Testing for equivalence using cross-national cognitive interviewing

Rory Fitzgerald, Sally Widdop
CCSS, City University

Michelle Gray, Debbie Collins
National Centre for Social Research
Abstract

In recent years there has been significant progress towards achieving greater equivalence in cross-national surveys especially in respect to sampling, non response measurement, fieldwork procedures, data provision and other areas (Jowell et al, 2007). However, efforts to achieve equivalence in construct measurement have lagged behind.

Experience from cross-national surveys, such as the European Social Survey (ESS), has shown that it can be very difficult to write and translate questions that will have the same meaning in different countries. A cross national research project has recently taken place that developed a tool designed to try and increase equivalence prior to data collection. Using cognitive interviewing, questions for both the ESS and the Budapest Initiative were tested in seven countries and six languages.

After briefly outlining the methodology employed the paper focuses on how the results of cognitive interviewing might be used to improve both the questionnaire design process and subsequent attempts to produce functionally equivalent translations. The tool employed to do this is a newly devised typology of ‘error sources’.

The key research question examined in the paper is therefore whether cross-national cognitive interviewing can be used effectively to differentiate different error sources associated with cross-national questionnaire implementation. The typology covers: 1) source question issues - which are found across all countries. 2) translation problems - where translated questions are not functionally equivalent to source questions; 3) source question design and its interaction with translation - where the question appears to work well in the source questionnaire but has features in its design which make translation difficult and 4) cultural issues - where concepts being measured do not exist in all countries. Examples of each of these error types are demonstrated in the paper. The paper concludes with a critical review of the methodology employed on the project and in particular of the typology.

Contact: Rory Fitzgerald, Deputy Director, Centre for Comparative Social Surveys, New Social Sciences Building, City University, Northampton Square, London EC1V 0HB; Tel: +44 (0) 207 040 4903; email: r.fitzgerald@city.ac.uk

This work was supported by the Economic and Social Research Council (grant reference number: RES-000-22-2269).
Introduction

In recent years there has been significant progress towards achieving greater equivalence in cross-national surveys especially in respect to sampling, non response measurement, fieldwork procedures, data provision and other areas (Jowell et al, 2007). However, a key area that lags behind in this respect is the achievement of equivalence of constructs – the foundation of all questionnaire development. Unfortunately, solid evidence of how effectively questionnaires work cross-nationally is scarce. But recent data collected by researchers working on the European Social Survey (ESS) has shown mixed success in developing cross-national questionnaires that measure attitudinal constructs equivalently (Saris, 2007: 71, Harkness, 1997: 91).

Among the most widely used criteria for assessing questionnaire quality are; reliability, validity, the extent of item non-response, relative bias and response effects, (mis)understanding of questions, and problems in the interaction between interviewer and respondent (Saris, 2007). A large body of research has been undertaken into the types of questions that are particularly error-prone in relation to one or more of these criteria, several of which have tested alternative formats and wordings by means of ‘split ballot experiments’ (Schuman & Presser, 1981; Krosnick & Fabrigar, forthcoming). In addition ‘Multi-Trait Multi-Method’ (MTMM) studies have evaluated the effects of question design on reliability and validity (Andrews, 1984; Költringer, 1995; Scherpenzeel 1995; Scherpenzeel & Saris, 1997). Recent findings from the program of MTMM work on the ESS suggest large differences in measurement error between countries with a recommendation that these be corrected prior to commencing analysis to reduce the risk of incorrect conclusions being drawn (Saris, 2007).

However, as sophisticated and useful as these techniques are, they offer only limited insight into the differential causes of measurement errors. Consequently, their role in helping to ‘fix’ weak questions is limited, at least at present, to providing some general rules for emerging structural design concerns (for example regarding the relative benefits of 7 and 11 point scales). At present, the onus falls on end users to adjust the data according to whatever algorithms are made available to them. It would of course be preferable, where possible, for measurement errors to be avoided in the first place and to ensure that the conceptual basis of a question is sound prior to fielding.

In order to improve this situation and to find better ways of tackling the problem of achieving equivalence during the questionnaire design phase, researchers from the European Social Survey and the Budapest Initiative1 formed a collaborative work group. The aim of this group was to develop an effective methodology for employing cognitive interviewing (CI) in cross-national questionnaire design (Miller et al, 2008). Cognitive interviewing is “the practice of administering a survey questionnaire while collecting additional verbal information about the survey responses... [which is] used to evaluate the quality of the response or to help determine whether the question is generating the sort of information that its author intends” (Beatty, 2004:45). Despite being fairly well established in national studies, cognitive interviewing has not yet made its mark effectively on comparative survey work (Smith, 2004:450) no doubt in part because the costs and resource implications can appear insurmountable (Harkness et al, 2003:29). The project (referred to as the CI project) offered an exciting opportunity to address these issues.

---

1 The Budapest Initiative is the name of a work group who came together to try and improve questions to test health measures. The initiative was named after the location of the first meeting.
The CI project was largely successful in its aim of conducting comparable cognitive interviews based around a common interviewing protocol and harmonised methodology in six countries. It also managed to overcome many of the implementation difficulties of previous cross-national cognitive interviewing projects (for a description of these difficulties see Miller, (2007)). A related paper reports on the methodological lessons learnt from the CI project and the recommendations for future cross-national cognitive interviewing that emerged from it (Miller et al, 2008).

In this paper we will highlight the lessons learnt from a truly qualitative analysis of a subset of questions included in the CI project which were designed for the fourth round of the European Social Survey (ESS). We will consider the role that cognitive interviewing – could play in cross-national research design.

We will describe the methodology used to conduct a comparative analysis of cognitive interviewing data collected across a range of countries, providing a detailed example of the analysis of one question. We will then set out our argument that for the cross-national questionnaire designer to be able to use the results of cognitive interviewing effectively and to identify a strategy for resolving the problem, it is necessary to understand the overarching source of the error(s) identified. Drawing on experience from question development during previous rounds of the ESS, a four item typology of some of the main error sources in cross-national questionnaire design was developed. Using this typology we will demonstrate the range of problems cross-national cognitive interviewing can reveal, and in turn show how this provides a framework for developing remedies.

We will close this paper by arguing that cognitive interviewing can, when conducted to rigorous standards, be a highly effective pre-testing tool in cross-national questionnaire design. Its unique contribution to the pretesting process means that for large scale high quality surveys it should be included wherever possible. The cost associated with this activity in terms of both time and money is justified when improvements in the validity and reliability of survey measures can be made. This is an especially important consideration for time series projects when questionnaire items are intended to be repeatedly administered.

First though, we will provide a brief summary of the ESS and its approach to cross-national questionnaire design. It is necessary to understand the procedures by which questions for the ESS are developed prior to testing because the range of problems found will in part be dependent on these prior efforts.

The European Social Survey and its questionnaire design approach

Background to the ESS

The European Social Survey was established in 2001. It is a biennial, academically led cross-national attitude survey, with one half of its questionnaire repeated at each round. The other half consists of two (or three) rotating modules, the topics and authors of which are selected via round-by-round Europe-wide competitions. The questionnaire is around one hour in duration, and is conducted using face to face interviewing in all participating countries. Equivalent, high quality random probability samples of residents aged 15+ in each nation are selected for interview and response rate targets are high or very high (Jowell & Eva, 2008). The project is now entering its fourth round with three cross-national data sets already available online and many of the 19,000+ registered users of the data already producing papers, articles and books.
The survey covers over 25 European countries in each round. Although welcome, such wide coverage poses challenges that are especially pertinent to its questionnaire design. For example, the entrance of Turkey into the ESS in Round 2 as a Muslim country raised immediate issues about the Judaic-Christian assumptions behind existing questions on religion. Similarly, questions on ‘democracy’ cue in different issues among the ‘new’ democracies of Eastern Europe from those in Western Europe. Most recently, questions on pensions have been posing problems because of large differences in provision between countries (Fitzgerald & Jowell, 2008).

**ESS Questionnaire Design**

It is worth noting that not all cross-national surveys employ the same approach as the ESS and both the methodology and amount of effort spent on questions prior to pretesting is likely to be an important determinant of the types of problems discovered by cognitive interviewing.

The ESS questionnaire is developed in British English and subsequently translated to achieve equivalent meanings rather than word-for-word transformations (Harkness, 2007). Such an approach is the most common method of questionnaire design in cross-national research. With the single exception of a translation into one other language for a large scale pilot towards the end of the design process, all design and evaluation of the ESS questions takes place in British English, the working language of the ESS. Almost all questions are closed and are administered in the same format in all countries with the same answer categories following an ‘Ask the Same Question’ (ASQ) approach in all countries (ibid). This is where the various language versions of the questionnaire are derived from translations of a single language ‘source questionnaire’. The success of this approach depends upon the suitability of source questionnaire content and formulation as well as on the quality of the translations. A small number of concepts – such as occupation and education - require country-specific questions that are later coded to a standard classification. All concepts and dimensions in the ESS are ultimately represented in the integrated dataset in an identical format for all countries, facilitating easy comparison.

