (Tucker 1997; Conrad and Blair 1996) and that this is particularly problematic when undertaking cross-cultural or cross-national cognitive testing (Kudela et al. 2004; Miller et al. 2005). However, standardisation can limit the ability of the interviewer to collect sufficient information to understand why the respondent answered the test question as he/she did (Gerber and Wellens 1997; Beatty 2004; Willis 2004; Willis 2005). Our approach was to provide interviewers with a summary of the measurement aims of the ESS test questions and an indication of the key issues to be explored in the cognitive interview (a set of standardised parameters), but not to provide scripted (standardised) probes for each test question. The probes used for each test question can be found in the Appendix. The summary information (measurement aims for each test question and the key issues to be explored in the cognitive interview, e.g., comprehension of particular terms) enabled interviewers to develop “spontaneous” probes to investigate issues emerging from respondents’ narratives about how they went about answering the test question, providing richer data on the sources of error (Beatty 2004; Willis 2005; Beatty and Willis 2007). This approach also avoided the need to translate probes equivalently, which can be problematic (for example, see Levin et al. 2009). Furthermore, to reduce the risks of bias resulting from directive, interviewer-driven probing (Conrad and Blair 2009), interviewers started with a general probe such as “How did you come up with that answer?” before turning to more specific issues such as “What did term X mean to you when you answered the question?” Such general probes can capture “cultural” issues that shape the respondent’s understanding of the survey question and are therefore useful in providing context for interpretation (Gerber and Wellens 1997).

The adoption of a spontaneous probing approach relies on interviewers’ being sufficiently skilled to “notice potential problems and choose appropriate follow up probes” (Beatty and Willis 2007, p. 297). Countries involved in this study had a varying range of experience, and training was provided to those with limited prior experience in cognitive interviewing methods along the lines of that proposed by Willis (2005). The training focused on nondirective probing, and included participants’ practising these skills in the presence of a skilled interviewer who gave feedback. Moreover, all participating countries were briefed on the measurement aims of the test questions, were involved in the

development of the shared interviewing protocol, and had regular contact with the coordinators during fieldwork. Interviews were undertaken in six languages indigenous to the countries participating in the study.

Quality control is important in cognitive interviewing studies involving numerous research teams in data collection and analysis, such as cross-national studies, to ensure consistency in approach (see, for example, Willis 2005a; Goerman 2006). For this study, direct assessment of interviews conducted in non-English languages was not possible as the native English-speaking coordinating team members were not fluent in any of the other five languages covered by the study, and this is acknowledged as a weakness. Instead within the resource constraints of this study, central quality control measures consisted of: the use of agreed protocols; initial training of interviewers in the interviewing protocols; encouraging interviewers to chart and review their initial interviews after early attempts to complete the data summarisation process (see below); and to share these with the coordinators, who provided feedback on the level and richness of the detail provided.

2.4. Analysis Methods

When analysing the cognitive interview data we were attempting to unearth evidence of question-performance in terms of any problems that the question structure, content and survey context may have caused respondents. We also looked for evidence that the questions were meeting the measurement aims. The data are qualitative in nature, reflecting respondents' accounts of their thought processes, understanding of the survey response task presented, and the factors that shape their responses (Collins 2007; Knafl et al. 2007). A rigorous, critical-realistic qualitative approach to the analysis of the cognitive interview data was adopted (see, for example Guba and Lincoln 1994; Miles and Huberman 1994; Madill et al. 2000).

An analysis protocol was developed to ensure consistency and transparency across countries and included the following key features:

- (1) Audio recording of interviews
- (2) Verbatim transcription or detailed notes of each interview made by the interviewer in the language in which the interview was conducted
- (3) Data reduction using a standard template, completed in English to ensure that data interrogation was accessible to all members of the research team
- (4) A committee approach to analysis and interpretation to ensure that key points were not lost or misconstrued in the process of translation of the analysis findings into English or through summarisation.

A data reduction template was devised on the basis of the framework data management process. Primarily used to organise qualitative interview data, this matrix-based analytical tool facilitates rigorous and transparent data management and allows the analyst freedom to conduct across and within case interpretation (Ritchie and Spencer 1994; Spencer et al. 2003). The charts were set up with the column headings reflecting the issues to be explored, with each row representing one respondent. Each cell contained a summary of the key points whilst retaining enough information to clearly communicate the findings, as well as references back to the original sources (recordings and notes) (see Figure 2). The completed matrix was reviewed by all members of the research team prior to the detailed analysis.

Question 1: CARD 1. Using this card please tell me which of the three statements on this card, about how much working people pay in tax, you agree with most? CODE ONE ANSWER ONLY 1. Higher earners should pay a greater proportion in tax than lower earners 2. Everyone should pay the same proportion of their earnings in tax 3. High and low earners should pay exactly the same amount in tax 6. (None of these) 7. (Don't know)			
Serial number, sex, age, level of education, country	Q1: RECORD RESPONDENT'S ANSWER: How respondent came up with their answer? AND/OR What they were thinking? How the respondent understood each answer option – what did each one mean to them?	Q1: Whether the statement the respondent chose reflects the tax system in their country? Whether the respondent understood the difference between the three options?	Q1: Probe: Who the respondent thought "working people" are. Probe: What the respondent understood by "high earners" (give examples). Probe: What the respondent understood by "low earners" (give examples).
UKNatCenJM01, Male, 60 years, Left school aged 14	Answer: Higher earners should pay a greater proportion in tax. Respondent said that he thought this was "fairer": explaining "Low earners have still got to keep their families going, just like others, but they have less money. People who earn more should pay more. Simple". Respondent understood greater proportion to be according to the amount that they earn.	Summary of the raw data	Column content depends on question complexity but could be questions on specific issues, such as comprehension, recall etc or focused on areas explored in probing.

