

Expert workshop on mixed mode data collection in comparative social surveys

16th September 2005

Background

This paper reports on an expert workshop on mixed mode data collection comparative social surveys that formed part of a short-term project funded by the ESRC's National Centre for Research Methods. The workshop was designed as a follow-up to a one-day conference on the same subject held the day before. Both events were hosted by the Centre for Comparative Social Surveys at City University on the 15th and 16th September 2005.

The aim of the workshop was to provide a forum for the key players in the conference to discuss and digest the main issues to have arisen and to identify a focus for future research in the field – especially (though not exclusively) in a cross-national context.

City University's Centre for Comparative Social Surveys, which co-ordinates the European Social Survey (ESS), is, of course, particularly interested in the prospects for mixed mode research in a European context. Research by the survey's Central Co-ordinating Team (CCT) led by Caroline Roberts - in collaboration with Gallup Europe and the University of Essex - has already begun to explore the potential impact of mixing modes on data quality, as well as ways of mitigating that impact. The ESS will have to decide at some point whether all, some or no countries should be able to switch to different modes of data collection for future rounds and, if so, under what conditions. Having recently been awarded a new grant from the EC to address these questions in more detail, a further aim of the workshop was for the ESS team to seek guidance on the design of this research. However, while the focus of this part of discussions would be on the ESS, the overarching aim was to be able to draw conclusions that would be generalisable to other comparative studies.

Workshop Participants

The workshop was attended by guest speakers at the conference and members of the ESS Central Co-ordinating Team and its Methods Advisory Group:

1. **Norman Bradburn (Chair):** Professor Emeritus, University of Chicago and senior fellow at the National Opinion Research Center (NORC). Member of the ESS Methods Group.
2. **Edith de Leeuw:** Associate professor at the Department of Methodology and Statistics at Utrecht University and senior lecturer in Methods and Statistics at the international Utrecht University College and at the Institute for Psychometrics and Sociometrics (IOPS) in The Netherlands.
3. **Don Dillman:** Regents Professor and The Thomas S. Foley Distinguished Professor of Government and Public Policy in the Departments of Sociology, Community and Rural Sociology and the Social and Economic Sciences Research Center at Washington State University in Pullman.
4. **Gillian Eva:** Researcher, Centre for Comparative Social Surveys, City University. Member of the European Social Survey's Central Co-ordinating Team.
5. **Rory Fitzgerald:** Senior Research Fellow, Centre for Comparative Social Surveys, City University. Member of the ESS's CCT.
6. **Robert Groves:** Director of the University of Michigan Survey Research Center, Professor of Sociology at the University of Michigan, Research Professor at its Institute for Social

Research and Research Professor at the Joint Program in Survey Methodology, at the University of Maryland.

7. **Joop Hox:** Professor of Social Science Methodology at the Faculty of Social Sciences of Utrecht University.
8. **Annette Jäckle:** Senior Research Officer, Institute for Social and Economic Research, University of Essex. Consultant to the ESS mixed mode research.
9. **Roger Jowell:** Professor and Director of Centre for Comparative Social Surveys at City University and Principal Investigator of the European Social Survey. Formerly Founder Director of the National Centre for Social Research (NatCen).
10. **Jon Krosnick:** Frederic O. Glover Professor in Humanities and Social Sciences, Professor of Communication, Professor of Political Science and Professor of Psychology at Stanford University, California.
11. **Lars Lyberg:** Statistics Sweden, Chief Editor of the Journal of Official Statistics. Member of the ESS Methods Group.
12. **Peter Lynn:** Professor of Survey Methodology, Institute for Social and Economic Research, University of Essex. Consultant to the ESS mixed mode research. Formerly Head of NatCen's Methods Centre.
13. **Caroline Roberts:** Senior Research Fellow, Centre for Comparative Social Surveys, City University. Member of the ESS's CCT.
14. **Ineke Stoop:** Social and Cultural Planning Bureau, The Netherlands. Member of the ESS's CCT.