In developing the ESS core questionnaire, attempts were made to draw on validated questions from other cross-national studies (CCT, 2002). This proved more difficult than anticipated and it was often necessary to adapt existing items or develop new ones. A range of techniques were used to develop and test the core items, and the same methods are used each round for the rotating modules. They include:

- expert papers by substantive specialists;
- a series of reviews by a multi-disciplinary specialist panel;
- consultation with national teams;
- using the Survey Quality Predictor program to estimate the likely reliability and validity of new items;
- a large-scale two-nation quantitative pilot followed by extensive analysis of item non-response, scalability, factor structure and expected correlations;
- translation according to a committee approach following the ESS TRAPD procedures;
- and split ballot Multi Trait Multi Method (MTMM) experiments.

The scope of question design work in the ESS is, at present, restricted by the scant resources available for pre-testing and prior to the CI project there had been no cognitive testing of new items developed for the rotating modules.

---

The questions considered in this paper had already been subject to a series of different drafting stages following expert review over a period of around five months prior to being tested in the CI project. However, they had not been assessed by the ESS National Coordinators from each country nor had they been subject to the large scale pilot.

Before discussing the analysis of some of the ESS questions and considering the types of problems that were discovered we will describe the methodology employed in the CI project.

**The cross-national cognitive interviewing project**

A key concern that has preoccupied previous cross national cognitive interviewing projects is how to ensure that comparisons of cognitive interview findings for the same set of test questions across different countries, with different research teams, are based on data that have been collected and analysed in the same way. Researchers need to be confident that differences between countries reflect a real difference in how the survey question was interpreted rather than a difference in how the cognitive interview was conducted, how the data were collated or how they were analysed. As in other areas of cross-national comparisons this requires that the principle of equivalence is employed in designing the research (Jowell et al, 2007). This does not necessarily mean that identical inputs are always required but it does require prioritisation of this if comparability over different methodological habits and approaches in each country is to be achieved.

This ‘need for standards’ in cross national question testing echoes a debate that has been going on in the qualitative research arena over the past 20 years about the appropriateness and need for standards, guiding principles or evaluative criteria to help assess ‘quality’ in qualitative research (see for example, Spencer et al, (2003a) chapters 3, 5 and 6 for a summary). The CI project set out with the aim of developing such guiding principles for the conduct of cognitive interviewing projects in cross national settings and the CI work group generally accepted the necessity of this in order to ensure equivalence. At the initial planning meeting the aims of the survey questions to be tested were discussed and the objectives and design of the cognitive testing agreed. Protocols were developed in an attempt to try to ensure consistency in the implementation of cognitive interviewing across participating countries. These protocols covered:

- sampling and recruitment;
- interviewing procedures and topic coverage;
- and analysis methods.

This paper is principally concerned with analysis methods. However, both sampling and interviewing procedures also impact upon data quality and so we pause to note a few key points about these.

**Sampling and recruitment**

The sampling protocol used for this study involved convenience sampling to fill specific quotas. The quotas were selected to reflect respondent characteristics that were thought to influence the way in which test questions would be answered. Since

---

3 A similar but slightly different approach was used for testing the questions provided by the Budapest initiative. See Miller et al, (2008) for a discussion of this.
the sample had to meet the needs of testing both the ESS questions and as well as those from the Budapest Initiative the sampling design had to reflect the general population and those with specific health conditions. Countries were asked to conduct a minimum of ten interviews and to obtain a range of different types of respondent, filling agreed quotas. In practice, several countries were able to achieve more than the minimum, as Table 1 shows.

Table 1: Composition of achieved cognitive interviews by country

<table>
<thead>
<tr>
<th>Country</th>
<th>Total</th>
<th>Gender</th>
<th>Age (in years)</th>
<th>Education</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>18–29</td>
<td>30–69</td>
<td>70+</td>
<td>Lower than degree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Men</td>
<td>Women</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Bulgaria</td>
<td></td>
<td>10</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Germany</td>
<td></td>
<td>10</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>UK</td>
<td></td>
<td>29</td>
<td>15</td>
<td>14</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>Portugal</td>
<td></td>
<td>8</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Spain</td>
<td></td>
<td>18</td>
<td>10</td>
<td>8</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Switzerland</td>
<td></td>
<td>17</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>US-English</td>
<td></td>
<td>30</td>
<td>11</td>
<td>19</td>
<td>3</td>
<td>19</td>
</tr>
<tr>
<td>US-Spanish</td>
<td></td>
<td>13</td>
<td>3</td>
<td>10</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Total</td>
<td>135</td>
<td>61</td>
<td>74</td>
<td>32</td>
<td>58</td>
<td>45</td>
</tr>
</tbody>
</table>

This hypothesis-driven approach to sampling is consistent with seeing cognitive interviewing as a qualitative method. The aim is to map the range of problems respondents have when attempting to understand and answer survey questions, to describe the cause of such problems and to identify the characteristics of respondents for whom the questions proved difficult. Such a sampling approach does not lend itself to later quantitative analysis of the data.

Interview protocols

The interview protocol agreed on at the meeting to test the ESS questions used principally non-scripted probes. Interviewers were given a series of areas to cover in the interview but were not expected to use the same verbatim prompts. The interviewing protocol provided a summary of the measurement aims of the ESS questions being tested and indicated the key issues to be explored in the interview. This information enabled interviewers to develop ‘spontaneous’ probes to investigate how respondents went about trying to answer the question whilst ensuring that they were addressing the objectives of the cognitive testing. Although fully scripted probes help to keep interviews consistent they are quite limiting in terms of what they reveal about respondents’ thought processes.

---

4 The Budapest Initiative questions were concerned with testing measures designed to tap respondents overall health states. They included measures such as cognition, hearing and movement.

5 Ideally we would have liked all countries to have conducted more interviews, however resource constraints meant that for many countries ten was all that could be undertaken.

6 This reflected differential resources and expertise available within participating organisations across countries. For example in the US several organisations were involved in the project and in the UK more interviews were possibly since funding had been provided by the ESRC.
The issues to cover in the protocol were developed by the ESS researchers who had worked with the question design teams to develop the modules. They reflected issues of concern that had emerged through the process of expert review. An example is shown below in Figure 1.

**Figure 1: Example of an ESS question from the interviewing protocol (question aim and interviewer probes also shown)**

**ESS question**

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.

<table>
<thead>
<tr>
<th>Extremely inefficient</th>
<th>Extremely efficiently</th>
<th>(Don’t know)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>88</td>
</tr>
</tbody>
</table>

**Question aim**

The intention of this question is to examine respondent perceptions of how ‘efficiently’ the income tax authorities do their job. This will enable the data analyst to examine any link between efficient tax collection and support for the welfare state. The aim is NOT to see whether respondents think the tax system itself is efficient (e.g. it is efficient at taking money from the rich to give to the poor for example). Instead, we want to know whether those charged with the job of tax collection perform their duty efficiently (i.e. on time, accurately, taking all the relevant facts into account).

**Interviewer probes**

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

**INTERVIEWER - FIND OUT:**

- What the respondent chose the number they did (i.e. 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 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
Analysis methods

Our contention is that cognitive interview data are intrinsically qualitative: they are accounts of respondents’ thought processes and as such require a qualitative analysis and not a quantitative one. Quantitative analysis is not suitable because the sampling is based upon convenience methods, with small sample sizes and the interviewing protocol is designed to collect these respondent accounts (Collins, 2007). Spencer et al, (2003b) propose the following ‘hallmarks’ of quality in any qualitative analysis process (see Table 2). The protocol for the analysis of cognitive interview data used on this project has attempted to incorporate these.

Table 2: Hallmarks of Quality

<table>
<thead>
<tr>
<th>Qualitative analysis quality hallmarks</th>
<th>Cognitive interview data analysis – adaptation of hallmark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis remains grounded in the data (i.e. ideas and concepts emerge from the data rather than being imposed on it).</td>
<td>In the context of analysing cognitive interview data this would mean assessing whether the test question has met its measurement objectives.</td>
</tr>
<tr>
<td>The data reduction process is transparent (i.e. that there is a clear route back to the raw interview data).</td>
<td>Within case analysis is focused on individual respondents, looking for example, at whether the respondent uses the same recall device when answering questions concerned with behaviour over the past 12 months, or changes his or her definition of a term, depending on the question. Between case analysis compares answer strategies, for example, across respondents.</td>
</tr>
<tr>
<td>The process permits within and between case (respondent) analyses. This facility will allow greater analytical power, assisting both thematic analysis and the identification of associations between phenomena.</td>
<td></td>
</tr>
</tbody>
</table>

Protocol development

The research team for this project developed a protocol for analysis to ensure consistency and transparency across countries. This was felt to be particularly important in cross national research, though issues of consistency and transparency apply to any research project.