Fig. 2. Illustration of framework matrix used for analysis of cognitive interview data

Framework facilitates comprehensive analysis, being capable of identifying all the answer strategies used and problems encountered by respondents that were captured during interviews. In addition, it assists in determining the significance of problems, inferring whether such problems are spurious (e.g., where the respondent went off at a tangent), can be seen as random error or are evidence of a more systematic error, stemming from the structural form and implementation of the question or questionnaire. More details about Framework and the analysis steps undertaken in this project can be found in Fitzgerald et al. (2009).

Once the key findings had been compiled and summarised, each country was asked to confirm whether the findings were an accurate and complete summary of what happened in the interviews. This process was an important way of "validating" the research findings (Enerstvedt 1989; Conrad and Blair 2009).

3. Performance of the Cross National Error Source Typology

In this section we illustrate the application of the CNEST to the cognitive interview data collected for the ESS, with reference to the four types of error CNEST defines: source question problems; translation problems resulting from translator error; translation

problems resulting from source question design; and cultural portability. Whilst we provide one example for each error source, the CNEST can and did identify multiple sources of error for individual survey questions. The questions being tested were a selection of some of the most problematic items from the Ageism and Welfare modules developed for the 2008 ESS.

3.1. Source Question Problems

Source question problems will emerge if all or part of the source question has been poorly designed resulting in measurement error. Such difficulties were found for a question about age and status.

Test Question

Some people say that certain age groups have a high or low status, while other people say there is no real difference. By status I mean the position or standing an age group has in society. I am going to ask you how high or low you think most people in [country] would say different age groups are in terms of their status.

CARD Firstly, using this card, please tell me how you think most people in [country] would rate the status of those aged 15-29

Extremely low status

0

1

2

3

4

5

6

7

8

9

Extremely high status

10

(Don't know)

88

Question aim: to see which age groups respondents would see as having the highest status and which they would see as having the lowest. Respondents were effectively being asked to rank the groups across three questions each focused on a different group.

This question was asked along with two others in a battery where respondents were asked to rate the status of “those aged between 15 and 29,” then those “aged between 30 and 70” and finally “those aged over 70.” Analysis of the cognitive interview data for this question identified several problems: the concept of status was problematic, as was the task of answering about the opinion of “most people.” However, here we focus on the difficulty respondents from all countries had with the age group in the question: a problem with the source question found in all countries.

The interview protocol was explicitly designed to explore respondents’ abilities to generalise across each age group since prior expert review had suggested that this might be a problem. The task of generalising across the age band 15–29 was found to be especially challenging. This resulted in a tendency for respondents in all countries to fixate on one end of the age band (normally the lower end) rather than its entirety, although in Bulgaria this tendency was less widespread. Figure 3 summarises the range of problems identified.

The problems posed by considering such a wide and diverse age group were succinctly summarised by one Portuguese respondent:

the ages of 15 and 29 years old encompass very different people and to attribute a status to this group as a whole is not very correct.

	Bulgaria	Germany	Great Britain	Portugal	Spain	Switzerland
Age range too broad*		⊙	⊙	⊙	⊙	⊙
Wanted to split age band into 2 or more*			⊙			
Hard to generalise*		⊙	⊙	⊙	⊙	⊙
Focused on younger end when answering	⊙	⊙	⊙	⊙	⊙	⊙

Fig. 3. Summary of range of problems found with source question. *Explicit difficulty mentioned by respondents

Respondents of all ages fixated on the younger end of the 15–29 age group when answering and commented on the difficulties experienced in generalising across this group. In summary, the detailed analysis and verification of cognitive interview findings revealed consistent findings across countries which suggested there was an inherent problem with the source question that had been replicated through translation.

3.1.1. Finding a Solution

Three alternative age band descriptors were considered: (1) eliminate the younger ages (15–19) from the age band to encourage respondents to disregard younger teenagers; (2) refer only to “adults under 30”; and (3) be less specific and refer to “people/adults in their 20s”. The recommendation made was to refer to “people in their 20s” because the cognitive interviews indicated respondents in all countries saw this group as more homogenous. Changes were made to the test question based on the results of cognitive interviewing as well as evidence from other aspects of pretesting.

ESS Round 4 Final Question

I'm now going to ask you some questions about the social status* that people in different age groups have in society. By social status I mean prestige, social standing or position in society; I do not mean participation in social groups or activities.

CARD I'm interested in how you think most people in [country] view the status of people in their 20s, people in their 40s and people over 70. Using this card please tell me where most people would place the status of... **READ OUT...**

		<i>Extremely low status</i><i>Extremely high status</i>										
E5	...people in their 20s?	00	01	02	03	04	05	06	07	08	09	10

* Annotation: "Social Status": in the sense of prestige, social standing and position in society.

3.2. Translation Problems Resulting from Translator Error

In the CNEST “translation errors” occur when translated questions are not functionally equivalent to source questions. They may result from genuine human error or from the fact that the decision as to which translation of a phrase or word to use is suboptimal, resulting in a loss of functional equivalence.

