UNIVERSITY OF SURREY

social researchUpdate

Analysing Cognitive Interview data to Improve Cross-National Survey Questions

Camilla Priede, Elina Ruuskanen, Anniina Jokinen, and Stephen Farrall

Camilla Priede is a Research Assistant in the School of Law, Sheffield University. Elina Ruuskanen and Anniina Jokinen are researchers at the European Institute for Crime Prevention and Control, affiliated with the UN. Stephen Farrall is a professor of criminology in the School of Law, Sheffield University. All work on the FP7-funded Euro-JUSTIS project, designing new survey measures of citizens' confidence in the criminal justice system.

- This *Update* examines the use of Cognitive Interviewing (CI) to improve cross-national surveys.

- During Cognitive Interviews, respondents' understanding of a survey question and the codes presented with them are explored via a series of short probes designed to elicit data on what the respondent thought the question was about, and how they reached their answer.

- A key challenge of cross-national interviewing is to ensure that questions and concepts are understood along similar lines in all countries surveyed.

- CI produces both quantitative and qualitative data that can be used to revise and improve questions in cross-national surveys.

Cross-national social surveys are increasingly used by researchers seeking to make comparisons between countries in terms of social, economic or political trends (Smith 2009). Cross-national surveys however, require careful interpretation of the results; do similar results in different countries mean there is no variation between the countries; likewise do dissimilar results mean people in different countries think and feel differently about the matter in question? Variations may reflect cultural or linguist differences in the interpretation of survey items, rather than substantive ones. Pretesting is one way of assessing the extent to which survey questions have similar meanings in different languages and cultures, and one way of pretesting is to conduct cognitive interviews.

This *Update* outlines the use of Cognitive interviewing (CI) to improve survey questions. We do so via a set of questions which were developed for use in the European Social Survey. There are several works which refer to the method and practice of CI; however few offer guidance on how to analyse the data or to use it to revise survey questions for use in an international survey and so we focus on this aspect of the technique. In this context, CI involves administering draft survey questions to respondents in order to collect verbal information about respondents' thought

processes (Beatty and Willis 2007). This information is used to inform decisions about revisions to survey questions to ensure that they 'work' with all respondents (Conrad, et al 2000; Willis 2005). CI was developed in the 1980s as part of the exploration of the Cognitive Aspects of Survey Methodology movement, and has since become widely used to test survey questions (Tourangeau, et al 2000). It is only recently that CI has been used in cross-national surveys (Fitzgerald et al. 2009; Willis and Zahnd 2007), and there are no clear guidelines for handling the data generated in this way in an international context. In our interviews respondents were asked survey questions followed by a series of verbal probes after each item. Probes were designed to explore elements of a question which may be problematic or understood in different ways, such as the timeframes that respondents were considering and the meanings of specific terms and concepts. For a fuller review of the CI technique refer to Willis (2005).

Both qualitative and quantitative data can be generated from CI. Indeed, for CI to be used successfully to revise survey questions, both qualitative and quantitative data generated should be examined. This is particularly pertinent in the case of cross-national survey research where survey designers face a triple challenge when determining how questions are understood: ensuring that they are understood in the same way by all participants in all situations in all countries.

