Sailing in unchartered waters: structuring and documenting cross-national questionnaire design

Fitzgerald, Rory

Veröffentlichungsversion / Published Version
Arbeitspapier / working paper

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Leibniz-Institut
für Sozialwissenschaften

GESIS *Papers*

2015|05

Sailing in Unchartered Waters

Structuring and documenting cross-national questionnaire design

Rory Fitzgerald

GESIS-Papers 2015|05

Sailing in Unchartered Waters

Structuring and documenting cross-national questionnaire design

Rory Fitzgerald

GESIS – Leibniz-Institut für Sozialwissenschaften 2015

GESIS-Papers

GESIS – Leibniz-Institut für Sozialwissenschaften
Survey Design and Methodology
European Social Survey
Postfach 12 21 55
68072 Mannheim
Telefon: +49 (0)621 1246 - 249
Telefax: +49 (0)621 1246 - 100
E-Mail: angelika.scheuer@gesis.org

ISSN: 2364-3781 (Online)
Herausgeber,
Druck und Vertrieb: GESIS – Leibniz-Institut für Sozialwissenschaften
Unter Sachsenhausen 6-8, 50667 Köln

Abstract

In recent years there have been important improvements in many areas of cross-national survey methodology. Yet one of the most central areas, questionnaire design, has received less attention. This article introduces, illustrates and evaluates a questionnaire design and documentation template that aims to better structure and document cross-national questionnaire design, largely unchartered territory to date. The article will demonstrate how the template conceptually structures the process of design, facilitates communication between the multiple actors involved and brings cross-national pretesting findings together into a coherent framework. The ways in which the template makes the measurement aims clear and the design process transparent, are also explored.

Keywords: questionnaire design, cross-national survey, metadata, ESS

1 Documenting Questionnaire Design

This article introduces and evaluates an innovative attempt by the European Social Survey (ESS) to better structure and document questionnaire design, a key stage of the cross-national survey lifecycle. The ESS is a biennial cross-national survey that has been conducted in over 30 European countries since 2002.

Prior to introducing the ESS questionnaire design template the rationale for such a design tool is discussed in the context of documenting the entire survey lifecycle. Specific cross-national questionnaire design issues are also discussed. The challenges of documenting questionnaire design are considered and the current availability of questionnaire documentation reviewed.

1.1 Documenting the Survey Lifecycle

The efforts to document questionnaire design discussed in this paper follow on the tail of wider initiatives to document the entire survey lifecycle. Documentation of every stage is required as the survey process is fraught with the potential for introducing error. Secondary data analysts need access to information about the data production process to evaluate if it has reduced precision or caused erroneous conclusions to be drawn. As awareness of the range of potential error sources has grown, for instance with the 'total survey error' approach (Biemer and Lyberg 2003; Lyberg and Biemer 2008), there has been a move away from focusing only on traditional quality indicators, such as response rates, and the adoption of a more holistic approach. This shift has led to demands that the survey production process be recorded, in order that the influence of different error sources can be considered. Starting with the original aims and project design, the ultimate goal is to document every step of the survey lifecycle, through to the publication of the final dataset, metadata and paradata (Mohler, Pennell and Hubbard 2008). Documentation of the questionnaire design phase is a key part of this process.

Providing metadata and paradata is particularly important in cross-national surveys, where the varying contexts, languages, methodological variations and organisational challenges have the potential to have differential impact on the data collected, perhaps reducing 'equivalence' between countries (Jowell, Kaase, Fitzgerald and Eva 2007). There have been improvements in the transparency of many stages of cross-national surveys in recent years with large quantities of metadata and paradata collected and made available (Kolsrud, Skjak and Henrichsen, 2007; Mohler, Pennell and Hubbard 2008; Fitzgerald and Jowell, 2010; Stoop, Billiet, Koch and Fitzgerald 2010). These include project specifications, sample designs, translation documentation, questionnaires in multiple languages, fully labelled survey datasets, data documentation reports, interview paradata and fieldwork quality control reports. In addition there are efforts to standardise the 'data' collected in these areas via the internationally coordinated Data Documentation Initiative (DDI) (http://www.ddialliance.org/what) to promote interoperability across the social sciences.

Gradually increased metadata and paradata have become available for various parts of the survey lifecycle. Surprisingly, this has rarely included questionnaire design information, one key area of interest for most substantive data users. As far back as 1986 Converse and Presser criticised the level of information made available about questionnaire development. They pointed out, for instance, that although pre-testing reports often note the methods used, '...(the) background behind certain concepts, (information) about why certain new questions took the form they did, about why certain well-tried questions from other surveys were preferred to others (was) rarely made clear' (Converse and Presser, 1986: 51). In the questionnaire design field there have been calls for efforts to pool information across surveys, including the whole range of pre-testing methods (Presser et al 2004). Researchers in the US have made some headway here by establishing a data bank of cognitive interviewing pre-testing findings (Miller 2006). Yet even this important resource is limited to cognitive interviewing findings and little else has really changed in recent years (Presser et al, 2004; Harkness, Villar and Edwards 2010).

1.2 The Case for Documenting Questionnaire Design

The questionnaire is at the heart of the social survey and the ability of the instrument to measure with validity and reliability is critical.

In order to inform their work, secondary data analysts require information about the questionnaire design and associated decision making. Basic information might include: which concepts the survey seeks to measure; why certain concepts were excluded; which questions seek to measure the same concept in combination; why questions took the form they did and which questions originated from earlier surveys. Yet often there is little or no information available about how proposals or decisions were arrived at. In addition it often feels like questionnaire design is too often focused on the questions themselves, with insufficient consideration of the overarching measurement objectives and concepts. These discussions become even more important in a cross-national survey where the items have to be asked across very different contexts and where translation is often required from a single language 'source' questionnaire. There is potentially an additional benefit to using structured documentation that can improve the quality of questionnaire design itself. Fowler (1995) argues that the first step in good questionnaire design is to be explicit about the survey content and question objectives, but notes that 'One of the hardest tasks for methodologists is to induce researchers, people who want to collect data, to define their objectives' (Fowler 1995: 9). The template discussed in this article aims to ensure that researchers make their aims explicit something which is common in the natural sciences.