3.2.1. Simple Error

We identified several simple translation errors resulting from human error. An example is shown below.

Test Question

CARD Using this card please tell me, on a scale of 0-10, how efficiently you think the income tax authorities in [country] carry out their work. 0 means extremely inefficiently, and 10 means extremely efficiently.

Extremely inefficiently

012345678910

Extremely efficiently

(Don't know)
88

Question aim: to examine respondent perceptions of how 'efficiently' the income tax authorities do their job.

The Portuguese team inadvertently omitted the word “income” in their translation of the phrase “income tax authorities.” This meant that respondents in Portugal were asked about “tax authorities” in general rather than “income tax authorities” specifically. Ideally such errors would be detected during the translation process but in practice we know this does not always happen, so introducing cognitive interviewing earlier can help. All of the other participating countries’ translations included the word “income.”

This omission impacted on the answers obtained, with only one Portuguese respondent specifically referring to the “income tax authorities.” Others spoke more generally of the “tax authorities” or referred to services and duties associated with tax authorities in general. In the other European countries (with the exception of Bulgaria, where the cognitive interviewing data in the charts was unclear on this distinction), some respondents referred to the income tax authorities specifically in their answers, which was interpreted as evidence that this specific authority was being considered. Since a broader range of tax authorities was considered by Portuguese respondents, equivalence with the others countries was compromised.

3.2.2. Functionally Equivalent Translation Not Realised

The next example illustrates a suboptimal choice of translation.

Test Question

CARD 3 The next few questions are about welfare and public services in [country]. Using this card please tell me how much you agree or disagree that “the system of public services in [country] prevents large-scale poverty”.

1. Agree strongly
2. Agree
3. Neither agree nor disagree
4. Disagree
5. Disagree strongly
6. (Don’t know)

Question aim: to examine respondent perceptions about the impact the system of public services in their country was having in terms of preventing large-scale poverty.

The question asked respondents to consider the “system of public services” in their country. Respondents’ understanding of this term was probed during the cognitive interview. Figure 4 presents the two components of such a system that were mentioned by respondents in all countries.

The two components in Figure 4 outline the broad similarities found across almost all countries in terms of understanding of “the system of public services.” However, in Germany some respondents did not appear to understand the term at all. A closer examination of the German translation for “the system of public services” (“öffentliche

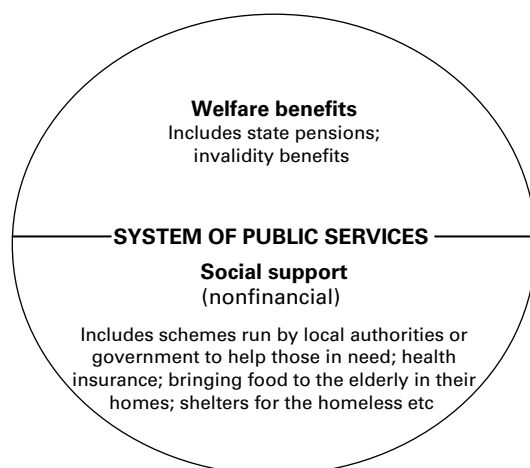


Fig. 4. Components that make up the system of public services according to the respondent

Dienstleistungen”), which literally back translates as “public services,” would appear to be equivalent to the source questionnaire. However, it was either incomprehensible to German respondents or associated exclusively with the “public social security benefits/achievements” (“öffentliche Sozialleistungen”) dimension. Respondents tended to question what was meant by the term with one highly educated respondent answering “don’t know” and commenting that she could not imagine what “öffentliche Dienstleistungen” might mean. The German researchers concluded that the term “öffentliche Dienstleistungen” was not frequently used in everyday language, being too formal.

3.2.3. Finding a Solution

This translation error illustrates the problems that can be experienced when a directly equivalent term to the one used in the source question cannot be found in another language (country). It also highlights the importance of taking steps to promote equivalence between countries. This could be facilitated by providing additional information in the source questionnaire to help translators find an equivalent term. For this question in particular, providing examples of the range of services the question designers want the respondent to think about would help to reduce reliance on a slightly abstract term. In the end the question was made clearer to promote equivalence between countries by

- referring to “social benefits and services” and removing the reference to “the system”;
- giving as examples “health care, pensions and social security”; and
- providing an annotation for social security “. . . cash benefits of one sort or another, such as sick pay, unemployment benefits, child benefits etc”.

The final question wording as fielded in ESS Round 4 was as follows:

ESS Round 4 Question

I am now going to ask you about the effect of social benefits and services on different areas of life in [country]. By social benefits and services we mean things like health care, pensions and social security*.

CARD 30 Using this card please tell me to what extent you agree or disagree that social benefits and services in [country]....**READ OUT...**

		Agree strongly	Agree	Neither agree nor disagree	Disagree	Disagree strongly	(Don't know)
D22	...prevent widespread poverty?	1	2	3	4	5	8

* Annotation: “Social security” meaning cash benefits of one sort or another, such as sick pay, unemployment benefits, child benefits etc.

3.3. Translation Problems Resulting from Source Question Design

As noted in Section 1, when this type of problem is discovered it suggests that although the question could function reasonably well in the source (and possibly some target) language(s), there is something inherent in its design that makes translation particularly difficult.

Evidence of this type of error can lead to doubt regarding the “portability” of the item into one or more translated versions, for linguistic rather than contextual reasons. In most cases this type of error would require the source question to be amended, but where this is not possible additional translation guidance might reduce the problem.