At the initial planning meeting all participating countries discussed and agreed on the analysis shown in Figure 2.
The notes template was organised by test question and the key measurement issues to be explored (i.e. the aims of the test as reflected in the interviewing protocol). The template was useful in providing a standard way for interviewers to report back on findings. As well as containing details on respondents’ understanding of the task or question, for example, interviewers could also record their observations or explanations for what happened. The views of the interviewer were clearly identified (typed in uppercase or entitled ‘interviewer note’). Verbatim quotes could also be recorded in the template (written in quotation marks or italics) allowing the research team to keep their findings grounded in the data.

The data reduction template (chart) used for the ESS questions was the Framework analysis method, described below. Each country was provided with a matrix template to complete. The column and row headings were prescribed, with the column headings reflecting issues to be explored in the cognitive interview (the cognitive probes). Once countries started completing the matrix (after each cognitive interview had been completed), the ESS team looked at the level of information being included and provided feedback. This was important, particularly for countries with little or no prior experience of cognitive interviewing, as it helped identify any problems, such as where more probing was needed in the interview or more detail was required in the matrix.

---

7 Recording could be audio or audio-visual, depending on availability of resources and existing organisational practices.
8 This was an optional requirement and some countries went straight to completing the data reduction template, when resources were limited.
9 The template is useful for studies employing cognitive probes (rather than think aloud) and is based on the standard question and answer model.
Description of Framework

Framework, developed by the National Centre for Social research in the UK primarily for use in organising qualitative interview data, is a matrix based analytical method that facilitates rigorous and transparent data management such that all the stages involved in the ‘analytical hierarchy’ can be systematically conducted (see Ritchie & Spencer, 1984). It also allows analyst to move back and forth between different levels of abstraction without losing sight of the ‘raw’ primary data.

The name Framework comes from the thematic framework that is central to the method. The thematic framework is used to classify and organise data according to key themes, concepts and emergent categories. Main themes are identified and sub divided into a series of related sub-topics that evolve and are refined through familiarisation with the raw data. Once judged comprehensive, each main theme is ‘charted’ in its own matrix, where every respondent is allocated a row and each column denotes a separate subtopic. Data from each case are then synthesised with the appropriate parts of the thematic framework. Table 3 summarises the main steps involved in carrying out this analysis for qualitative data.

Table 3: Steps involved in carrying out a case and theme analysis

<table>
<thead>
<tr>
<th>Step</th>
<th>Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Familiarisation with data</td>
<td>Reading transcripts or notes, listening to interview recordings</td>
</tr>
<tr>
<td>2) Identification of initial concepts/themes</td>
<td>Highlight, summarise, provisionally label</td>
</tr>
<tr>
<td>3) Categorisation leading to description</td>
<td>Iterative process of refinement, starting close to the data and becoming more abstract and interpretative, asking questions of the data such as:</td>
</tr>
<tr>
<td></td>
<td>• is this a different manifestation of that?</td>
</tr>
<tr>
<td></td>
<td>• is this a subset of that?</td>
</tr>
<tr>
<td></td>
<td>• is this of the same order as that?</td>
</tr>
<tr>
<td>4) Seeking explanations: informed by hunches and hypotheses, reflections during fieldwork and analysis and other research or theories</td>
<td>Detailed within case analysis</td>
</tr>
<tr>
<td></td>
<td>Comparison between cases</td>
</tr>
<tr>
<td></td>
<td>Repeated interrogation of the data</td>
</tr>
<tr>
<td></td>
<td>Moving back and forth between cases searching for rival explanations</td>
</tr>
</tbody>
</table>

Tailoring Framework to cognitive interview data analysis

We were attracted to using Framework as an approach for analysing cognitive interview data because it provided a rigorous and transparent way of organising and interpreting it. In addition, it was felt that this method would enable us to examine the data yielded from a cross national study in a more systematic and detailed way than methods of analysis previously employed in other projects. In order to command the confidence of questionnaire designers it is essential that the analysis process is

---

10 For this project the matrix was set up in Excel. However, a standalone Computer Assisted Qualitative Data Analysis Software (CAQDAS) package is now available (see http://www.natcen.ac.uk/framework/index.htm for more information).

11 On a practical note, we were also able to utilise the experience and expertise of NatCen’s Questionnaire Development and Testing (QDT) Hub, who had developed the use of Framework for analysing cognitive interview data and previously run training courses on its use.
transparent and can be easily replicated. Furthermore surveys like the ESS insist on transparency during all stages of their design process. Using a charting system such as Framework is an ideal way to meet this standard and we would argue that such an approach should become the norm for large scale publicly funded surveys. The days when researchers report that they ‘did the analysis’ without describing how it was done should surely become a thing of the past.

Framework required some adaptation to reflect the specific objectives of cognitive interviewing. The basic analytical process is the same as for qualitative data. Steps 2 and 3 in Table 3 above - creating categories and classifying data within them - is a somewhat different process for cognitive interview data because the interview is much more focused or structured than a standard qualitative interview is. We are primarily interested in the cognitive processes respondents’ use when attempting to answer a survey question or complete a questionnaire. In qualitative research, the categories would emerge from the data: in cognitive testing some of the analytical categories are implicitly built into the data collection process. The question and answer model informs our choice of scripted probes or areas to cover in the interview, and these in turn form the main analysis categories. However, the framework is flexible and can be augmented to capture other issues that emerge from the data that do not fit within the cognitive framework. This is similar in some ways to the ‘logical issues’ categories developed by Willis et al, (1999) in their standardised coding scheme. Figure 3 illustrates how the framework matrix for analysing cognitive interview data might be set up.
### Figure 3: Illustration of framework matrix for analysis of cognitive interview data

**Overview of analysis process used**

As mentioned earlier, countries were asked to complete the Framework data reduction template (or chart) in English for each test question. This meant that these documents could be reviewed and scrutinised by all members of the research group. However, it is likely that there was some loss of data quality due to reliance on the translation skills of single members of the research team in each country. The template reflected the issues that were to be explored in the cognitive interview (i.e. the interviewing protocol).

---

11 All members of the research group – i.e. all researchers from participating countries – were competent English speakers.

---

**Overview of analysis process used**

As mentioned earlier, countries were asked to complete the Framework data reduction template (or chart) in English for each test question. This meant that these documents could be reviewed and scrutinised by all members of the research group. However, it is likely that there was some loss of data quality due to reliance on the translation skills of single members of the research team in each country. The template reflected the issues that were to be explored in the cognitive interview (i.e. the interviewing protocol).
In advance of the joint analysis meeting, countries produced a bullet point summary of key findings (in English) for each test question. These were shared with the rest of the research group and provided a starting point for discussion at a joint analysis meeting. The main points were then collated and the overall findings were summarised and agreed upon by the group. This further data reduction phase was done by committee at the meeting to ensure that key points were not lost or misconstrued in the process of translation (to English) or through summarisation. This summary document of all countries’ findings along with individual country’s bullet points and complete data reduction templates (charts) were used for further analysis after the meeting.

Following the joint analysis meeting the ESS research team in collaboration with those at NatCen divided the ESS questions between them and conducted a more detailed analysis. The analysis adopted an organised and structured approach, which relied heavily on the information collected at the joint analysis meeting as well as a thorough interrogation of the data in the charts from all European countries. The main aims were: to try and identify how respondents understood the question overall and how they went about answering it; identifying and classifying the nature of any problems found and assessing how ‘serious’ they were in terms of the impact that they could have on data quality if they were retained in the same format. However, we also noted where respondents expressed no difficulty and looked for evidence of when a question was working well. This contributed to our assessment of how serious the problems were by considering the relatively successful elements of the question in the context of any other problems found.

The stages of analysis were as follows:

1. Examinations of evidence of overt response problems - non-response to original survey question (i.e. ‘don’t know’ or ‘refused’).

2. Evidence of other respondent behaviour suggesting confusion and/or response problems. This included behavioural and verbal activities such as hesitation or requests for repeats.

3. Identification of any other issues that may help to explain or account for findings. This included contextual information identified by the country representatives at the joint analysis meeting such as possible inaccuracies with translation or an overall impression that respondents were uncomfortable answering a question.

4. Assessment of respondent’s understanding of the question overall (in each country) and how they came to answer it. For every pattern of ‘correct’ or ‘incorrect’ understanding or processing of questions that was found at the joint analysis meeting the data were checked across all countries in a systematic fashion to ascertain which countries also had evidence of this finding. For attitudinal cognitive testing, ‘correct’ and ‘incorrect’ refers not to

---

13 Only when something particularly divergent or interesting emerged from the joint analysis meeting did we compare our findings with the American charts. Our emphasis was on the European data since the questions were designed for European administration. It is hoped to examine the data from the US in more detail later.

14 The collection of this information did not form part of the interviewing protocol. However, anecdotal evidence emerged during the joint analysis meeting and was included as part of the analysis process. For future projects we would recommend that this information is collected systematically, as part of the interviewing protocol.
the attitude that the respondent holds but rather whether they understood the question in the way intended by its designers.