In the past, most publicly funded national surveys were developed by a Principle Investigator, possibly with a team of collaborators, who designed the questionnaire, oversaw the collection of the data, analysed the data and produced the final report. There was therefore less need to document the measurement aims and questionnaire design process, since those performing the analysis had designed the questionnaire themselves. This approach was also adopted by many cross-national surveys. However more recently data from most publicly funded surveys are made available to researchers from outside the original research team and this is often a requirement of the funding. The research undertaken and the data collected are now recognised as public goods, the data have to be made publicly available and '...a serious attempt is at last being made to reduce the barriers between the producers and users...' (Kolsrud et al, 2007: 139).

1.2.1 The Challenges of Documenting Questionnaire Design

Perhaps documentation of questionnaire design has lagged behind documentation of other parts of the survey lifecycle because formally structuring and documenting it is challenging and time-consuming. By its very nature, effective questionnaire design requires a mix of creativity and the adoption of scientifically demonstrated best practice. On the one hand there is accumulated knowledge to guide questionnaire development (for examples see Converse and Presser 1986; Fowler 1995). At the same time each questionnaire is unique, sometimes requiring bespoke solutions. Sudman and Bradburn, for example, argue that even well-established experts cannot produce a 'perfect' questionnaire and will always need to pre-test instruments prior to fielding (Sudman and Bradburn 1982). The iterative and detailed nature of questionnaire design becomes even more challenging when large teams of designers and expert reviewers are involved, often the case in large scale cross-national endeavours.

1.3 Cross-National Questionnaire Design

Most large cross-national surveys develop a source questionnaire in a single language whilst also attempting to account for cross-national issues to aid translation (Harkness, Van de Vijver and Johnson 2003a). On most large scale cross-national surveys the source questionnaire is designed to be 'field ready' in one language and country (as on the ESS, for instance). It is also used as the basis for translation into various different target languages (Harkness 2007). In these situations there are five main questionnaire documentation types that can be produced (Table 1). This article is concerned with the process of documenting cross-national questionnaire development from the initial specification of the measurement aims until the finalisation of the source questionnaire (documentation types 1-4 from Table 1). However the ability of the source questionnaire to facilitate stage 5 is also briefly considered.

Table 1: Questionnaire Documentation Types for cross-national social surveys

	Document Name	Description
1	Source questionnaire	The source questionnaire, used as the basis for translation into all other target languages, should always be made publicly available.
2	Target language questionnaires	All translated questionnaires fielded should always be made available.
3	Measurement aims of the fielded questions	A document that outlines the overall measurement aims of the module, the underlying concepts required to realise the overall measurement aims, the question items intended to tap the concepts and any expected relationships between variables. This does not include development work but simply documents the final questionnaire.
4	Development description	A summary of the design process, starting with the original measurement aims, that goes through the various development and pre-testing stages right up to the point of questionnaire finalisation. Documentation may be limited to decisions taken or might include actions considered and the reasons for rejection. It is 'data' generated about the production of the final questionnaire. It might, for example, demonstrate particular difficulty in operationalising a concept, which in turn highlights a possible source of error.
5	Translation process	The translation process needs to be detailed, demonstrating the ease, or otherwise, of attempting to achieve equivalence with the source questionnaire. Any adaptations required from the source questionnaire should be documented.

1.4 Questionnaire Documentation in Cross-National Surveys

In cross-national surveys, the source questionnaire has traditionally been made available to data users. More recently, cross-national surveys have also started to make questionnaires available in all fielded languages. It is far less common that surveys provide information about the underlying questionnaire measurement aims or the design itself. Amongst the major cross-national surveys it appears only the ESS currently provides this (Based on a search of the websites of SHARE, EVS, WVS, ISSP and EuroBarometer, August 2012).

In terms of questionnaire development, there is a strong case, in public surveys, for providing a description of the measurement aims of the questionnaire in addition to providing the final questionnaires themselves. Some ESS data users for instance have requested this information. Providing the questionnaires and measurement aims ensures that the rationale for including certain questions and the expected interrelationships between them are clear. Arguably it is also important to document the iterations of the design process itself as key design decisions are made. Data users are then better informed as to why certain questions were chosen over others, how pre-test findings influenced the final design of the questions and why certain areas or items were included or excluded. It can also help to explain why existing questions, that might have been validated in a single country study, have (or have not) been adapted for use cross-nationally. In the longer term, such documentation might be included in an 'evidence database' for questionnaires, so that future designers can build on existing information and resources.

How much of the questionnaire design process should be documented? Mohler and colleagues call for a balance between what needs to be curated in case it is lost, what needs to be easily accessible whilst also avoiding a '...lifetime engagement' (Mohler, Pennell and Hubbard 2008:. 406). This article now moves on to outline and evaluate attempts to document the questionnaire design process for modules in the ESS.

2 Questionnaire Design for the ESS

The questionnaire design template was developed for teams developing ESS rotating modules, which are awarded following an open competition. The original application of the successful 'Question module Design Team' (QDT) serves as the starting point for the questionnaire design process and its documentation. The successful teams develop their module in close collaboration with methodological experts from the ESS Core Scientific Team (CST) (Fitzgerald and Jowell 2010).