Cognitive interviews and Euro-JUSTIS

To illustrate CI, we draw on a series of interviews carried out in four countries (England, Finland, Bulgaria and Italy) and designed to test questions about confidence in the criminal justice system, as part of the Euro-JUSTIS project. Data from these interviews were analysed in
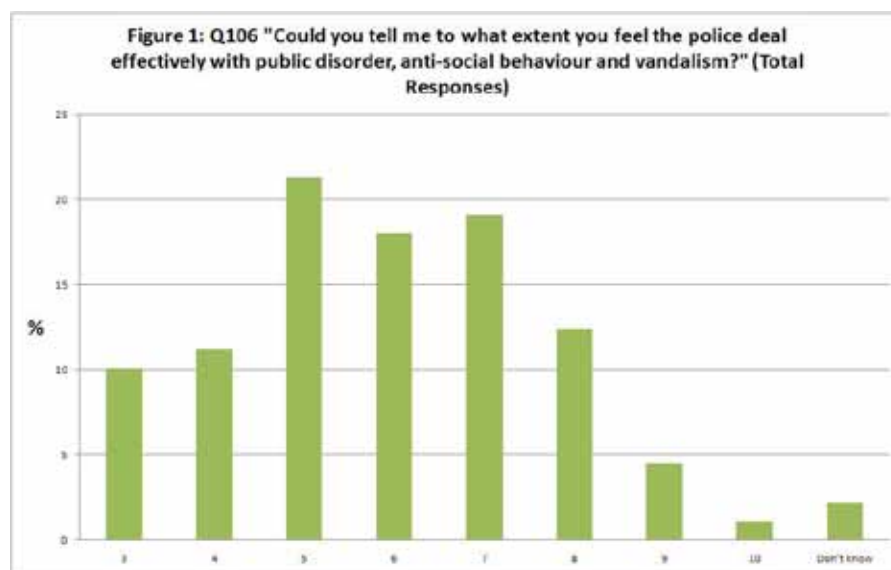
three ways. Firstly, distributions of responses (including non-responses) were examined to assess whether there were any questions that were producing either a skewed distribution or a high proportion of refusals or don't knows. This is standard practice in the pretesting of questionnaires (Presser et al. 2004) and enables the spotting of questions which just do not work.

Additionally, the data were analysed and problems identified with survey questions coded using a framework adapted from one developed by De Maio and Landreth (2004). This allowed for an examination of the number of problems per question per country, and thus provided something of an indication of the extent of problems. Finally the qualitative data was examined to understand more fully the origin of the problems encountered with the survey questions. This analysis also enabled us to ensure that key terms in the survey questions were understood in the same way by all respondents. Sometimes a question, or a term in a question can be understood in a way which is not strictly speaking 'problematic', but is still not the way in which it was intended to be understood, meaning that the question requires revision.

Why different types of data analysis are necessary

There are cultural and linguistic differences that do not necessarily emerge when examining only the quantitative data and in these cases a closer look at the qualitative data is required. Examples of the kinds of issues we encountered are provided below, as well as examples of questions that were altered on the basis of findings from the cognitive interviews. We will illustrate this with reference to one of the questions that was tested. Q106 initially seemed easy for most respondents to answer (Figure 1). The distribution looks normal, and there were few Don't Knows. However, the distribution of responses was somewhat different in England and Finland (Figure 2); the Finns seemed to rate the police as being more effective than the English respondents.

Previous research suggests that Finns have a high confidence in their criminal justice system (Lappi Seppälä et al. 1999), and it could be that they found their police more effective than the English found theirs. However, a closer examination of the scales and people's responses during the CI revealed a more entrenched problem. Question 106 used a 0-10 scale. The Finnish

Figure 1: Q106 "Could you tell me to what extent you feel the police deal effectively with public disorder, anti-social behaviour and vandalism?" (Total Responses)

researchers noticed that for many of the questions which used a 0-10 scale the Finnish respondents tended to use the higher end of the scale. One explanation for this could be the Finns' high confidence, but it is also possible that in the minds of the Finnish respondents, the scale was associated with the one that is used when marking school exams in Finland. In the Finnish school system, a scale from 4-10 is used to rate performance (so that 4 = fail and 10 = excellent). In this context, 5 means that you have passed the exam, but very poorly, whereas in the UK this may be seen as a middling option. There is further evidence on this, as Finnish respondents with somewhat critical views chose answer options 6-7 (which are associated with rather poor performance in school, i.e. passable, or D). This is highlighted in the following quote where a Finnish respondent explains why she picked 7 when answering Q106.

> I don't think they're doing it effectively enough [dealing with public disorder, anti-social behaviour and vandalism]. I think that if the police were more in sight in busy places such as railway and metro stations, it would create a safe atmosphere where it wouldn't even occur to anyone to pee in public elevators, draw graffiti or litter.