Introduction – clarifying the goals of mixed mode research

The workshop was chaired by Norman Bradburn, who opened the proceedings with his own reflections on issues arising from the one-day conference. This led to a discussion of the reasons as to why there had been a recent renewal of interest in research into mode effects. It was attributed to a number of causes. In particular, concerns about rising costs in data collection and the diminishing returns associated with falling response rates have led many to explore the potential benefits of combining different modes. Some National Statistical Institutes and market research agencies, for example, already offer respondents a choice of interview mode, in an effort to encourage survey participation. Meanwhile, survey researchers have had to reconsider the impact of mode on response as new additions to the range of data collection methodologies have become more available. This is particularly true for Internet surveys, which have highlighted the importance of visual design features in self-completion questionnaires for response accuracy and, thereby, reawakened interest in mode effects. In comparative surveys, wide variation in survey practice and the experience of using different modes has added further impetus to the need to explore multi-mode alternatives in studies that have traditionally insisted on a single mode of data collection. At the same time, recent reviews of research in the field (Holbrook, Green, & Krosnick, 2003) have highlighted some of the limitations of earlier research findings on mode effects and with it, the need for more sophisticated designs in mode comparison studies to isolate the causes of differential measurement error.

But an issue that emerged early on in the workshop and which recurred throughout the day was the lack of agreement on what kind of mixed mode future should be pursued and thus, what should be the focus of research endeavours – particularly in relation to mixing modes in comparative studies. In order to develop an agenda for future research into mixing modes of data collection, it is necessary to determine which of the motivators for mixing modes was likely to dominate. In the

short-term, the mostly likely mixed mode scenario for the ESS will be to allow certain countries to switch to telephone interviewing (either instead of, or in addition to face-to-face interviewing). However, longer term, more sophisticated mixed mode designs might be considered, such as sequential designs that minimise costs, yet aim to maximise response rates, or designs that permit respondents to choose their preferred mode of data collection. These different types of mixed mode design indicate different priorities for the methodological research.

Related to this issue, Norman Bradburn highlighted different strategies for different types of research into data collection mode:

- 1) to understand different types of mode effect and to redesign questionnaires to minimise effects;
- 2) to understand better the problems associated with using different channels of communication in different survey modes (i.e. visual vs. aural stimuli);
- 3) to know the size of the effects in order to make decisions about the trade-offs to be made in survey design between different sources of survey error.

For survey designers deciding *whether* to mix modes, each of these strategies is pertinent. Yet for survey designers concerned about *how* to mix modes, it would be important to pursue additional avenues of research – for example, to supplement our understanding of the relationship between mode preferences, response propensities by mode and how these relate to measurement error. During the course of the workshop, this and a number of other gaps in the field were discussed.

Key issues in mode research

Discussions around a number of recurring themes revealed areas in need of new and supplementary studies to build on existing knowledge in the field.

1. Mode preferences and measurement error

Not much is currently known about people's preferences for different data collection modes. What modes would respondents prefer to use when participating in a survey? Meta-analyses of mode preference data have found that people tend to "over-prefer" the mode in which they were interviewed, but when mode of interview is controlled for, there is an overall preference for mail surveys (Groves). It is likely that these findings are now out of date, yet the apparent popularity of the Internet as a mode of data collection may well reflect an overall preference among respondents for self-completion. More research into public attitudes to data collection modes would shed light on this issue and might help guide survey designers in making mode choices.

How do mode preferences relate to measurement error – if at all? Are certain respondents more inclined to give socially desirable responses irrespective of mode? Or are they likely to be more open when responding in their preferred mode of participation? Do "satisficers" have a preference for certain methods over others – for example, telephone interviews, which tend to be conducted at a faster pace and allow them to answer without due reflection? Or do different modes encourage these response effects to different extents?

A further dimension to this issue is the extent to which mode preferences vary by country. Are any cross-national differences that do exist a function of variations in survey practice and the experience of using different modes? Or do they relate more to cultural norms surrounding, for example, cold telephone calling or interactions with strangers knocking on the door. In the ESS context, such cross-national differences have led to the decision in some cases to permit the scheduling of interviews by telephone as a means of securing respondent participation. Would offering respondents a choice of mode at the data collection stage further enhance response rates?

2. Interview length

Offering different survey agencies/ countries or respondents a choice from a range of data collection modes will be a realistic option only once it is known that a questionnaire can practicably be administered in each of the modes on offer. In the case of the ESS, the average face-to-face interview lasts around 60 minutes. This poses a practical barrier to the use of alternative modes, because of the problems of securing a respondent's co-operation in a phone survey of that length, or of motivating someone to self-complete a lengthy paper-and-pencil questionnaire or Internet survey. Not enough is known, however, about the extent to which modes are differentially sensitive to questionnaire length (and people's tolerance of long interviews), so any survey considering the feasibility of mixing modes will need to examine this problem.