The precise timetable and pre-testing arrangements vary by round according to the resources available. However it essentially involves a series of iterations of expert review which are complemented by input from national coordinators and a series of pre-tests (including Survey Quality Predictor programme coding, cognitive interviewing and quantitative tests and a full scale 2 nation pilot). The process of development between selection of the QDT and finalisation of the source questionnaire usually lasts around one year or more. In the first three rounds of the ESS, documentation of module development was limited to the original proposal, the final questionnaire and translated questionnaires. In order to improve the process of questionnaire design, to respond to demands from data users for information about the aims of the questionnaire and to respond to the metadata agenda, an ESS questionnaire design template was developed and implemented from Round 4 onwards.

2.1 The ESS Questionnaire Design Template

The ESS questionnaire design template has five key aims (Fitzgerald 2009). The first is to document design from original proposal to final module. The second is to encourage designers to work within a conceptual structure. The third is to facilitate clear communication between the multiple cross-national actors involved. The fourth is to allow pre-testing findings to be considered simultaneously. The fifth is to ensure that data users understand the rationale for asking particular questions, along with the structure and content of the final module. In order to try and realise these aims the template was developed in four steps (ibid):

1) Evidence of documentation provided by other cross-national surveys was sought. No examples of publicly available comprehensive questionnaire documentation were found.

2) A section detailing the overall aims of the module and one that provides information on the likely analysis framework was included. This follows Fowler (1995) who recommends that a good list of question objectives and an analysis plan in produced.

3) Existing literature on approaches to questionnaire design were considered in structuring the template, especially those related to attitudinal measurement, which is a central focus of most ESS rotating modules. In particular the approaches outlined by Fowler (1995: 14), Saris and Gallhofer (2007) and Harkness, Mohler and Van de Vijver (2003b) were influential. Saris and Gallhofer make a distinction between 'concepts by intuition' which can be measured by a single item and 'concepts by postulation' which can only be measured indirectly through multiple question items. In order to structure the template conceptually, separate sections were included for simple concepts (single items, concepts by intuition) and complex concepts (multiple items, concepts by postulation).

4) The structure of the template built on the initial ESS rotating module application form, which has a section on the theory behind the proposed module, a section where the team demonstrate how they propose to achieve their measurement objectives and a section outlining any methodological or practical difficulties foreseen with implementation.

The ESS questionnaire design template for Round 5 is in Appendix A. This empty template is first provided to the QDTs for them to populate. It contains 4 sections: theoretical background; an outline of the concepts to be measured; a section outlining the complex concepts to be measured and a section for simple concepts. Each concept has to be described. Sometimes a complex concept might be constituted from two or more sub-concepts (which in turn can require multiple items to measure them). The QDT is asked to specify the expected relationships between the concepts and the sub-concepts. Finally, the template asks for the proposed items that will measure each concept and sub-concept.

The template is designed in such a way as to keep the development of the proposed questionnaire items close to the text outlining the measurement aims. Feedback from the CST, either individually or collectively, is then added to the template. Sometimes the resulting 'conversation' between the QDT and CST about the questionnaire is also included under the relevant section excluding internal discussions within each of these teams. The aim is to try and record the reasons for taking (or rejecting) major decisions, rather than documenting every option considered. However, what might appear as a minor issue in the early stages of design can become more significant later, making decisions about what to document difficult.

2.2 Examples from the ESS Questionnaire Design Template

In this section, examples from the development of the ESS Round 5 module 'Trust in the Police and Courts' are provided, based on text in the template drafted by the QDT (Jackson et al. 2009) and the CST questionnaire design subgroup[1]. It would be impractical to include the full text from the templates in this article (see www.europeansocialsurvey.org) so summaries have been drafted. However it should be noted that although the template references are to the 'ESS' the intellectual content discussed here is primarily from the Trust in Justice Question module Design Team (ibid) along with input from the ESS Core Scientific Team (CST).

Section A: Theoretical background

In this section, the Trust in the Police and Courts QDT outlined the overall rationale for their module. The module was originally entitled 'Trust in Criminal Justice: A Comparative European Analysis'. The first aim of the module was "...to assess national levels of trust in justice and the legitimacy of (the) legal authorities across Europe" (ESS, 2011a:2). The second aim was to test two different models of cooperation with the criminal justice systems based on the work of Tyler (ibid). The first model states that trust in justice fosters police legitimacy that, in turn, fosters public compliance with the law and cooperation with legal authorities. The second model examines instrumental factors based on rational choice (Jackson et al. 2009). The assumption in this case is that people are governed by self-interest, in the form of sanctions or incentives. People will obey the law when they judge it likely that they would be caught and punished if they (had) committed a crime" (ESS 2011: 1).

The Jackson et al QDT outlined that their approach involved data at three levels. First a small number of primary indicators (level 1) focused on trust and legitimacy and collected at the individual level (categorised as simple concepts in the template). Then a set of secondary indicators (level 2) categorised as complex concepts in the template and again focused on trust and legitimacy and collected at the individual level. Finally level 3 contextual indicators obtained from external non survey sources at the country level to aid interpretation (ESS 2011:2). This information made the measurement aims of the module clear. In addition, outlining this proposed model of analysis facilitated other decisions

[1] Members of this group included Roger Jowell, Rory Fitzgerald, Sally Widdop, Jaak Billiet, Willem Saris, Dorothee Behr and Brita Dorer. National Coordinator input was overseen by Nicolas Sauger.

about the module. In addition the QDT explicitly referenced the work of Tyler in the template (ESS 2011a). Data users can therefore consult this work to see how closely the QDT followed it.

Any key changes to the overall measurement aims are recorded in this section too. For example, during the early development of the 'Trust in Justice' module a change in scope was agreed. It was decided to focus only on the police and courts rather than the entire justice system (e.g. prison services were excluded). The title of the module was amended accordingly and this was recorded in the template (ESS 2011b: 1).