5. Identification of key findings. The structure of this stage varied depending on the question content and/or structure of the question i.e. attention may have been given to respondents understanding of key phrases or words in the question and response options or their ability to make links between different parts of the question, or to differentiate between response options. The key findings were summarised with evidence that was systematically collected from each country and linked to the error source typology (see the section describing the Step by Step Analysis). Relevant unique case identifiers were recorded, any country specific variations noted and explanations given where possible. Comparison between cases, repeated interrogation of the data and searching for rival explanations were all key features of this stage of the analysis. This stage involved heavy reliance on the data included in the framework charts and involves an implicit acceptance that the description in English accurately reflects the data from the original interview. This is not without possible dangers since the translations were done by a single researcher in most cases. An additional stage of verification was therefore built in (see stage 8).

6. Identification of overall conclusions (including application of the error source typology to classify errors found).

7. Production of a list of recommendations of changes to improve the questions.

8. Country verification. The research teams from the European countries were asked to verify the findings from their country. This ensured that the analysis remained grounded within the countries where the questions had been tested\textsuperscript{15}. This stage was especially important because we did not know enough about the participating countries to judge whether our interpretation of their data was correct or not. One of the ten golden rules of cross-national research is to never seek to analyse data about a country one knows little if anything about without engaging researchers from that country to help you (Jowell, 1998).

Each question was considered individually and the analysis plan consulted alongside the notes from the joint analysis meeting, the bullet point lists and data charts produced by each participating country.

Before we illustrate the full analysis process for one of the test questions, we consider the types of error cognitive interviewing can uncover as part of a cross-national survey pretesting strategy.

\textbf{The error source typology}

The ESS prides itself on its transparency during all stages of its design, execution and archiving. This extends to ensuring that all known deviations are published openly on the ESS website. Drawing on these known examples, previous rounds of ESS questionnaire design and pretesting, as well as work by Saris (2007) and Harkness (2003), a pattern of repeated problems began to emerge. The ESS team developed an

\textsuperscript{15} Unfortunately, it was not possible to do this comprehensively for all the test questions. We would recommend that this stage should always be built into the analysis process.
error source typology which was used in this project to categorise the problems found with the questions that were tested. It was hoped that once the source of an error was clearly identified it would become easier to identify a method by which to ‘fix’ the problem.

The typology was developed specifically for the CI project whilst fieldwork took place. It was presented to participants at the joint analysis meeting and was based on the different types of error identified from reviewing the first three rounds of the ESS. These are discussed further below.

a) **Source question** – all or part of the source question has been poorly designed leading to problems which are found across all countries.

This paper will not rehearse the different components that make up good and bad questions (see for example DeVaus, 2002; Groves et al, 2004). We would expect that one of the main sources of error cognitive interviewing would highlight would be poorly designed questions in the source questionnaire.

It is likely that this type of error will always be found in testing projects where the questionnaire is administered in the source language and it might also be found in questionnaires administered in other languages. A translation may inadvertently ‘replicate’ the error in a source questionnaire into the target language questionnaires although this should not necessarily happen where translations are designed to be functionally equivalent rather than word for word replications.

b) **Translation error** - where translated questions are not functionally equivalent to source questions. This second category reflects evidence that has arisen over three rounds of the ESS where translation teams have either made avoidable errors or where despite best efforts a functionally equivalent translation was not realised, even though there is no cultural reason to suggest it was not possible to do so.

A clear example from ESS Round 1 was when ‘be wealthy’ was inadvertently translated as ‘être en bonne santé’ (be healthy) in the French questionnaire. The question should have been as follows:

“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?”

Also in Round 1 in Denmark "Crime" was translated as "lovovertrædelse", which roughly back translates as "breach of the law". This resulted in very few Danes agreeing with the statement compared to other countries. The question should have been as follows:

“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 any crime, they should be made to leave!”

These ‘translation errors’ came to light only when data analysts began quarrying the data and alerted the Central Coordinating Team. They compromised comparability to such an extent that the items had to be removed from the combined data set. If such errors are discovered at the pretesting stage it might help to identify questions where translation guidance should be provided or enable potential problems to be flagged to translation teams in advance. Furthermore, if each country were to conduct their own small scale cognitive pre-test they would be likely to pick up on many, if not all,
examples of this type of error. Translation errors would never arise in countries using the source questionnaire in the source language (i.e. British English in the ESS).

c) Source question and its interaction with translation - where the question appears to work well in the source questionnaire but has features in its design which make translation difficult leading to measurement problems.

The ESS includes a program of quantitative work designed to improve the quality of its questions and to learn lessons about the types of question format that work best cross-nationally. It has become increasingly clear that there are certain types of question formulation, which work better cross-nationally than others, although it is still early in terms of the development of this work.

In cases where this type of error is identified, it might be that the question format is also sub optimal in the source language but the critical difference is that there is something in the question’s design that is especially problematic cross-nationally making a functionally equivalent translation difficult to achieve.

Multi Trait Multi Method split ballot experiments have suggested that 7 point anchored response scales have lower quality than 11 point anchored scales (Saris, 2004; CCT, 2008). Similarly, queries received during the translation process have revealed that translation teams often struggle to find functionally equivalent translations to fully labelled scales, suggesting that ‘anchored’ scales, where only the end points are labelled, might be more effective. In ESS Round 4, for example, the scale shown below caused some translators difficulty when trying to find functionally equivalent scale descriptions to the British English (Behr et al, 2008).

“Which option on this card best describes whether or not you can discuss personal issues such as feelings, beliefs or experiences with any 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

Whilst the British English formulation itself has some vagueness (e.g. how many issues count as ‘a few’ compared to the number that count as ‘some’) this became even more pronounced once translated into 25 or more different languages.

d) Cultural issues – when the concepts being measured do not exist in all countries or do not exist in a form similar enough to allow equivalent measurement or to pose particular questions (i.e. you can’t simply write a better question or improve the translation). This would also include circumstances where concepts have such varied salience between countries that it creates serious measurement difficulties. It may on occasion be possible to tap the required concept by using less direct measures but the current questionnaire approach could not produce equivalence.

The prevalence of this type of error is likely to depend greatly on the geopolitical coverage of the study. A world survey would face this type of challenge more frequently than a study conducted on a single continent. As the ESS covers most of geographic Europe this has posed increasing challenges - especially as the project includes countries like Russia, Turkey and Israel in addition to those in the EC.
When the ESS core questionnaire was designed many data analysts wanted to include measures of ethnicity via direct, harmonised questions. However it quickly became apparent that asking about such a concept in this manner was fraught with difficulty stemming from cultural differences. The ESS team rejected using a single closed question to obtain this information concluding that “…ethnic identity is not equally ‘real’ in each country. The majority group in Western societies does not use the term "ethnic group" and refer mostly to geopolitical realities such as the nation, or a sub-nation or region. It seems to be more salient for immigrants who refer to their origin. The same category that has the quality of an "ethnic identity" in one context (e.g. European in New Zealand) may have none in another context ("European" in Europe)” (CCT, 2002). Ultimately it proved possible to provide the analyst with a series of measures to tap ethnicity including parental country of origin, nationality, mother tongue and religious denomination. Here we can see that cultural issues were identified as a major barrier to direct measures of ethnicity leading to rejection of the approach taken to tap this concept used by other cross-national surveys. This is a clear example of culture and context acting as a challenge to cross-national measurement.

Other cultural differences have been identified during ESS questionnaire design that have thwarted attempts to measure specific concepts as intended or even at all. For example in ESS Round 4 the module on attitudes towards welfare had been expected to include items to measure respondent receipt of welfare benefits. However, it proved impossible to devise such a set of items because of the extensive differences between welfare systems across Europe. Similarly, in ESS Round 2 developing questions for a medical module that asked about a ‘GP’ caused serious problems since the concept of a 'general family doctor' simply did not exist in the same way in all European countries.

Cross-national cognitive interviewing will hopefully provide evidence indicating the precise manifestation of these types of errors, which will enable decisions to be made about whether specific concepts can be measured across countries and if so, what the best approach to achieve this would be.

**Step by Step Analysis**

In this section we will consider one of the ESS questions in depth to show how the analysis was conducted and the error source typology applied.

**Example: 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

**Stage 1: Evidence of overt response problems**

We looked across all cases in each country for respondents who answered ‘don’t know’ or refused to answer. In Bulgaria, Germany, Portugal, Spain and Switzerland all respondents gave an answer – there were no ‘don’t know’ or refusals recorded. In the UK one respondent (JF05) answered ‘don’t know’. The data suggested that they
could not understand the difference between response options 2 and 3 (although they didn’t explicitly say this) and this may have led to them answering this way.

Stage 2: Evidence of other respondent behaviour suggesting confusion and/or response problems

We also looked for evidence of other respondent behaviour which may have indicated confusion or other difficulties in answering this question. At the joint analysis meeting both the Bulgarian and German research teams reported that there were long periods of hesitation from respondents before answering this question. Whilst hesitation is not a problem in itself it may indicate that the question is difficult to comprehend or requires a lot of processing in order to come up with an answer.