Feedback during the translation stage of this project amplified findings from expert review prior to testing, which had suggested that finding a good translation for “moral” would be difficult in some languages. Furthermore “moral” was not generally used as an adjective to describe people and expert review also suggested this was an “awkward formulation”.

Test Question

CARD I am now going to ask you some questions about how those aged between 15 and 30 are seen by **other** people in [country]. Using this card, please tell me how likely is it that other people in [country] view those aged 15 to 30 as **moral***

Not at all likely											Extremely likely	(Don't know)
0	1	2	3	4	5	6	7	8	9	10	88	

* Annotation for translators: Moral in the sense of upstanding, law abiding, decent etc.

Question aim: to assess whether or not a series of stereotypes applied to the ‘under 30’ age group. This age group was chosen because the questions were originally developed with older people in mind and we wanted to ascertain how respondents processed this in relation to a younger age group. We were also interested in exploring what respondents understood by the concept of morality.

Cognitive testing revealed that contrary to expectations most countries were able to find a functionally equivalent translation and respondents generally understood the question. However, in one country (Germany) a dimension found in the other countries was lost. Figure 5 shows differences in the understanding of “moral” in the test countries.

Germany used the word for “respectable” (“anständig”) as it was felt to be closest to the intended meaning of “moral” in English. The direct translation was not felt to be suitable in this type of question. Analysis of the cognitive interview data indicated that a crucial dimension of the concept of “moral” was lost in translation, that of knowing right from wrong. German respondents seemed to focus on an individual’s behaviour, their behaviour towards others in society and compliance with rules necessary when living with others. For example, one respondent said this meant “*wie kann ich mich in der Gesellschaft manierlich bewegen*” (to be able to act well-mannered in society). It appears that the awkwardness of the source question led to difficulties for the translation team.

	Moral	R's understanding – dimensions of moral
Source (British English/ Great Britain)		• People having moral codes or values • Knowing right from wrong • Treating others how you would like to be treated • Living by society's rules with a code of conduct • Family background/upbringing
Country – language	Translation	
Bulgaria	“moral” (“Морал”)	As source
Germany	“respectable” (“anständig”)	As source but not including • Knowing right from wrong
Portugal	“to have ethical principles” (“moral”)	As source
Spain	“ethical” (“éticos”)	As source
Switzerland (Swiss French)	“to have moral sense” (“ont un sens moral”)	As source

Fig. 5. Comparison of understanding of “moral”

At a joint analysis meeting the consensus was that regardless of whether the word “moral” existed in other languages, it was not the easiest word to translate and use in this context. Although evidence suggested that it was only in Germany that dimensions were lost, this might have ended up being a problem in other countries in the main stage ESS.

3.3.1. Finding a Solution

Intervention by translators alongside cognitive testing can identify such issues at an early stage and allow the questionnaire designer to amend the source questionnaire. In this instance the final decision made for ESS Round 4 was to ask whether or not people thought each age group had “high moral standards”, a less problematic formulation based on ownership of a value set rather than a personal character description. In addition an annotation to aid translators was added for moral in terms of “being ethical, honest and embracing social norms”. Note that this example differs from the previous translation example. In this example the source of the error is the source question (the difficult term “moral”) rather than with suboptimal translation itself that applied to the more straightforward translation example.

ESS Round 4 Question							
I have just been asking about <u>your</u> views. I am now going to ask you how you think <u>most people in</u> [country] view people of different ages.							
CARD 52 Using this card please tell me how likely it is that <u>most</u> people in [country] view those in their 20s... READ OUT...							
		Not at all likely to be viewed that way			Very likely to be viewed that way		(Don't know)
E17	...as having high moral standards?	0	1	2	3	4	8
* Annotation: "High moral standards": refers to being ethical, honest and embracing social norms.							

3.4. Cultural Portability

The ESS testing produced few examples of cultural portability issues. This is perhaps not that surprising given that earlier questionnaire design phases should have identified these and any culturally inappropriate questions would have been discarded. However, one of the few examples of a cultural barrier to equivalent measurement which was found is discussed below. Although this is a subtle example, it is nonetheless clearly a cultural portability issue.

Test Question CARD Using this card please tell me which of the three statements on this card, about how much working people pay in tax, you agree with most CODE ONE ANSWER ONLY 1. Higher earners should pay a greater proportion in tax than lower earners 2. Everyone should pay the same proportion of their earnings in tax 3. High and low earners should pay exactly the same amount in tax Question aim: to identify respondent preference amongst three different tax collection systems.
--

This question was found to be problematic in all countries, suggesting a source questionnaire problem. However, the research team additionally classified this as a cultural portability problem because the salience of the problem was directly related to different cultural contexts. Many respondents felt that the question wording implied that some level of tax-system knowledge was required in order to answer effectively. Whilst this applied to some extent in all countries, it was regarded as a particular problem in Spain

and Switzerland where there was evidence that respondents felt less knowledgeable about their tax systems than elsewhere. For example, respondents in Spain reported that Option 2 reflected the tax system in their country (when in fact Option 1 did) and respondents in Switzerland exhibited low levels of confidence in answering. The Swiss research team pointed out that in Switzerland it is the sole responsibility of the head of the household to complete tax returns, so it is likely that other members of the household who are not involved in this will only have minimal knowledge of the tax system. This assumption was also reflected in the responses given by some Swiss participants.