Section B: Overview of the Concepts to be Measured

In this section of the template QDTs are asked to provide an overview of the concepts required and to make explicit anticipated relationships. In certain respects this provides a 'sketch' of an analysis plan, helping to make possible measurement models clear. The Trust in Police and Courts QDT provided the diagram in Figure 1 in their first template (ESS 2011a: 4). The CST queried whether there should be a single model for the police and courts or separate models for each of these institutions. The team confirmed they were not (immediately) planning to combine the institutions into a single model (ESS 2011b: 5).

Section C: Complex Concepts

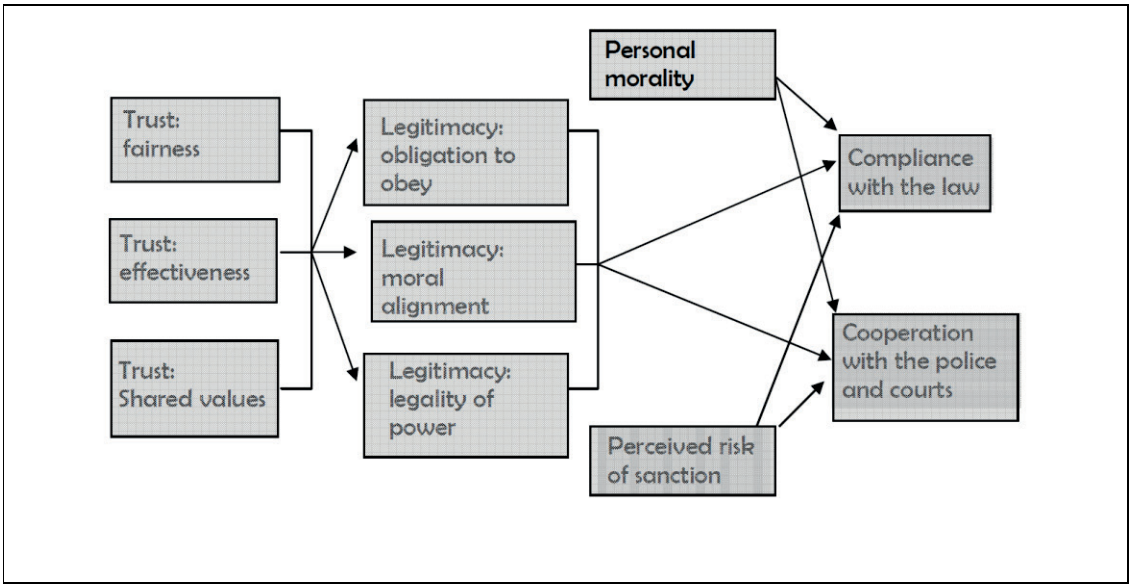
This section summarises the development of the 'Trust in Police distributive fairness' complex concept based on information from numerous templates.

'Stage 1: Conceptual Specification'. The first stage in ESS questionnaire development involves the QDT submitting their proposals for the module, identifying key concepts, definitions and measurement aims. After some initial discussions and requests for clarification, Jackson and his colleagues added the following to the template: "...we also break down trust in the police and trust in the courts into four sub-concepts each of which is measured using multiple indicators: trust in effectiveness, trust in procedural fairness, trust in distributive fairness, and trust in shared values. These are the level-2 indicators of trust in justice" (ESS 2011a: 3).

The QDT named one concept "Trust in Police distributive fairness" (ESS 2011a: 8) and described it as capturing 'the idea that the police treat all members of society equally' (ESS 2011b: 12). In addition background information was provided by the QDT. They outlined that this concept had been developed in previous work (Sunshine and Tyler 2003; Reisig, Bratton and Gertz 2007). Some of the items previously fielded were provided with the two of these loading most strongly onto the factor (Reisig et al 2007). The QDT proposed to operationalise this complex concept with 3 items related to equal treatment, in terms of quality of service, enforcing the law consistently and making sure people receive the outcomes they deserve (ESS 2011a: 8).

Figure 1 summarises the subsequent development of this concept from the first draft questions through to the final items fielded. An overall summary follows.

Figure 1: ESS Police and Courts Measurement Model provided by the Jackson QDT (ESS, 2011a; p4)



Summary of the Development of the Trust in Police Distributive Fairness Concpet

In summary, this complex concept aimed to measure whether respondents trust that the police in their country distribute justice 'equally'. There were some key stages in the development of the concept which are outlined in more detail in Figure 2.

First, the original items from a previous questionnaire on this topic (see Reisig et al 2007) were abandoned, as it was felt that they did not tap the target concept. Instead, the concept was operationalised by asking whether specific groups are treated equally compared to others in related groups. After considering the inclusion of a comparison between the young and old, it was decided to focus only on the rich and poor and those from different ethnic groups, with this change driven by limited questionnaire space. Second, the 'agree / disagree' format used in previous surveys to measure this concept was abandoned, following best practice recommendations (Fowler 1995). Documenting this approach means that data users can understand the rationale for abandoning this traditionally popular approach. Third, the 11 point scale format, frequently used in the ESS, was subsequently abandoned in favour of a more intuitive scale. Data users have a record as to why a nominal question format was used. Fourth, in line with changes across the module, it was decided to ask respondents to consider the police in their country, rather than think only of the local police about their local area. The reason for this decision was documented. The rationale for adding the complex translation annotation for 'police in [country]' was that it will assist data users in understanding the translation adaptations implemented in different countries. This information might also help those designing a cross-national questionnaire on this topic in future. Fifth, the pilot results suggest that there was sufficient quality in the factor structure to include just two items to measure the concept. The documentation makes the pilot findings on this explicit. Finally, following pilot analysis and advance translation, key words were amended and / or annotated.