Stage 3: Identification of any other issues to explain or account for findings

We asked countries to report on anything else that they felt might have had an impact on the responses given. The research teams from Portugal, Spain and Switzerland mentioned problems translating the source question and respondents low level of knowledge about the tax system in their country.

- Translation of the source question

In Portugal and Spain the translators added ‘should’ so the question read: ‘Using this card please tell me which of the three statements on this card, about how much working people should pay in tax, you agree with most?’ This was to make the question stem match the response options which already included the word ‘should’. This was an omission by the question design team in the source questionnaire, which was identified through cognitive testing. The word ‘should’ clearly needed to be included in all countries.

A slightly more problematic translation issue was reported by the Swiss team. They noticed that respondents struggled with the differences between response options 1 and 2. Closer examination of the translation revealed inaccuracies in the translation of option 2 which meant that it was too similar to option 1 and therefore not equivalent to the source questionnaire. Option 1 appeared to be pretty close to the source ‘Les hauts revenus devraient payer une plus grande proportion d’impôt que les bas revenus’ but option 2 ‘Tout le monde devrait payer la même proportion d’impôt par rapport au revenu’ appeared to be suggesting a similar approach to option 1 - again by stressing that tax should be paid in proportion to income when in fact this option referred to the same proportion being paid.

- Respondent’s level of knowledge

The other issue observed at this stage of analysis was the impression gained by the Swiss and Spanish interviewers in particular that respondents were not confident in answering this question and felt that it required sufficient knowledge about tax systems in order to answer properly.

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 respondents.
In Spain, a slightly different concern emerged. Respondents did not always know which tax system existed in their country and on occasion thought that it was best reflected by option 2 rather than 1. Although the question designers did not require respondents to understand the tax system in their country, there was concern that respondents with no understanding of the tax system in their own country may have struggled to understand this question. It also indicates perhaps that Spanish respondents might be basing their interpretation of the response options (and their answer) on what they know happens in their country – rather than thinking about the question more generally or hypothetically as intended. Alternatively, it might indicate that they attempted to identify the option which best reflected the tax system in their country and then chose that option without attempting to understand the others - although there was no firm evidence of either of these things at this stage of the analysis.

Stage 4: Assessment of respondent’s understanding of the question overall (in each country) and how they came to answer it.

This in many ways was the most important part of the analysis process. In cognitive interviewing we are trying to examine respondents thought processes as they answer the question. It is therefore important that the data charts indicate how the information came to light. For example, it may emerge spontaneously from a respondent or as the result of an interviewer probe - either way it is valuable data and should be recorded.

In order to complete stage 4 of the analysis we examined the charts from each country that detailed the reasoning given by respondents to explain how they came to their answers. The data revealed two main answer strategies used by respondents; 1) they answered based on what they thought was fair or 2) they answered based on what was seen as realistic, practical or feasible in their country. This suggests that the question was broadly working as intended by the question designers.

- Respondent’s answers were based on what they thought was fair

This was the most common strategy adopted by respondents in all countries. This was good news considering that this was what the question designers hoped people would base their answers on. In some cases this meant that respondents chose option 1 because they felt that lower earners could not pay more than they do now and because of a general sense of social justice. For example, a UK respondent (GM01) said: “It seems fairer that way because low earners...they've still got to keep their families going” and in Switzerland respondent 11 commented: “people who have more money have to pay a greater proportion than lower earners because they have more resources than lower earners”. In other cases it meant that respondents chose option 2 because of a sense that it was only fair that tax was applied equally to all. For example, Bulgarian respondent 09 said “all people rich and poor should be taxed with the same share” and Spanish respondent 01 said that he thought it was unfair that “because one has higher assets one should pay a higher proportion” and so he chose option 2.

- Respondent’s answers were based on what was seen as realistic, practical or feasible in their country

This strategy was less commonplace but was used by respondents in all countries. The general feeling was that those who earn more can afford to pay more - this was mentioned by GM04 in the UK, respondent 11 in Switzerland, respondent E4 in Portugal and respondent 04 in Germany who said “they [higher earners] have
enough [so] they won’t have problems in paying the money”. Other respondents such as E3 and E6 from Portugal said that different proportions should be applied to those who earn different amounts in other words “there should be different tax groups applied in proportion of what each one earns” (E3). In addition, respondents 02 and 05 in Germany both observed that option 3 was “rubbish” and “isn't even feasible”.

A specific concern that was raised at the start of the project from the question designers was that respondents’ answers would simply mirror the tax systems they have in their own country without consideration of the other response options. However, evidence from the cognitive interviews suggests that this fear was unfounded with the possible exception of Spain (see stage 3). Most respondents acknowledged the tax system that existed in their country and if this matched the code they chose they stressed that this was not the reason why they picked it. For example, respondent 11 from Switzerland chose option 1 - not because this was what was used in Switzerland but because he thought it was fairest. Other respondents acknowledged what tax system was used in their country but chose different codes e.g. UK respondents SA04 and SA07 both chose option 2 but identified that option 1 reflected the tax system in the UK.

Stage 5: Identification of key findings – summarised with information and examples from each country and linked to the error source typology

At this stage of analysis we worked through specific elements of the question that had been focused on during the interviews. Particular attention was given to three key areas: 1) differentiating between the response options 2) understanding the response option chosen and 3) problems with specific words or phrases such as ‘proportion’, ‘higher earners’, ‘lower earners’ and ‘working people’.

- Differentiating between the response options

There was clear variation between countries in terms of respondents’ acknowledgement and understanding of the differences between the response options, suggesting a lack of equivalence between countries.

In all European countries we found evidence of respondents reporting directly (or for whom it could be inferred from their answer) that they recognised the difference between the response options (Switzerland - CH04, Germany – 01, 02, Spain – 33, Bulgaria – 01, 02, 05, UK - GM05, SA04). In Portugal, for example, respondent E3 commented “It doesn’t make much sense for a person earning 500 Euros to pay the same thing as a person earning 5000 Euros…I believe that there should be different tax groups applied in proportion of what each one earns”.

At the same time, there is clear evidence from all countries of confusion about the difference between the answer options provided. In addition, the answer options concerned differed between countries. As noted earlier, in Switzerland, the difference between options 1 and 2 was not always clear to respondents whereas in the UK, Bulgaria, Spain and Germany the difference between options 2 and 3 was more problematic. In Spain for example, it was reported that respondents had difficulties distinguishing between ‘same proportion’ (option 2) and ‘same amount’ (option 3).

- Difficulty understanding the differences between options 1 and 2

In Germany, the UK, Bulgaria and Portugal we found evidence of a good level of understanding of the difference between options 1 and 2. In Spain too, respondents
generally made a distinction between the first two categories although there were exceptions (e.g. SP13). Swiss respondents had far greater difficulties. For example respondents 21, 24, and 31 seemed confused and respondent 21 also changed their answer during probing. Examination of the translation suggests that functional equivalence may not have been achieved\textsuperscript{36}. We would classify this as a translation error (using the error source typology described earlier in this paper).

- **Difficulty understanding the differences between options 2 and 3**

The difference between options 2 and 3 was more problematic and difficulties were experienced by respondents in all countries except Portugal. For example, Spanish respondent 04 said “...I didn't see any difference...they are saying the same [thing]”. This sentiment was echoed by respondents in Bulgaria, Germany, Switzerland and the UK. On occasion, UK respondents (e.g. JF01, JF04 and JF05) struggled with the distinction between options 2 and 3. This suggests a problem with the design of the source question. Unfortunately, there is not enough data recorded in the charts to suggest why respondents could not make a clear distinction between these options. However, the fact that they could not means that any efforts to better distinguish between these codes would be beneficial.

- Understanding the response option chosen

In order to ascertain whether respondents had understood the response options we looked at the data given by respondents for their chosen response and checked whether they had comprehended it in the way envisaged by the question designers. This process was completed for each country.

- **Evidence of understanding from respondents who chose option 1**

In Portugal there was clear evidence of understanding of option 1 amongst respondents who chose this answer. For example, respondent E1 refers to ‘...the higher the persons income, the higher the proportion that he / she should pay.’ Similarly in Germany (respondents 08 and 10), the UK (GM04, JF03), Bulgaria (BG02) and Spain (09, 16 and 18) there was also evidence that respondents understood the meaning of option 1.

However, other respondents were less clear. In Spain the data is unclear on whether or not respondents truly understood that the percentage being paid would be different between options 1 and 2. In addition, due to the problems Swiss respondents had differentiating between options 1 and 2 there were no clear examples of respondents having understood option 1. This problem was compounded when we found evidence of respondents who did not understand any of the options (e.g. respondents 01and 06).