3.4.1. Finding a Solution

In terms of trying to reduce the effect of these different levels of knowledge across countries the recommendation was to simplify the question so it would be suitable for cross-national implementation. One possibility was to only include the first two answer options. Another suggestion was to give examples of what each option meant in practice or to provide monetary examples. In the end, considering the evidence from cognitive testing (and quantitative pilot data), examples were added for each of the answer options. Whilst it was anticipated this would help in all countries, it was expected to be particularly beneficial in countries where knowledge about the tax system was lowest.

ESS Round 4 Question

D35 CARD 34 Think of two people, one earning twice as much as the other. Which of the three statements on this card comes closest to how you think they should be taxed?

CODE ONE ANSWER ONLY

They should both pay the <u>same share</u> (same %) of their earnings in tax	1
so that the person earning twice as much pays double in tax.	
The higher earner should pay a <u>higher share</u> (a higher %) of their earnings in tax	2
so the person earning twice as much pays more than double in tax.	
They should both pay the <u>same actual amount</u> of money in tax	3
regardless of their different levels of earnings.	
(None of these)	4
(Don't know)	8

3.5. Summary of Problems and Solutions

As we have shown, the CNEST is a useful tool to identify sources of error and provides specific evidence of the problems experienced by respondents. This evidence can then be used on its own or in combination with the results of other pretesting techniques to find

Table 5. Solutions to sources of error

Source of error	Solution
Source question	Change the source question to make it clearer and/or easier for respondents, or better suited to measure the intended concept.
Translation – simple error	Amend the translation(s) in the countries affected. Advise other translators to avoid the error.
Translation – source question design	Change the source question and, if needed, provide guidance for translation including annotations for specific words or phrases.
Cultural portability	Change the source question to ensure that the item will work across all countries included in the survey, or accept that the concept cannot be measured cross-nationally in a structured interview via an ASQ approach.

solutions to these problems. Table 5 shows the solutions that can be applied at the most general level to “fix” problems from each error source.

Of course, the precise changes that need to be made to improve a question ultimately depend on a wide range of factors, including, but not limited to, the content and form of the question itself, the measurement aims of the question designer as well as consideration of the number of countries the question will be administered in.

4. Reflections on Applying the CNEST

It is perhaps not that surprising that the CNEST appeared to map nearly all the problems found with the ESS test questions since it was developed on the basis of evidence from three prior rounds of ESS questionnaire development and analysis.

Applying the CNEST was not always straightforward. As noted, questions often had multiple errors and sometimes the same error had multiple sources. In other instances it took a number of attempts at classification to be sure that the correct category or categories had been applied. For example the cultural example above was first thought to simply be a source question design problem. But further interrogation of the data indicated that there was a cultural portability issue present too. There were also a few instances where it was sometimes difficult to decide on the *type* of error found. For instance, a question on the extent to which “the system of public services supports work-life balance”, triggered a much broader range of considerations in Britain than in other countries. From the data available it was unclear whether this reflected differences in the provision of work-life balance support between countries or a translation problem, although iterative discussions and review of the data led to final agreement about the error source. There were no instances where there was a suggestion to create a new category within the CNEST.

Attempts to apply the CNEST in other studies including those with different pre-testing methodologies, would be beneficial to further assess the completeness of this tool.

5. Conclusion

Cross-national questionnaire design is a complex task and researchers are faced with many more potential sources of error than when designing a questionnaire for a single country study. The challenges are especially acute in large-scale studies where it is impractical to develop questionnaires in multiple languages throughout the questionnaire design process. It is therefore important that tools are developed that assist researchers to design and assess source questionnaires. It is hoped that the CNEST might prove to be such a tool.

Interestingly, researchers in the U.S. have also developed typologies to assess sources of error uncovered during cross-national and cross-cultural behaviour coding and cognitive interviewing. Not only are these typologies similar in many respects, they also share similarities with the CNEST. This is remarkable considering that the development of the CNEST was informed by a broader set of inputs. Taking this into account, it is unsurprising that the CNEST has some additional and unique distinctions.

The initial attempt at applying the CNEST to the ESS Round 4 cognitive interviewing findings has been encouraging, suggesting that the CNEST is complete, although of course based on just a limited set of questions. Using a robust and transparent cognitive interviewing methodology was a prerequisite for having confidence in this process and is arguably essential in any cross-national pretesting. However, until the CNEST has been applied on other studies and with respect to other pretesting outputs, evaluations remain preliminary.

Most importantly, perhaps, is the clear evidence that identifying the source of the error has a very practical application. Making the error source explicit helped to directly inform the plans for amendment and improvement. Whether this involved identifying simple translation mistakes, poor source questions or cultural portability issues, being clear on the error source led to transparent, evidence-based recommendations for improving cross-national source questionnaire design in the ESS. Furthermore, due to the value of these findings, cognitive interviewing and the CNEST will be utilised in future rounds of the ESS.

Appendix

Test Questions and Probing Areas

INTERVIEWER – READ OUT. . .

Q7 CARD 4

Some people say that certain age groups have a high or low status, while other people say there is no real difference. By *status* I mean the position or standing an age group has in society. I am going to ask you how high or low you think most people in [country] would say different age groups are in terms of their status.