3 Evaluation of the Questionnaire Design Template

The first aim of the ESS questionnaire design template was to document the process of design from proposal to final module. This has been achieved by using the template to document key decisions. The example in this paper demonstrates how both the module overall (e.g. the change in the module title and reference to which level of police to include) and a particular concept (trust in distributive justice) evolved.

The second aim was to encourage questionnaire designers to work within a conceptual structure. The template facilitates this by keeping the items and discussion about them close together 'physically', encouraging continual cross-referencing. This is particularly critical for a cross-national survey where the underlying concepts need to be well defined.

The third aim was to facilitate clear communication between the multiple actors involved in the process, allowing pre-testing findings to be considered simultaneously. Detailed communication between the different actors can be found within the templates on the ESS website. The template provides a record of earlier discussions, enabling these to inform later design decisions. The fourth aim was to facilitate the triangulation of pre-test findings. The example outlined earlier showed how pre-test findings were considered together.

The fifth aim was to ensure that data users understand the rationale for asking particular questions, along with the structure and content of the final module. This aim was achieved with full documentation of the module (see ESS 2011k) which includes: the original measurement aims of the module; the specification of concepts to meet those measurement aims; the specification of question items to measure the required concepts; the items that will be administered and an outline of how the question items might be analysed.

It is also worth noting that the template has now been used successfully by the ESS CST for a variety of modules with different conceptual approaches. This includes those such as the 'Trust in the Police and Courts' (Round 5) and 'Welfare in a changing Europe' (Round 4), which both have a very structured latent variable based measurement model. The Work, family and well-being module (Round 5) on the other hand was not comprised of a series of latent variables but rather looser conceptual groupings. These have also been comfortably accommodated in the template with the measurement intentions made clear.

The final documentation of the module seems to be of interest to data users. The documentation provided for each of the Round 4 modules, the first round for which this documentation is available, had been downloaded over 7500 times by August 2012. There are however some clear disadvantages with the current paper based template. The large number of template versions (created as the module evolves) and the various iterations included in them make post hoc reconstructions of item development a complex and time consuming process. In addition, the amount of time taken to document each step, both by the QDTs and CST, is demanding and sometimes results in incomplete documentation of certain stages, for example the detailed discussions in meetings. The process becomes particularly time consuming once a draft questionnaire has to be maintained in parallel with the template. In addition, a paper document cannot easily be tailored in order to meet the needs of different stages, for example allowing only certain sections (e.g. only the questions) to be presented. In future it might be beneficial to document only the changes implemented at each iteration and exclude any areas discussed but not acted upon. Finally, technical decisions regarding questionnaire ordering do not fit neatly into the current template and where documented must appear as an annex.

4 Conclusions and Outlook

In the natural sciences the aims of research are stated in advance. This means the criteria for evaluation of the research are made explicit allowing for replication. At the very least, therefore, the final conceptual documentation that has been produced for the ESS to accompany its questionnaires marks a significant step forward and could be considered for wider adoption by large-scale surveys. Smaller surveys or those unable to resource full documentation could document the finally fielded questions in a conceptual structure.

The current paper based ESS questionnaire documentation system may be too cumbersome and time consuming to be sustainable long term. Yet the knowledge collected during the questionnaire development process is too valuable to be lost and needs to be captured to prevent a continual process of 'reinventing the wheel' (Mohler and Johnson 2010: 25). This article has demonstrated a successful attempt to document a process that has often been under reported. So unchartered territory has now at last been mapped. However, the journey is far from over and it is now time to habitualise the documenting of decisions that inform the questionnaire design phase of a survey. The ESS CST has started work on creating an electronic database that documents each step of the questionnaire design process, which will be suitable for the ESS and other large-scale surveys. The new tool will be based on the current question design template but structured to make its outputs usable during the design process, perhaps by producing questionnaires alongside the template or allowing bespoke output files to be produced and printed (e.g. each design step for one concept). Ultimately the ESS CST hopes to link such a databank with those from different stages in the survey lifecycle. This might include translation databanks and databanks containing the finally fielded questionnaire in all languages. However, the ultimate prize for large-scale cross-national surveys would be to enable the end data user to trace the development of a questionnaire item from the original design right through to the (translated) finally fielded item. That prize, however, remains in unchartered territory.

5 References

Biemer, P., & Lyberg, L. (2003). Introduction to Survey Quality. New Jersey: Wiley.

Converse, J., & Presser, S. (1986). Survey Questions. Handcrafting the Standardized Questionnaire. Beverly Hills: Sage.

DDI. http://www.ddialliance.org/what Information abstracted on 19th May 2011.

European Social Survey. (2011a). Round 5 Module on Trust in the Police and Courts - Question Design Template Draft 1. London: Centre for Comparative Social Surveys, City University. Available from www.europeansocailsurvey.org.

European Social Survey. (2011b). Round 5 Module on Trust in the Police and Courts - Question Design Template Draft 1 – with comments. London: Centre for Comparative Social Surveys, City University. Available from www.europeansocailsurvey.org.

European Social Survey. (2011c). Round 5 Module on Trust in the Police and Courts - Question Design Template Draft 2. London: Centre for Comparative Social Surveys, City University. Available from www.europeansocailsurvey.org.

European Social Survey. (2011d). Round 5 Module on Trust in the Police and Courts - Question Design Template Draft 2 - with comments. London: Centre for Comparative Social Surveys, City University. Available from www.europeansocailsurvey.org.

European Social Survey. (2011e). Round 5 Module on Trust in the Police and Courts - Question Design Template Draft 3. London: Centre for Comparative Social Surveys, City University. Available from www.europeansocailsurvey.org.