- **Evidence of understanding from respondents who chose option 2**

There is mixed evidence of understanding amongst respondents who chose option 2. In the UK, Bulgaria, Germany, Spain and Portugal there was some evidence that this option was understood. Bulgarian respondent 09 said that this option was the fairest because everyone should pay the same share; Spanish respondent 04 explained that he picked this option because he thought it was unjust that people with higher assets should pay a higher proportion. In the UK, respondent GM06 chose it to stop instances where high earners got taxed so much that they actually earn less than

\textsuperscript{36} The Swiss translation of the codes was discussed in detail earlier in this paper.
lower earners. In Spain and Switzerland, there was no clear evidence in the data about whether or not respondents choosing option 2 understood it. As noted earlier in the Swiss cases, it seems that those who chose option 2 tended to be very confused about the difference between options 1 and 2.

- **Evidence of understanding from respondents who chose option 3**

This option was not selected in Switzerland, Portugal, Germany, Bulgaria or the UK. It was chosen by a few respondents in Spain but without clear understanding it. Respondents SP03 and SP08 chose it but only SP03 seemed to be clear on its meaning. Evidence for understanding of this option is best obtained from respondents who did not choose it but commented or reflected upon it. However this was only possible in some countries due to a lack of data. In the UK, this option was described as “ridiculous” (SA08) as lower earners would loose out and higher earners would benefit. Echoing this sentiment, a respondent in Portugal called option 3 “outrageous” (E6). Those UK respondents who could describe what it meant talked about it in terms of paying the same or a fixed amount of money regardless of what people earn (JF07, GM04, GM07) rather than a percentage or a proportion (as with the other two options). For example, GM07 said that option 3 meant “everybody should pay an equal, fixed amount”.

- **Problems with specific words or phrases**

Testing revealed that there was evidence to suggest that certain words used in this question were problematic. However, it is important to note that the difficulties identified concerning respondents understanding of key words may be more indicative of a weakness with the interviewing approach than with a lack of understanding by respondents. We do not know, and indeed cannot possibly know for certain, if the explanations that respondents came up with about what they understood by ‘high’ or ‘low’ earners for example, was actually what they were thinking about at the time of answering the original survey question, or whether it was something they thought about afterwards in response to probes designed to gauge their general understanding. In future, we would recommend a clear mechanism for recording the ‘route’ by which the answer was elicited in the charts, for example by prefixing responses to direct probes with a ‘p’ or having a special column in the charts for responses to the non-directive introductory probes. Even when the route is clear however, it might not always be possible to determine.

We will now consider the key words or phrase in turn, highlighting the problems raised by each.

- **‘Proportion’**

Swiss respondents (e.g. 11, 20, 21 and 24) had difficulty understanding this term and ‘greater proportion’ as used in option 1 was not understood at all. In Portugal too, there was confusion about whether the use of ‘proportion’ in option 2 (Everyone should pay the same proportion of their earnings in tax) also referred to a percentage. This may in part be attributed to the translation used in Portugal where a Portuguese equivalent of the English word ‘part’ was used instead of proportion or percentage, suggesting a possible translation error.

- **‘Tax’**

On occasion, some respondents asked whether ‘tax’ referred to income tax OR to wealth tax (investments or assets). Upon reflection, it is obvious to us that this was not clear in the source question.
- ‘Higher’ and ‘lower’ earners

The German research team observed that uses of these words were inconsistent throughout the question which would not make the respondent’s task any easier. References should either be made to ‘higher’ and ‘lower’ earners or to ‘high’ and ‘low’ earners but not to both. In general, respondents understood the difference between the concepts of ‘high’ and ‘low’ earners but the relative emphasis given to monetary amounts earned or the occupations concerned differed. One way in which respondents understood ‘higher’ and ‘lower’ earners was in terms of the types of occupations that people had e.g. a lawyer compared to a cleaner, others related it to the salary that people earned. This pattern of giving occupational or monetary examples was found in all countries.

- ‘Working people’

In all countries, except Bulgaria, people mainly understood the term ‘working people’ to simply mean ‘those doing paid work’. However, in Bulgaria, one interpretation was that ‘working people’ could mean those who are available for work because they are over 18 and another was that it meant all those who are expected to work sooner or later – even if they are not working now. Crucially, neither of these interpretations considered working people to be those who are in work at the moment.

In Switzerland, Portugal and the UK there was also evidence that the term held class connotations i.e. respondents were thinking of ‘working people’ in terms of working people who held certain manual occupations. This was not intended in the question design (suggesting a problem with the source question) and was not found in Germany or Spain.17

Stage 6: Conclusions (including application of the error source typology)

At this stage of the analysis, we summarised key findings and linked these to the error source typology identified earlier in this paper.

It is first important to note that there were cases where respondents in all countries had understood the question, the response options and were able to answer this question without any major difficulties. However, as we have seen, this was not always the case and some respondents really struggled. Two main types of error were uncovered at this question.

- Problems with the source question

Item non response, the level of confusion exhibited at this question and indications that respondents (such as those from Spain and Switzerland) felt they were being tested on their knowledge about tax systems all suggest that respondents might benefit from being informed that there are no right or wrong answers to this question. This can be classified as a problem with the source question because this range of problems was found in all countries (although in some more than others). In addition there were instances where respondents asked if proportion was being referred to in option 2 or not. In general, we can conclude that this was evidence of cognitive ‘overload’ rather than a general problem in distinguishing between the two options (although evidence of this was also found in all countries). In short - respondents were simply being asked to do too much. This is clearly a problem with

17 There was a lack of data to draw any conclusions about this in Bulgaria.
the source question and efforts should be made to reduce the complexity of the question and perceived respondent burden.

Finally, the terms 'higher', 'lower', 'high' and 'low' were all used in the wording for this question. This may have, inevitably, confused respondents. The references used in the question should be consistent and refer to one or other of the terms but not to both. Furthermore, evidence form cognitive testing shows that the word 'tax' was unclear to some respondents and the ambiguity of the phrase 'working people' also caused problems leading to differing interpretations of the term including connotations with social class.

- Translation error

In Switzerland respondents were not able to distinguish between options 1 and 2. Since this particular problem was not found elsewhere it suggests a translation error. This was later confirmed by the Swiss team.

The process of cognitive testing using multiple research teams has provided us with an extra level of expert review highlighting errors that were missed in earlier stages.

Stage 7: Recommendations of changes to be made to improve the questions

Based on the conclusions drawn we produced a list of recommendations of changes that could improve the question. These are shown below:

a) Simplify the question and ask respondents to simply choose between options 1 and 2 (delete option 3).
b) Add the percentage symbol (%) to the question – to help reduce confusion about what proportion means.
c) Omit the phrase 'working people' and instead refer to ‘those in paid work’ or ‘earners’ to avoid ambiguity.
d) Be explicit about the type of tax being referred to – if necessary add annotations to clarify this to translators. Annotations are footnotes that are added to the ESS source questionnaire to aid translation. They aim to avoid ambiguity by providing definition and clarifications of the concept behind questions, especially where the words themselves are unlikely to have direct equivalents in other languages.
e) Add an ‘interviewer read out’ instruction to reassure respondents that there are no correct answers to the question. This would be especially important in countries (such as Spain and Switzerland) where knowledge of the tax system was reported to be low.
f) Add a translation note to encourage translators to make sure that the distinction between options 1 and 2 is clear.
g) Only use the terms 'higher' or 'lower' for consistency and to avoid confusion.
h) Consider adding a scenario to the question to help respondents understand the implications of their choice (whilst being aware that the specific example may well have a large impact on respondent choices).

Stage 8: Country verification

The final stage of analysis involved consultation with research teams from the European countries who were asked to verify the findings from their country. Most of the findings from this question were approved as being accurate. One useful correction was provided by the Spanish team. The ESS research team had suggested that respondents had difficulty understanding the difference between options 2 and 3.
because of a translation problem (this conclusion was based on the fact that understanding of these options by Spanish respondents differed so much in comparison with UK respondents). However, the Spanish research team suggested that in fact ‘…there is no a translation problem but a lack of knowledge. Several participants recognised explicitly they didn’t know how the tax system is about in Spain’. This lack of knowledge was present in the interview data and this interpretation was adopted as the conclusion on this matter.

The recommendations from the cognitive testing project as well as findings from a quantitative pilot, consultation with the question designers and expert review comments were all triangulated to produce the final question wording to be used in Round 4 of the ESS. This is shown below.

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 same share (same %) of their earnings in tax so that the person earning twice as much pays double in tax.

The higher earner should pay a higher share (a higher %) of their earnings in tax so the person earning twice as much pays more than double in tax.

They should both pay the same actual amount of money in tax regardless of their different levels of earnings.

(None of these)

Examples of errors

Source question example

We will use the question shown below to illustrate how cross national cognitive interviewing can reveal error with the source question.

Question 7

CARD 4 Firstly, using this card, please tell me how you think most people in [country] would rate the status of those aged 15-29?