Here the data shows that in Finland,

the 0-10 scale is somewhat skewed, and that in reality, the Finns used a 4-10 scale when answering this question. Although the other countries also did not use the lowest three values, it appears that the scale was conceptualised in a different way in Finland to the other countries – so what was middling for other countries was negative for the Finns.

We discovered that Q106 was not completely unproblematic in England either. Examining this question, eight problems relating to comprehension were found in the English data, suggesting that there was some difficulty with how the question was understood by those respondents. One of the issues we probed was the meaning of "public disorder, anti-social behaviour and vandalism". The question was designed to discover how well the participants thought the police dealt with 'low level' crimes, but many English respondents did not understand "public disorder" in this way. When the interviews were carried out, there was controversy in the UK media about how the Metropolitan Police handled the public disorder associated with the G20 meeting held in London in April 2009. Several respondents saw "public disorder" as referring to how the police dealt with demonstrations and thus found

this question hard to answer as it conflated this aspect of policing with less serious offences ("anti-social behaviour and vandalism").

In Finland "public disorder, anti-social behaviour and vandalism" was understood in a fairly uniform way. Few respondents mentioned mass events such as concerts and demonstrations and most interviewees talked about breaking and destroying things, making things or places dirty, urinating in public places, fighting and so on. Thus, using qualitative data from the CI showed that English and Finnish respondents were thinking about somewhat different kinds of actions when answering the question. Being aware of these kinds of differences is important when making international comparisons, and finding out about them required CI and examining the qualitative data.

The data also helped to reveal when questions worked as intended. In our data many of the questions tested worked well and CI did not reveal many problems regarding terms and definitions. The term "respect" was used in questions such as "About how often would you say that the police in this area treat people with respect?" and "How likely do you think it is that the court would treat you with respect?". Respect was understood similarly in all four countries with most respondents interpreting it as something similar to "treating you as a human, being listened to and taken seriously or being professional and polite".

In Bulgaria, however, although there was an equivalent concept of respect, the analysis of the CI data demonstrated that thought has to be given to the particular term which is used in each question to define the type of respect needed. The term "respekt" has become established as a foreign loanword in Bulgarian. For some of the respondents this term was identical with the Bulgarian word "uvazhenie". The term

Figure 2 Q106 UK and Finland responses
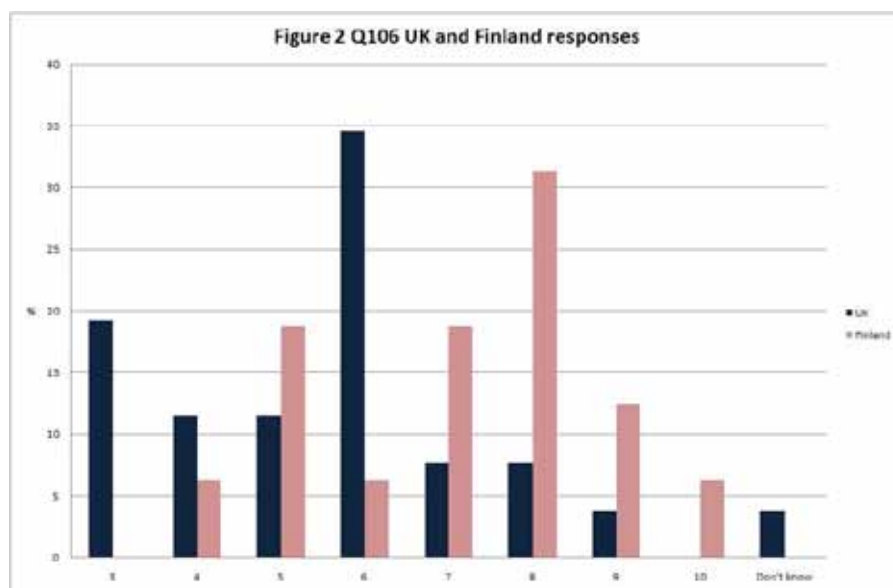
"respekt" when used in Bulgarian as well as meaning "uvazhenie" also implies showing regard for the rights of people and equal treatment of all people, regardless of the situation or their personal characteristics. In some of the questions tested, the two different words for the term "respect" could be used inter-changeably, whereas in others they could not be. Of key importance was the impression that "respekt" is a term related to institutions, while "uvazhenie" refers to individuals. The fact that some Bulgarian respondents distinguished between "respekt" and "uvazhenie" in terms of the legitimacy of institutions suggests that the final questions would need to use "respekt", as this usage is closer to our own interests in the social institutional roles played by the courts and police. Only CI could have brought out these kinds of linguistic differences.