European Social Survey. (2011f). Round 5 Module on Trust in the Police and Courts – Question Design Template Draft 3 – with comments. London: Centre for Comparative Social Surveys, City University. Available from www.europeansocailsurvey.org.

European Social Survey. (2011g). Round 5 Module on Trust in the Police and Courts – Question Design Template for ESS National Coordinators. London: Centre for Comparative Social Surveys, City University. Available from www.europeansocailsurvey.org.

European Social Survey. (2011h). Round 5 Module on Trust in the Police and Courts – Action points for module following feedback from ESS National Coordinators. London: Centre for Comparative Social Surveys, City University. Available from www.europeansocailsurvey.org.

European Social Survey. (2011i). Round 5 Module on Trust in the Police and Courts – Questions from the Round 5 Pilot Questionnaire. London: Centre for Comparative Social Surveys, City University. Available from www.europeansocailsurvey.org.

European Social Survey. (2011j). Round 5 Module on Trust in the Police and Courts - Question Design Template – with pilot analysis. London: Centre for Comparative Social Surveys, City University. Available from www.europeansocailsurvey.org.

European Social Survey. (2011k). Round 5 Module on Trust in the Police & Courts - Final Question Design Template. London: Centre for Comparative Social Surveys, City University. Available from www.europeansocailsurvey.org.

Fitzgerald, R. (2009). Facilitating triangulation of pretesting techniques through improved documentation of (cross-national) questionnaire development. Paper given at the European Survey Research Association Conference, Warsaw, 29th June-3rd July.

Fitzgerald, R., & Jowell, R. (2010). Measurement equivalence in comparative surveys: The European Social Survey (ESS)—from design to implementation and beyond. In J. Harkness, F. Van deVijver & P. Mohler (eds.), Cross-Cultural Survey Methods. New Jersey: Wiley.

Fowler, F. (1995). Improving Survey Questions. Design and Evaluation. CA: Sage.

Harkness, J., Van de Vijver, F., & Johnson, T. (2003a). Questionnaire design in comparative research. In J. Harkness, F. Van de Vijver & P. Mohler (eds.), Cross-Cultural Survey Methods. New Jersey: Wiley.

Harkness. J., Mohler, P., & Van de Vijver, V. (2003b). Questionnaire design in comparative research. In J. Harkness, F. Van de Vijver & P. Mohler (eds.), Cross-Cultural Survey Methods. New Jersey: Wiley.

Harkness, J. (2007). Improving the comparability of translations. In R. Jowell, C. Roberts, R. Fitzgerald & G. Eva (eds.), Measuring Attitudes Cross-Nationally. London: Sage.

Harkness, J., Villar, A., & Edwards, B. (2010). Translation, Adaptation and Design. In J. Harkness, B. Edwards, M. Braun, T. Johnson, L. Lyberg, P. Mohler, B-E Pennell & T. Smith (eds.), Survey Methods in Multicultural, Multinational, and Multiregional Contexts. New Jersey: Wiley.

Jackson, J., Hough, M., Farrall, S., Aromaa, K., & Keijser, J. (2009). Trust in Criminal Justice: A Comparative European Analysis. Successful module application extracted from www.europeansocialsurvey.org, 20th May 2011.

Jowell, R., Kaase, M., Fitzgerald, R. & Eva, G. (2007). The European Social Survey as a measurement model. In R. Jowell, C. Roberts, R. Fitzgerald & G. Eva (eds.), Measuring Attitudes Cross-Nationally. London: Sage.

Kolsrud, K., Skjak, K., & Henrichsen, B. (2007). Free and immediate access to data. In R. Jowell, C. Roberts, R. Fitzgerald & G. Eva (eds.), Measuring Attitudes Cross-Nationally. London: Sage.

Lyberg, L., & Biemer, P. (2008). Quality assurance and quality control in surveys. In E. de Leeuw, J. Hox & D. Dilman (eds.), International Handbook of Survey Methodology. New York: Taylor and Francis.

Miller, K. (2006). Q-BANK: Development of a Tested-question Database. Proceedings of the Section on Government Statistics, American Statistical Association, (p 1352-1359).

Mohler, P., Pennell B-E., & Hubbard, F. (2008). Survey Documentation: Towards Professional Knowledge Management in Sample Surveys. In E. de Leeuw, J. Hox & D. Dilman (eds.), International Handbook of Survey Methodology. New York: Taylor and Francis.

Mohler, P. & Johnson, T. (2010). Equivalence, Comparability, and Methodological Progress. In J. Harkness, F. Van de Vijver & P. Mohler (eds.), Cross-Cultural SurveyMethods. New Jersey: Wiley.

Presser, S., Couper, M., Lessler, J., Martin, E., Martin, J., Rothgeb, J., & Singer, E. (2004). Methods for Testing and Evaluating Survey Questions. Public Opinion Quarterly 68(1), 109-130.

Reisig, M., Bratton, J. et al (2007). The Construct Validity and Refinement of Process-Based Policing Measures. Criminal Justice and Behaviour, 34,1005-1027.

Saris, W., & Gallhofer, I. (2007). Design, evaluation and Analysis of questionnaires for survey research. NewYork: Wiley.

Sudman, S., & Bradburn, N. (1982). Asking questions: a practical guide to questionnaire design. San Fransisco: Jossey-Bass.

Stoop, I., Billiet, J., Koch, A., & Fitzgerald, R. (2010), Survey nonresponse in Europe: Lessons from the European Social Survey. London, Wiley.

Widdop, S. (2009). Using a Questionnaire Appraisal System to Improve Cross-National Questionnaire Design. Paper given at the European Survey Research Association Conference, Warsaw, 29th June-3rd July.