It should be noted that other problems were identified through testing this question, such as with the understanding the concept of status. However here we focus on the difficulty respondents from all European countries had with the age group specified in the question: a problem with the source question which could be identified in all countries. The interview protocol was explicitly designed to test for respondent’s ability to generalise across each age group since expert review of the questionnaire prior to this pretesting suggested this might be a problem.
Generally speaking, it appeared that on the surface respondents in all countries were able to give an answer at this question. During cognitive probing however, respondents in the UK, Germany, Portugal, Switzerland and Spain reported difficulty generalising across an age band which was seen as wide and diverse\(^\text{18}\). One respondent in Portugal (E2) for example remarked that “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”. Similarly a respondent in the UK found this question awkward to answer commenting that teenagers, defined by her as 15-19 year olds, are very different to 20-30 year olds. The universal sense that respondents found the task of generalising across these ages mentally challenging resulted in a tendency for respondents to fixate on one end of the age band (normally the lower end).

Respondents tended to base their answer on the younger ages, for example considering “youth”, students or those in education, youngsters who were still supported by their parents and people who had not yet entered the labour force. There was consistent evidence of this answer strategy found among Swiss, Portuguese, Spanish, German, British and Bulgarian respondents; however there was less evidence of lower age group fixation in Bulgaria. On occasion respondents in the UK reported a desire to score younger ages differently to older ages, again supporting evidence that the age band used in the question was too wide and therefore problematic. Additionally, negative stereotypes associated with teenagers in the UK were a contributing factor to the difficulty UK respondents faced when asked to give an overall score for this age group.

In summary, consistent findings across countries point to the fact that there was an inherent problem with the source question which was subsequently replicated through translation.

Discussions surrounding amendments to the age band to make the question easier to answer did not always produce consistent suggestions either within or across countries, highlighting additional challenges the questionnaire design team would face in trying to change the age band specified in the question. For example, would the optimal solution be to (1) eliminate the younger ages (15-19) from the age band altogether to encourage respondents to disregard younger teenagers? (2) change the source question so that it refers to ‘adults under 30’? Or (3) be less specific and refer to ‘people / adults in their 20’s’? The final decision made was to refer to ‘people in their 20s’ because it was thought that this age group would be easier to make generalisations about.

Analysis of this question across countries also revealed that respondents struggled with the instruction in the question, which asked them to say how ‘most people’ in [country] would rate the status of people in the age band. This could be cognitively challenging for respondents or possibly even misheard or overlooked. It should be noted that it was often difficult to ascertain when respondents answers’ reflected what they thought of as ‘most people’s’ view and when they were answering thinking about their own view, or a combination of the two. Findings from the cognitive testing however did suggest that all three strategies were adopted by respondents in all countries and overall there was strong evidence to suggest that the correct strategy (thinking about the views of ‘most people’) was adopted less frequently than expected. Depending on the requirements of the question design team, one solution could be to split the question into two parts: first asking respondents about what they

\(^{18}\) Respondents in Bulgaria did not specifically report this difficulty but there was evidence to suggest that they answered whilst thinking about the younger end of the age group, which could be inferred as a difficulty experienced when generalising.
personally think and secondly asking respondents about what most people think. However, this would add length to the questionnaire and may lead to additional problems related to social desirability and how people would want to be seen in the eyes of the interviewer.

Unfortunately, it was not possible to address this problem for ESS Round 4. Amending the question to ask about a respondent’s own view when they themselves were in that group would be problematic and the question design team confirmed they wanted to measure a societal view. An introduction was added to the question to emphasis that we were interested in ‘...the social statuses that people in different age groups have in society.’

Translation Errors

Errors that are classified as ‘Translation errors’ occur when translated questions are not functionally equivalent to source questions. These can occur as a result of genuine human error or because of difficulties choosing the appropriate phrase or wording meaning that total functional equivalence is not achieved.\textsuperscript{19}

\textit{Human error}

An example from the CI project which illustrates human error in translation is shown below.

\textbf{Question 2}

\textbf{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.

<table>
<thead>
<tr>
<th>Extremely inefficiently</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>(Don’t know)</th>
<th>88</th>
</tr>
</thead>
</table>

It was discovered at the joint analysis meeting that in their translation of this question, the Portuguese team omitted the word ‘income’ from the phrase ‘income tax authorities’. This meant that respondents in Portugal were asked about ‘tax authorities’ in general rather than ‘income tax authorities’ specifically. All of the other participating countries’ translation ensured that the word income was included in the question.

Only Portuguese respondent E2 specifically referred to the income tax authorities in their response, others spoke more generally of the ‘tax authorities’ e.g. respondents E3, E4, E5 and E7. The remaining respondents did not mention the tax authorities specifically but mentioned the services that they provide or the duties they perform. The key problem here is that if Portuguese respondents were not thinking of the same tax collecting authority as respondents in other countries equivalence was probably lost.

\textsuperscript{19} Refer back to the section describing the error source typology for more details about translation errors.
Inappropriate choice of translation

A key example from the cognitive interviewing project which illustrates inappropriate choice of translation is Question 3 (below).

**Question 3**

**CARD 3** 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)

It is first important to consider how the term ‘system of public services was understood by respondents across participating countries. In the UK, this term was largely understood as referring to social support and benefits agency and how they help prevent poverty (GM05, SA02, JF04), social services or security (GM05, GM06) and government run schemes for the homeless (GM06). In other European countries it referred to agencies which provide social help, health insurance, invalidity benefit and state pensions (this was found in Switzerland, Portugal and Spain); the benefits system (found in Bulgaria, Spain and Switzerland) and systems or schemes provided by local authorities or government to help those who need it (found in all European countries). These things were also found amongst the cases in Germany where the welfare system, people who bring food to the elderly or the authorities ‘who help the people’ (respondent 09) were mentioned. However, there was clear evidence that German respondents did not understand this term, which was somewhat surprising and led to closer examination of the German translation for the ‘system of public services’.

The German research team translated the term ‘the system of public services’ as “öffentliche Dienstleistungen”. The literal back translation of this is ‘public services’, which would also work well in British English but in German was either incomprehensible to respondents or associated with ‘public social benefits’ (‘öffentliche Sozialleistungen’). Respondents tended to question what was meant by the chosen translation, suggesting for example, that it “didn’t seem to mean anything in Germany” (respondent 06). In addition, respondent 01 (a highly educated respondent) answered ‘don’t know’ to this question and commented that they could not imagine what ‘öffentliche Dienstleistungen’ could mean.

The German researchers explained that the term ‘öffentliche Dienstleistungen’ is not often used in everyday language. This suggests that it is too formal and may not have been appropriate for use in a survey context. The German team also reported that there was not an equivalently comprehensive term in German to capture the English equivalent of the term ‘the system of public services’. However, the TRAPD approach to translation which was used in Germany does not require identical terms to be used and since Germany does have a public service system it should ultimately have been possible for them to convey this to respondents. It would seem that translators do sometimes struggle with moving away from a ‘word for word’ approach to translation.

The translation errors observed for this question indicate the problems that can be experienced when a directly equivalent term is not found in different countries. It also highlights the importance of taking steps to achieve equivalence in all 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 an ambiguous and quite technical term. This was the eventual solution employed.

Source question & interaction with translation

As noted earlier in this paper, 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 in one or more target languages.

Evidence of this type of error can lead to doubt regarding the ‘portability’ of the item into one or more translated versions. In most cases this type of error would require the source question to be amended although on occasion, additional translation guidance might prove sufficient to help reduce problems.

Question 15

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*?

<table>
<thead>
<tr>
<th>Not at all likely</th>
<th>Extremely likely</th>
<th>(Don’t know)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>88</td>
</tr>
</tbody>
</table>

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

Feedback during the translation stage of the CI project suggested that finding a good translation for ‘moral’ would be difficult. The research teams from the participating countries suggested that ‘moral’ was not generally used as an adjective to describe people and expert review also suggested this was an ‘awkward formulation’.

In general cognitive testing revealed that most countries were able to find a functionally equivalent translation. However in some countries certain dimensions to the meaning in British English were lost and even though this only applied in one or two countries, it suggests there might be wider problems when even more countries have to translate the word.

Testing of this question in the UK revealed that respondents understood ‘moral’ largely as intended, with references made to:

- people having moral codes;
- knowing right from wrong;
- treating others how you would like to be treated;
- and living by society’s rules with a code of conduct.

In Switzerland the word ‘moral’ was translated as ‘to have moral sense’ and although this did on occasion confuse respondents, understanding tended to be consistent across respondents and answers were quite close to those given in the UK. Swiss respondents comprehended the translation as; ‘to behave as you should do and to live
by the law’ with a strong link made between ‘to have moral sense’ and knowing right from wrong.

Similarly, the translation of ‘moral’ in Portugal did not prove to be problematic. The Portuguese word for ‘moral’ is very close to the English word, with a general meaning of ‘to have ethical principles’. Portuguese respondents understood the direct translation of ‘moral’ as a sense of knowing right from wrong, but also reported that they were thinking about religious principles, politeness and respect towards others.