Firstly, using this card, please tell me how you think most people in [country] would rate the status of those aged 15–29

Extremely low status													Extremely high status	(Don't know)
0	1	2	3	4	5	6	7	8	9	10			88	

Further areas to explore (Q7-9)

INTERVIEWER PROBE QUESTIONS 7 TO 9 TOGETHER AND FIND OUT:

- Whether the respondent was able to use the three age groups offered to answer these questions.
- What the respondent understands by “status”. Do they agree with the definition provided? (By status I mean the position or standing an age group has in society) Were they using this definition to answer the question? Or did they use their own different definition.
- How they came up with their answer to Question 7 (15–29 age group)?
- How they came up with their answer to Question 8 (30–70 age group)?
- How they came up with their answer to Question 9 (71 +)?
- If appropriate: How the respondents decided which age group had the highest and lowest status.
- Was the respondent thinking about all three age groups and making comparisons as they answered each item?
- If respondent refuses to answer – note this and find out why
- If respondent says “don’t know” – note this and find out why

INTERVIEWER – READ OUT. . .

Q2 CARD 2

Using this card please tell me, on a scale of 0–10, how efficiently you think the income tax authorities in [country] carry out their work

0 means extremely inefficiently, and 10 means extremely efficiently.

Follow-up questions (Q2)

- How did you come up with this answer? AND
- What were you thinking? AND/OR Why did you pick that number?

Further areas to explore (Q2)

INTERVIEWER – FIND OUT:

- Why the respondent chose the number they did (ie what this means in the context of the question).
- What the respondent understands by “efficient”.
- What the respondent understands by “carrying out their work”.
- Who the respondent thinks “the income tax authorities” are.
- What would the income tax authorities have to be like at carrying out their work for the respondent to have answered “extremely inefficiently”.
- What the income tax authorities would have to be like at carrying out their work for the respondent to have answered “extremely efficiently”
- (If applicable) The respondent’s reasons for NOT choosing a number at either end of the scale (0 or 10)
- If respondent says “don’t know,” “can’t pick a number” or “refuses to answer” – note this and find out why

The next few questions are about welfare and public services in [country].

Using this card please tell me how much you agree or disagree that “the system of public services in [country] prevents large scale poverty”

- ### Follow-up questions (Q3)

- How did you come up with this answer? AND
- What were you thinking when you gave that answer?

Further areas to explore (Q3)

INTERVIEWER – FIND OUT:

- Some examples of what the respondent thinks [country] might be like if there was large scale poverty/understanding of this term. What the respondent understands by the word “poverty”. Are they thinking of poverty in terms of not being able to afford food/basic shelter or relative poverty in that some people have much less than others (a large gap between rich and poor) even though they still have basic food and shelter?
- Whether the respondent thinks there is already large scale poverty in [country].
- What the respondent understands by “the system of public services”. Does the respondent think it only refers to the benefits system, or does it also cover the health system, the education system or possibly other public services such as the fire and police services?
- If respondent refuses to answer or says “don’t know” – note this and find out why.
- What the respondent understands by “prevents” in this question.

INTERVIEWER – READ OUT. . .

Q15 CARD 7

I am now going to ask you some questions about how those aged between 15 and 30 are seen by other people in [country]. Using this card, please tell me how likely is it that *other people* in [country] view those aged 15 to 30 as **moral**⁴

Not at all likely												Extremely likely	(Don't know)
0	1	2	3	4	5	6	7	8	9	10		88	

Follow-up questions (Q15)

- How did you come up with this answer? AND
- What were you thinking?

⁴ Moral in the sense of upstanding, law-abiding, decent etc

Further areas to explore (Q15)

INTERVIEWER FIND OUT:

- How the respondents made a judgement about how **others** view people aged 15 to 30 for each of the things read out.
- How respondents interpret moral (is it that they “have their own morality” or “that they follow the morality of the majority on their country”)
- Why respondents choose the number on the scale for their answers
- What “not at all likely” means to the respondent at this question
- What “extremely likely” means to the respondent at this question
- If respondent refuses to answer – note this and find out why
- If respondent says “don’t know” – note this and find out why

INTERVIEWER – READ OUT. . .

Q1 CARD 1

Using this card please tell me which of the three statements on this card, about how much working people pay in tax, you agree with most

CODE ONE ANSWER ONLY

1. Higher earners should pay a greater proportion in tax than lower earners
2. Everyone should pay the same proportion of their earnings in tax
3. High and low earners should pay exactly the same amount in tax
4. (None of these)
5. (Don’t know)

Follow-up questions (Q1)

- How did you come up with this answer? AND
- What were you thinking?

Further areas to explore (Q1)

INTERVIEWER – FIND OUT:

- How the respondent understands each answer option – what does each one mean to them?
- Whether the statement the respondent chose reflects the tax system in their country.
- Whether the respondent understands the difference between the three options.
- Who the respondent thinks “working people” are.
- What the respondent understands by “high earners” (ask for examples).
- What the respondent understands by “low earners” (ask for examples).
- If the respondent says “none of these” – note this and find out why.
- If the respondent refuses to answer – note this and find out why.
- If the respondent says “don’t know” – note this and find out why.