Implications for analysing cognitive interviews

In this *Update* we have illustrated how both quantitative and qualitative data derived from CI can be used when the aim is to create internationally comparable survey questions. To conclude, we suggest the following implications for future research:

- When planning an international survey, carry out CI in as many countries as practicable, because the data may reveal cultural and lingual differences that would otherwise be left undiscovered.

- Pay attention to both quantitative and qualitative data; questions that might seem unproblematic when looking only at the distributions of responses may still contain problems or differences relating to how certain terms are understood by respondents in different countries.

- … and vice versa: small problems with definitions or terms should not always mean the whole question is poor. This is where quantitative data becomes useful.

- Even though the questions tested were developed by quantitative researchers for quantitative research, it is necessary for the qualitative data to be treated as being of equal significance; to do otherwise is to miss out on the holistic picture offered by mixed-methods research.

References

Beatty, Paul C, and Gordon B Willis. 2007. Research Synthesis: The practice of Cognitive Interviewing. *Public Opinion Quarterly* 71 (2):287-311.

Conrad, Frederick G., Johnny Blair, and Elena Tracy. 2000. Verbal Reports are Data! A Theoretical approach to Cognitive Interviews. In Proceedings of the 1999 Federal Committee on Statistical Methodology Research Conference. Washington DC: Office of Management and Budget.

De Maio, Theresa J, and Ashley Landreth. 2004. Do Different Cognitive Interview Techniques Produce Different Results? In *Methods for Testing and Evaluating Survey Questions*, edited by S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin and E. Singer. Chichester: Wiley.

Fitzgerald, Rory, Sally Widdop, Michelle Gray, and Debbie Collins. 2009. Testing for Equivalence using cross-national cognitive interviewing. In Centre for Comparative Social Surveys Working Paper Series: City University London, Centre for Comparative Social Surveys

Lappi-Seppälä, Tapio, Jyrki Tala, Marjukka Litmala, and Risto Jaakkola. 1999. *Public Confidence in Courts. A Study on People's Opinions in 1998*. Helsinki: National Research Institute of Penal Policy

Presser, Stanley, Mick P. Couper, Judith T. Lessler, Elizabeth Martin, Jean Martin, Jennifer M. Rothgeb, and Eleanor Singer. 2004. Methods for Testing and Evaluating Survey Questions In *Methods for Testing and Evaluating Survey Questionnaires* edited by S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin and E. Singer. Hoboken: Wiley.

Smith, T. W. 2009. Editorial: Comparative Survey Research. *International Journal of Public Opinion Research* 21 (3) 267-270.

Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski. 2000. *The Psychology of Survey Response*. Cambridge University Press.

Waddington, P.A.J , and Ray Bull. 2007. Cognitive Interviewing as a Research Technique. *Social Research Update* (50) http://sru.soc.surrey.ac.uk/SRU50.html.

Willis, G., and E. Zahnd. 2007. Questionnaire design from a cross-cultural perspective: An empirical investigation of Koreans and non-Koreans. *Journal of Health Care for the Poor and Underserved* 18 (4 SUPPL.):197-217.

Willis, Gordon B. 2005. *Cognitive Interviewing: A Tool For Improving Questionnaire Design*. London: Sage.

social research UPDATE (ISSN: 1360-7898) is published by the Department of Sociology, University of Surrey Guildford GU2 7XH, United Kingdom. tel: +44 (0)1483689450
Edited by Nigel Gilbert (n.gilbert@surrey.ac.uk)

Spring 2010 © University of Surrey