Spain used the word ‘ethical’ to replace ‘moral’ when translating this question as the translators believed this word would match the intended meaning better and indicated that the direct translation “morales” in Spanish was not used in everyday language. Spanish respondents thought about ethical values, to know the difference between good and bad, to be consistent in your opinions and to be responsible and truthful. An ethical person was seen as a responsible person who fulfils their obligations and who knows what good values are. Despite concerns that this would be problematic the cognitive data again suggests broad comparability with the UK data.

Generally, there was equivalence of understanding of the term ‘moral’ between Bulgaria and the source questionnaire. However, an additional dimension not found elsewhere was also considered by some Bulgarian respondents. The Bulgarian team used a direct translation of ‘moral’. Respondents associated moral with obeying written and unwritten social norms and laws and spoke of those who were ‘moral’ as “clean-fingered” (BG04), the well educated, sophisticated and well behaved amongst society. Respondents also talked about people having their own morals (individual) and obeying society’s moral norms (the collective) such as not committing crime, and on occasion suggested that moral was associated with family background and upbringing. This association with education and sophistication was not found in other countries suggesting an additional dimension had been included.

Finally, feedback from Germany at the joint analysis meeting indicated that it was very difficult to find a functionally equivalent word for ‘moral’. A decision was made to use the German word for ‘respectable’ (“anständig”) as it was felt to be closest to the intended meaning of ‘moral’ in English, despite the fact that a direct translation of moral does exist in German. Cognitive testing of this question in Germany revealed that on the whole respondents understood this question in similar ways to respondents in other countries. German respondents talked about respect and compliance with the law and rules, knowledge of etiquette, honesty, decency and politeness. Despite best efforts to achieve equivalence through translation however, it appears that a crucial dimension that most other countries mentioned was lost in Germany; this was related to knowing right from wrong. German respondents seemed to focus on individual’s behaviour, their behaviour towards others in society and compliance with rules necessary when living with others. For example, Respondent 07 said "wie kann ich mich in der Gesellschaft manierlich bewegen" (to be able to act well mannered in society). There was no evidence found in the data from Germany to suggest that moral was understood in terms of an individual’s possession of a set of moral codes or values. It appears that the interaction between the source question and translation in the end meant that a slightly different question was being asked in Germany.

At the joint analysis meeting, countries, other than the UK, concluded that regardless of whether the word ‘moral’ existed in their own languages, it was not the easiest word to use in this context. The direct translation was often not used in everyday language and finding a functionally equivalent term or phrase was a big challenge. Although evidence suggested that it was only in Bulgaria and Germany that
dimensions were added or tested, we feared that this might also have ended up being a problem in other countries who did not participate in the CI project.

One solution, suggested by a few countries, was to replace ‘moral’ with ‘immoral’. However, the Spanish team reported that this has connotations with clothing styles and ways of behaving which might not be intended or equivalent to other countries. An alternative solution would be to ask about having ‘high moral standards’, which overcomes the problem of using moral as an adjective.

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 around ownership of a value set rather than a personal description.

Cultural issues

The ESS procedures specify that question module design team members are drawn from a number of different European countries, involve a multi disciplinary and multi national expert review panel in the question design process and invite research teams in the participating countries to comment on the questionnaire. All of this helps ensure that the instrument is developed with the range of cultures it will eventually be administered within firmly in mind. However, despite such efforts it is always possible that some issues ‘slip through’ these procedures and if pre testing is used during early questionnaire design stages or where cross-national input into the design is limited such issues might arise less frequently.

Fortunately, there were not many examples uncovered in the CI project that suggested there were serious cultural barriers to asking about any of the required concepts in the way proposed. Wherever possible we think that there should be methods built into the cross-national questionnaire design process to highlight such difficulties at an early stage rather than relying on methods like cognitive interviewing to uncover them.

One of the few examples of a cultural barrier to equivalent measurement, which was found during the CI project, is shown below. The test question asked respondents to choose which of the three descriptions of tax collections systems shown on a card they agreed with most.

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

It is important to note that this question caused problems in all countries with the notable exception of Portugal. Item non response and the general confusion reported underline that this is a difficult question dealing with a complex topic. Differing tax systems are not generally a source of discussion amongst the general population of Europe and this was reflected in the narrative around this item.
Whilst respondents’ awareness of the tax system that exists in their country was not a prerequisite for providing a meaningful answer at this question, there was evidence from all countries that respondents did not feel knowledgeable about this area. They therefore either failed to give an answer or provided an answer they had little confidence in.

This problem was particularly pronounced in cultures where knowledge of the tax system was low, most notably Spain and Switzerland. 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 this question. These responses are described in more detail in the section titled ‘Step by Step Analysis’.

As mentioned above having knowledge about your own country’s tax system was not a prerequisite for answering this question. However, feeling confident enough to give an informed opinion is likely to be necessary in order to provide a reliable answer. Split ballot piloting was also used in the UK and Bulgaria to test this question. It suggested there were problems with the reliability of the answers given due to low awareness about tax systems (Coromina & Saris, 2008). The CI project highlighted that in some countries this was likely to be more of a problem than in others, suggesting poorer measurement quality will be evident for this question where awareness of tax issues is lower, making comparisons of uncorrected estimates problematic.

In terms of trying to reduce the effect of this differential knowledge there was a clear recommendation that the question should be simplified, possibly by only asking about the first two answer options. Furthermore, giving some kind of example of what each option means in practice might also help by reducing the emphasis placed on the technical language used. It might also be worth making it clear to respondents that there is no right or wrong answer as this might help to reassure them. In the end it was possible to implement some but not all of these changes.

Conclusions

Designing questionnaires that produce measures with high reliability and validity is a complex and challenging task, particularly in a cross-national setting requiring translation into numerous languages.

Drawing on experience from the development of questions for previous rounds of the ESS, a four item typology of some of the main error sources in cross-national questionnaire design was developed. These concerned the source question, translation errors, the source question and its interaction with translation and cultural issues. Although there are various methods to correct for some of these errors post hoc it would clearly be optimal to be able to address them prior to fielding the questionnaire.

Cognitive interviewing was considered to be an effective method by which to identify such errors and their source prior to fieldwork. However, previous attempts to conduct cognitive interviewing across a wide range of countries, with differing levels of experience of the method and inconsistent practices, have been fraught with methodological and implementation complications that have seriously undermined equivalence and in turn confidence in the process (Miller, 2007). The CI project described in this paper and in Miller et al, (2008) has, with the hindsight gained from
previous studies, addressed many of these difficulties by insisting on a common methodology and total transparency at all stages of the process.

Perhaps most importantly the methodology employed on this project has enabled a more thorough, systematic and transparent analysis to be conducted than has previously been employed on such studies. This is a result of the key features of the analysis approach adopted, which included the use of the Framework analytical approach, Framework data reduction charts being completed in English, initial joint analysis by all participating countries and individual country verification. All of these elements have made it possible for a core group of researchers to undertake an in-depth exploration of the data from all participating countries, developing and testing hypotheses as well as moving between the raw data and higher levels of abstraction. This has facilitated a truly comparative analysis in the later stages of the CI project and reduced reliance on expert review. When comparing findings between countries, a single analysis of data across all countries was used rather than a series of separate analyses that were later harmonised.

Through this process we have found evidence of all four error sources in the typology, although some were more common that others. There were many examples of poor source questions suggesting that cognitive interviewing in the source language should be the first stage of testing prior to translation, otherwise such errors are often ‘replicated’ through the translation process. There were also examples of simple translation errors. It is unlikely that such errors will ever be completely eradicated, however stressing the importance of a committee approach to translation and ensuring it is implemented carefully along with a final ‘copy editing stage’ should minimise the quantity of avoidable errors. We also discovered more subtle translation errors where a technically correct translation had been used but was not understood by respondents. Providing clear translation instructions on specific questions would reduce the prevalence of such errors and cognitive interviewing in multiple languages could help to highlight these prior to final translation and fieldwork. There was evidence too that certain decisions taken when designing source questionnaires can make translation especially difficult and therefore increase the likelihood of lower reliability and validity in non-source language questionnaires. Identifying these elements of a question and where possible adaptations that could be made to the design to remove the difficulty, would undoubtedly improve the quality of the final survey. Again cognitive interviewing provided an effective mechanism by which to identify such difficulties. Finally, we also identified evidence of cultural barriers to the proposed question measures. Although fairly subtle with the specific questions tested in the CI project it was clear that this kind of error would lead to lower reliability in particular countries. Identifying cultural barriers at the earliest stages of design is critical.

Further work to see if the typology could be applied to all problems that were discovered will be necessary and part of this will also consider whether there are examples of problems which fell into more than one category of the error source.

In order to improve the quality of constructs in cross-national questionnaires a wide variety of pretesting techniques need to be employed. We recommend that greater attention to the sources of error in question design that can in turn suggest solutions would be beneficial. Since cognitive interviewing appears to be an effective mechanism for this task it can make a major contribution to improving equivalence in cross-national surveys.
References