Figure 2 Concept Development

Stage 1 Questions (ESS, 2011a; p8)

x. To what extent do you agree with these statements about the police in this area?

	Strongly agree	Tend to agree	Neither agree nor disagree	Tend to disagree	Strongly disagree	Don't k
A. Provide the same quality of service to everyone	1	2	3	4	5	6
B. Enforce the law consistently when dealing with all people	1	2	3	4	5	6
C. Make sure people receive the outcomes they deserve under the law	1	2	3	4	5	6

Issues Raised (Expert review by the CST) (ESS, 2011b; p10)	**Changes made for Stage 2**
a) The distributive angle, ie equal treatment of different groups, was not felt to be clear enough in the proposed items. For instance items A and B refer to 'everyone' / 'all people' but this could easily be missed by respondents. The items also were also felt to be too close to those in another concept (eg asking whether the police treat people fairly is quite close to whether they treat people consistently {excluding the possibility of course that they treat everyone unfairly})	New items were proposed focusing on how particular groups are treated in comparison to one another, to tap whether all members of society are treated equally.
b) Use of agree / disagree items critiqued, evidence suggests agree / disagree is a very poor scale (for example see Fowler, 1995; page 55).	Agree / disagree scales dropped.
c) The term 'Police in this area' is likely to cause problems in certain languages and cultures (based on difficulties with translating the term 'area' in previous ESS rounds). This highlighted the more general problem of which type of police are being referred to. The different organisational structures of the police cross-nationally complicated this issue further.	Local police referred to.

Stage 2 Questions (ESS, 2011c; p12)

1) To what extent do you think the local police provide the same quality of service to everyone?

   00 01 02 03 04 05 06 07 08 09 10

2) To what extent do you think the local police enforce the law consistently when dealing with all people?

Or:

3) To what extent do you think the local police treat people from different ethnic groups (or rich and poor?) differently?

   00 01 02 03 04 05 06 07 08 09 10

4) To what extent do you think the local police make sure people receive the outcomes they deserve under the law?

   00 01 02 03 04 05 06 07 08 09 10

Issues raised (Expert review by the CST) (ESS, 2011d; p13-14)	**Changes made for Stage 3**
a) Item 1. Quality of service is a very vague term, not well suited to the police context.	Issue not addressed at this stage.
b) Item 3. The CST like the approach in item 3 and would suggest asking about those in different age groups as well.	This approach, of tapping the concept by comparing the relative distribution of justice to different related groups was adopted going forward.
c) Item 4. It is not the responsibility of the police to deliver outcomes under the law but rather prosecutors and the courts.	Item dropped as this does not tap the concept.

Stage 3 questions (ESS, 2001e; p12)

1) To what extent would you say that the police in this area treat people from different racial or ethnic groups equally or differently? Choose your answer from this card, where 0 means they always treat them equally and 10 means they always treat them differently.

   Always treat them equally — Always treat them differently

   00 01 02 03 04 05 06 07 08 09 10

2) And to what extent do you think the police in this area treat people from the rich and poor differently? Please use the same card.

   Always treat them equally — Always treat them differently

   00 01 02 03 04 05 06 07 08 09 10

3) And to what extent do the police in this area treat young and old people differently? Still use the same card.

   Always treat them equally — Always treat them differently

   00 01 02 03 04 05 06 07 08 09 10

4) Overall, do the police in this area provide the same quality of service to everyone? Choose your answer from this card, where 0 means that the police provide exactly the same quality of service to everyone, and 10 means that they provide a much better service to some people than others.

   Exactly the same quality of service to everyone — Much better service to some people than others

   00 01 02 03 04 05 06 07 08 09 10

Issues raised with stage 3 questions (Expert review by the CST and NCs and use of the ESS Question Appraisal System (Widdop, 2009) and Survey Quality Predictor (see Saris 2007) (ESS, 2011f; p15-17)	Changes made for Stage 4
a) Term 'area' used again. Reason not documented.	Issue not documented at this stage.
b) Q2. 'Treat people from the rich or poor' sounds wrong. Should read 'treat the rich or poor'.	Grammar improved.
c) Q4. Suggest asking either about victims or those accused of a crime. Context is important.	See 'd' below
d) Q4. Quality of service is a very vague term. Perhaps ask about how the police treat people when they are reporting crimes (helps with 'c').	Question focused on how the police treat people when they report crimes.
e) Move away from an 11 point scale so that focus is simply on whether the police treat people equally or not (rather than the extent).	Adoption of non scale answer format.
f) SQP gave a predicted quality of 0.7 for all of these items which is rather good.	N/A as scale changed (see e).

Stage 4 questions – presented at plenary NC meeting (ESS 2011g; p14)

Now some questions about whether or not the police in this area[2] treat different groups of people equally.

D14 When people of different ages report crimes in this area do you think the police treat...**READ OUT...**

...old people better,	01
middle aged people better,	02
young people better,	03
or, are all age groups treated equally?	04
(Don't know)	88

D15 And when people with different incomes report crimes in this area do you think the police treat...**READ OUT...**

...People on high incomes better,	01
on middle incomes better,	02
on low incomes better,	03
or, are all income groups treated equally?	04
(Don't know)	88

D16 And when people from different ethnic groups[3] are reporting crimes in this area do you think the police treat...**READ OUT**

.... people from the majority ethnic group better,	01
from some minority ethnic groups better,	02
all minority ethnic groups better,	03
or, do they treat all ethnic groups equally?	04
(Don't know)	88

Issues raised with Stage 4 Questions (Expert review by CST and NCs) (ESS, 2011h; p7)	Changes made for Stage 5
a) D14, D15: 'middle' categories (02) confusing.	Middle category removed.
b) D15: use 'rich and poor' instead of 'high and low incomes' as more intuitive.	Rich and poor used.
c) Use of 'area' creates translation problems. Also, 'in this area' is not necessarily appropriate. People might not report crimes in the area where they live.	Country of interview used instead of area.