6. References

- Beatty, P.C. (2004). The Dynamics of Cognitive Interviewing. *Methods for Testing and Evaluating Survey Questionnaires*, S. Presser, J.M. Rothgeb, M.P. Couper, J.T. Lessler, E. Martin, J. Martin, and E. Singer (eds). New Jersey: Wiley, 45–66.
- Beatty, P.C. and Willis, G.B. (2007). Research Synthesis: the Practice of Cognitive Interviewing. *Public Opinion Quarterly*, 71, 287–311.
- Behr, D., Harkness, J., Fitzgerald, R., and Widdop, S. (2008). ESS Round 4 Questionnaire and Translation Queries. London: European Social Survey, Centre for Comparative Social Surveys, City University.
- Carrasco, L. (2003). The American Community Survey (ACS) en Español: Using Cognitive Interviews to Test the Functional Equivalence of Questionnaire Translations. Washington, DC: Statistical Research Division, U.S. Census Bureau.
- Collins, D. (2003). Pretesting Survey Instruments: An Overview of Cognitive Methods. *Quality of Life Research*, 12, 219–227.
- Collins, D. (2007). Analysing and Interpreting Cognitive Interview Data: A Qualitative Approach. *Proceedings of the 6th Questionnaire Evaluation Standard for Testing Conference*. Ottawa: Statistics Canada.
- Conrad, F.G. and Blair, J. (1996). From Impressions to Data: Increasing the Objectivity of Cognitive Interviews. *Proceedings of the American Statistical Association, Section on Survey Research Methods*, 1, 1–9. Alexandria, VA: American Statistical Association.
- Conrad, F.G. and Blair, J. (2009). Sources of Error in Cognitive Interviews. *Public Opinion Quarterly*, 73, 32–55.
- DeMaio, T.J. and Landreth, A. (2004). Cognitive Interviews: Do Different Methods Produce Different Results?. *Methods for Testing and Evaluating Survey Questionnaires*, S. Presser, J.M. Rothgeb, M.P. Couper, J.T. Lessler, E. Martin, J. Martin, and E. Singer (eds). Hoboken, NJ: John Wiley and Sons, 89–108.
- Enerstvedt, R.G. (1989). The Problem of Validity in Social Science. *Issues of Validity in Qualitative Research*, S. Kvale (ed.). Sweden: Studentlitteratur.
- Fitzgerald, R., Widdop, S., Gray, M., and Collins, D. (2009). Testing for Equivalence Using Cross-national Cognitive Interviewing. London: Centre for Comparative Social Surveys, City University Working Paper 2009/01.
- Forsyth, B.H., Stapleton-Kudela, M., Levin, K., Lawrence, D., and Willis, G.B. (2007). Methods for Translating an English-Language Survey Questionnaire on Tobacco Use into Mandarin, Cantonese, Korean and Vietnamese. *Field Methods*, 19, 264–283.
- Gerber, E. (1999). The View from Anthropology: Ethnography and the Cognitive Interview. *Cognition and Survey Research*, M. Sirken, D. Herrmann, S. Schechter, N. Schwarz, J. Tanur, and R. Tourangeau (eds). New York: Wiley.
- Gerber, E. and Wellens, T. (1997). Perspectives on Pre-testing: “Cognition” in the Cognitive Interview? *Bulletin de Methodologie Sociologique*, 55, 18–39.
- Goerman, P. (2006). Cognitive Interview Methodology Revisited: Development of Best Practices for Pretesting Spanish Survey Instruments. Paper presented at the

- Meeting of the American Association for Public Opinion Research, Montreal, Canada.
- Goerman, P.L. and Caspar, R. (2007). A New Methodology for the Cognitive Testing of Translated Materials: Testing the Source Version as a Basis for Comparison. Paper presented at the American Association for Public Opinion Research Conference, May 17–20, Anaheim, CA and submitted to 2007 JSM Proceedings, Statistical Computing Section [CD-ROM], Alexandria, VA: American Statistical Association, 3949–3956.
- Guba, E.G. and Lincoln, Y.S. (1994). Competing Paradigms in Qualitative Research. *The Handbook of Qualitative Research*, N.K. Denzin and Y.S. Lincoln (eds). Thousand Oaks, CA: Sage.
- Harkness, J. (2007). Improving the Quality of Translations. *Measuring Attitudes Cross-nationally: Lessons from the European Social Survey*, R. Jowell, C. Roberts, R. Fitzgerald, and G. Eva (eds). London: Sage.
- Harkness, J., De Vijer, F., and Johnson, T. (2003). *Questionnaire Design in Comparative Research. Cross-cultural Survey Methods*, J. Harkness, F. De Vijer, and P. Mohler (eds). Hoboken, NJ: John Wiley and Sons.
- Jowell, R. and Eva, G. (2009). Happiness is Not Enough: Cognitive Judgements as Indicators of National Well-being. *Social Indicators Research*, 91, 317–328.
- Knafl, K., Deatrick, J., Gallo, A., Holcombe, G., Bakitas, M., Dixon, J., and Grey, M. (2007). The Analysis and Interpretation of Cognitive Interviews for Instrument Development. *Research in Nursing and Health*, 30, 224–234.
- Kudela, M.S., Levin, K., Tseng, M., Hum, M., Lee, S., Wong, C., McNutt, S., and Lawrence, D. (2004). Tobacco Use Cessation Supplement to the Current Population Survey Chinese, Korean, and Vietnamese Translations: Results of Cognitive Testing. Final Report submitted to the National Cancer Institute, Rockville, MD.
- Kudela, M.S., Forsyth, B.H., Levin, K., Lawrence, D., and Willis, G.B. (2006). Cognitive Interviewing versus Behaviour Coding. Paper presented at the American Association of Public Opinion Research, Montreal, Canada.
- Levin, K., Willis, G.B., Forsyth, B.H., Norberg, A., Kudela, M.S., Stark, D., and Thompson, F.E. (2009). Using Cognitive Interviews to Evaluate the Spanish-Language Translation of a Dietary Questionnaire. *Survey Research Methods*, 3, 13–25.
- Madill, A., Jordan, A., and Shirley, C. (2000). Objectivity and Reliability in Qualitative Analysis: Realist, Contextualist and Radical Constructionist Epistemologies. *British Journal of Psychology*, 91, 1–20.
- Miles, M.B. and Huberman, A.M. (1994). *Qualitative Data Analysis: An Expanded Sourcebook* (Second Edition). Newbury Park, CA: Sage Publications.
- Miller, K., Willis, G.B., Eason, C., Moses, L., and Canfield, B. (2005). Interpreting Results of Cross-Cultural Interviews: A Mixed Methods Approach. *Methodological Aspects in Cross National Research*, J.H.P. Hoffmeyer-Zlotnik and J.A. Harkness (eds). ZUMA, Mannheim, 79–92. Available at http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/zuma_nachrichten_spezial/znspezial11.pdf?download=true