2 Translation annotation: 'In this area' – meaning the respondent's local area or neighbourhood. For respondents living in a town or city this would be a space larger than just the immediately adjacent houses or the street where they live. Countries need to find a phrase that captures this but is not limited to just the immediate surroundings.

3 Translation annotation: Ethnic group can also be thought of as the same race; refer to translation used in the core questionnaire.

Stage 5 Questions (Piloted in the UK and Bulgaria) (ESS, 2001i; p3-4)

Now some questions about whether or not the police in [country] treat different groups of people equally.

B12 When old people and young people report crimes, do you think the police treat …**READ OUT…**

…old people worse,	1
young people worse,	2
or are the old and young treated equally*?*	3
(Don't know)	8

B13 And when the rich and poor report crimes do you think the police treat…**READ OUT…**

…rich people worse,	1
poor people worse,	2
or are the rich and poor treated equally?	3
(Don't know)	8

B14 CARD 8 When people from different race or ethnic groups report crimes, do you think the police treat all groups equally or some groups worse than others? Choose the answer from this card that comes <u>closest</u> to your view.

<u>The police in [country] treat people from:</u>

All minority ethnic groups worse	1
Some minority ethnic groups worse	2
The majority ethnic group worse	3
All ethnic groups treated equally	4
(Don't know)	8

Issues raised with Stage 5 questions (Pilot, Advance translation) (ESS 2011j; 17-20)	**Changes Made for stage 6**
a) Translation error in Bulgaria. The translation used implied reporting by telephone which was problematic when visual contact is cued by the source question.	Annotation added so it is clear that the reporting is done in person.
b) The police 'treating' someone is rather unclear for translators.	Annotation for 'treat' added to provide context.
c) B14. 'Majority ethnic group' caused problems for respondents and translators.	Question reworded to refer to those of the same ethnic group as most [country] people.
d) Bulgarian interviewers reported respondents querying which police to refer to.	Translation annotation added.
e) Need to cut items to meet overall target for module.	B12 cut as less important theoretically. The focus was therefore only on ethnicity and wealth. Factor structure is adequate with fewer items.

Stage 6

The final documentation of the concept was prepared after the source questionnaire was issued and national teams commenced translation. The template was also used to finalise the translation annotations. One consideration for future rounds of the ESS is to provide translators with the final template, as well as the annotated questionnaire.

Final documentation (ESS 2011k; p 13-14)

SUB-CONCEPT NAME: Trust in police distributive fairness

This captures the idea that the police treat all members of society equally.

Expected relationship with other sub-concepts

We hypothesise that trust in distributive fairness will be more important to members of minority groups in predicting perceived legitimacy, compliance with the law and cooperation with the police.

Question item wording

ASK ALL

Now some questions about whether or not the police in [country] treat[4] victims of crime equally. Please answer based on what you have heard or your own experience.

D10 CARD 29 When victims report[5] crimes, do you think the police[6] treat[7] rich people worse, poor people worse, or are rich and poor treated equally? Choose your answer from this card.

Rich people treated worse	1
Poor people treated worse	2
Rich and poor treated equally	3
(Don't know)	8

D11 CARD 30 And when victims report crimes, do you think the police treat[8] some people worse because of their race or ethnic group or is everyone treated equally? Choose your answer from this card.

People from a different race or ethnic group[9] than most [country] people treated worse	1
People from the same race or ethnic group[10] as most [country] people treated worse	2
Everyone treated equally regardless of their race or ethnic group	3
(Don't know)	8

4 Treat in the sense of how the police respond to and deal with people.
5 Report in the sense of 'report in person' so that the police can see them.
6 The final annotation for police was introduced at the start of the module. See the final ESS Round 5 source.
7 Treat in the sense of how the police respond to and deal with people.
8 Treat in the sense of how the police respond to and deal with people.
9 See core questionnaire B36 for a translation of 'people from 'a different race or ethnic group'
10 See core questionnaire item B35 for a translation of 'people from 'the same ethnic group'

APPENDIX A

European Social Survey

ESS Round 5
Question Module Design Template

Module Title:
Module Authors:

SECTION A: Theoretical background
Describe the theoretical background of the module, its aims and objectives

SECTION B. Briefly describe all the concepts to be measured in the module and their expected relationships, either verbally or diagrammatically. Sub concepts do not have to be specified here (these are specified in section C).

SECTION C: Complex Concepts. For each complex concept listed in Section B, describe it in detail and specify the sub concepts as appropriate. Add more boxes to the template as required to describe all the complex concepts sub-concepts

COMPLEX CONCEPT NAME:

Describe the concept in detail, outlining the various sub concepts it comprises
Expected relationship with other complex and simple concepts

SUB CONCEPT NAME:

Describe the first sub concept in detail outlining any further sub concepts or specifying that it can be measured directly
Expected relationship with other sub concepts
Question item wording

SECTION D: Simple Concepts. For each simple concept listed in Section B, describe it in detail here. Add more boxes to the template as required. Once the conceptual structure is agreed with the CCT add the question wording for the proposed item.

SIMPLE CONCEPT NAME:

Describe the concept in detail
Question item wording

Author Note

Rory Fitzgerald has been a member of the ESS Core Scientific Team (CST) since 2004 and became ESS Director in 2012. His key expertise is in cross-national survey methodology, with a focus on questionnaire design, pre-testing, and non-response.

The author would like to thank Debbie Collins and Sally Widdop for their helpful comments on earlier drafts of this paper and the UK Economic and Social Research Council who fund my post.