- Miller, K., Fitzgerald, R., Caspar, R., Dimov, M., Gray, M., Nunes, C., Padilla, J.L., Pruefer, P., Schöbi, N., Schoua-Glusberg, A., Widdop, S., and Willson, S. (2008). Design and Analysis of Cognitive Interviews for Cross-National Testing. Paper presented at the International Conference on Survey Methods in Multinational, Multiregional, and Multicultural Contexts (3MC), Berlin, 25–29 June, Germany.
- Nápoles-Springer, A.M., Santoyo-Olsson, J., O'Brien, H., and Stewart, A.L. (2006). Using Cognitive Interviews to Develop Surveys in Diverse Populations. *Medical Care*, 44 (Suppl. 3), S21–S30.
- Patton, M.Q. (2002). *Qualitative Research and Evaluation Methods* (Third Edition). Thousand Oaks, CA: Sage.
- Ritchie, J. and Spencer, L. (1994). *Qualitative Data Analysis for Applied Policy Research. Analysing Qualitative Data*, A. Bryman and R. Burgess (eds). London: Routledge.
- Saris, W. and Gallhofer, I. (2007). Can Questions Travel Successfully? Measuring Attitudes Cross-nationally: Lessons from the European Social Survey, R. Jowell, C. Roberts, R. Fitzgerald, and G. Eva (eds). London: Sage.
- Schoua-Glusberg, A. (2006). Eliciting Education Level in Spanish Interviews. Paper presented to the American Association of Public Opinion Research, Montreal, Canada.
- Spencer, L., Ritchie, J., and O'Connor, W. (2003). Analysis: Practices, Principles and Processes. *Qualitative Research Practice*, J. Ritchie and J. Lewis (eds). London: Sage, 199–218.
- Tucker, C. (1997). Methodological Issues Surrounding the Application of Cognitive Psychology in Survey Research. *Bulletin de Methodologie Sociologique*, 11, 67–92.
- Widdop, S.E. (2007). *European Social Survey Codebook: Round 1 and Round 2*. London: Centre for Comparative Social Surveys, City University.
- Willis, G.B. (2004). Cognitive Interviewing Revisited: A Useful Technique, in *Theory? Methods for Testing and Evaluating Survey Questionnaires*, S. Presser, J.M. Rothgeb, M.P. Couper, J.T. Lessler, E. Martin, J. Martin, and E. Singer (eds). Hoboken, NJ: John Wiley and Sons, 23–44.
- Willis, G.B. (2005). *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Thousand Oaks, CA: Sage.
- Willis, G.B., DeMaio, T., and Harris-Kojetin, B. (1999). Is the Bandwagon Headed to the Methodological Promised Land? Evaluating the Validity of Cognitive Interviewing Techniques. *Cognition and Survey Research*, M. Sirken, D. Herrmann, S. Schechter, N. Schwarz, J. Tanur, and R. Tourangeau (eds). New York: Wiley, 133–153.
- Willis, G.B., Lawrence, D., Kudela, M., and Levin, K. (2005a). The Use of Cognitive Interviewing to Study Cultural Variation in Survey Response. Paper presented to the Questionnaire Evaluation Standards Workshop (QUEST) on Questionnaire Design and Testing, Heerlen, the Netherlands, 19–21 April.
- Willis, G.B., Lawrence, D., Thompson, F.W., Kudela, M., Levin, K., and Miller, K. (2005b). The Use of Cognitive Interviewing to Evaluate Translated Survey Questions: Lessons Learned. *Proceedings of the Federal Committee on Statistical Methodology Research Conference*. Arlington, VA.
- Willis, G.B., Lawrence, D., Hartman, A., Kudela, M.S., Levin, K., and Forsyth, B. (2007). Translation of a Tobacco Survey into Spanish and Asian Languages: The Tobacco

Use Supplement to the Current Population Survey. *Nicotine & Tobacco Research*, 10, 1075–1084.

Willis, G.B. and Zahnd, E. (2007). Questionnaire Design from a Cross-cultural Perspective: An Empirical Investigation of Koreans and Non-Koreans. *Journal of Health Care for the Poor and Underserved*, 18, 197–217.

Willson, S. and Miller, K. (2005). Using Qualitative Methods to Improve Measurement Error in Quantitative Data. Paper presented at the annual meeting of the American Sociological Association, Philadelphia, August 12.

Received June 2009

Revised May 2011