Some survey organisations impose a limit on the permissible length of phone interviews (e.g. Gallup's '18 minute' rule). But research has shown that people's willingness to respond to long surveys depends on their motivation and ability to participate which, to a large extent, will vary by survey topic. There may also be cultural variation in tolerance of interview length (e.g. norms regarding the duration of phone calls) and these should be investigated. Statistics Sweden conducted a study of long interviews by telephone in which participants were willing to engage for longer periods than is typically expected of phone interviews, often for over an hour. Using the right techniques, might all respondents be persuaded to make time for phone interviews? Or is tolerance for long phone calls peculiar to Sweden?

What factors cause respondents to break off an interview before it is over? There is some evidence that this might be related to question type (e.g. particularly sensitive questions), to other features of the questionnaire such as visual layout or the load time of certain questions in a web-based questionnaire, or even to natural conversational breaks in the interview. There is also evidence that interviewer motivation is a major contributing factor in maintaining respondents' interest in a survey and preventing break-offs. So studies of interview length should also explore the burden placed on interviewers in different modes and how this impacts on data quality. More generally, the role of the interviewer in survey quality cannot be ignored when considering the underlying causes of different types of measurement error (e.g. how interviewer behaviour influences responses to open-ended questions).

3. Cross-national implementation issues

For comparative surveys like the ESS that are considering mixing modes of data collection (either within or between countries), understanding the interaction between country and mode will be fundamental to future decisions about multi-mode designs. The following areas were identified as needing further research:

a. Variations in norms governing survey interactions

- Cross-national differences in mode preferences (discussed previously);
- Cross-national differences in tolerance for long interviews, in person and by telephone (discussed previously);
- Cultural differences in norms governing interactions with strangers on the doorstep and over the telephone; and other norms relating to the use of different modes for surveys.

How do each of these relate to response propensity by mode?

b. Practical issues

- Cross-national variation in survey costs by mode;
- Cross-national differences in mode penetration – i.e. coverage of different modes, access to survey technology (including subscription to fixed-line and mobile telephones, access to Internet at home and at work, availability of touchtone telephone technology for audio-

CASI). This also relates directly to the question of how to sample for mixed mode surveys. One solution that gets around problems of coverage of the sample frame is to use early interviewer contacts with sampled households to direct selected individuals to the alternative response modes. But given the costs involved in making first contact - around 80% of total costs because of the need for repeated call-backs (Groves) - this seems unlikely to be a sustainable solution for many mixed mode studies hoping to reduce data collection costs.

- Variation in survey agency experience of different modes in different countries: for example, prior to the ESS, face-to-face interviewing was virtually disappearing in certain countries. This had the effect of limiting the choice of field agency. This issue is also relevant to new technologies, for which there may be limited capacity or expertise in certain field agencies and in certain countries. By extension, local practices and survey 'habits' are likely to relate to public familiarity with certain modes, as well as to mode preferences. Resolving the problem of coverage may simply be a case of waiting for new modes (e.g. mobile phones, internet, etc.) to stabilise in different countries before they can realistically be used for surveying, but these issues are also relevant to the question of how much it would cost to establish the necessary infrastructure for a complex multi-mode design on a survey like the ESS.
- What is the demand for switching or mixing modes (or, indeed, willingness to do so) in different countries?

A review of current survey practice and expertise in different countries would be timely (to update and supplement previous studies of this kind, such as that carried out by the Office for National Statistics as part of a review of IALS, the International Adult Literacy Survey (ONS, 2000), and is an essential undertaking before further consideration is given to mixing modes cross-nationally.

c. Implications for measurement error

- Not much is known about cross-national variations in mode-related measurement error. How do country-level factors interact with mode to bias response? Are there discernible differences between countries in terms of the propensity to satisfice in surveys or to offer socially desirable responses to certain types of questions? Do the social desirability connotations of different topics, issues and behaviours vary by country? If so, can we identify patterns to variation?
- How do country-specific norms surrounding survey practices relate to response propensity and measurement error? If the interaction is great, then this may have implications for the type of mixed mode designs that are appropriate, so there are compelling reasons for prioritising this in new research.

4. How to handle measurement error

a. Mitigating mode effects

There were different views expressed on the issue of how to deal with the problem of differential measurement error inherent in mixed mode survey designs (a problem which is exacerbated when a mixed mode survey is carried out in a number of different countries). One view is that by understanding better the underlying causes of response biases, we can design surveys and questionnaires in ways that mitigate their overall impact on data quality. For example, Dillman (2000) advocates a 'unimode' method to develop questions that are functionally equivalent across survey modes (as opposed to being optimised for use in a single mode). But not enough is known about how to achieve this equivalence. For example, what causes comprehension problems and how can they be resolved in each of the different modes? Embellishing questions by adding words can help to clarify meaning in certain modes (e.g. face-to-face, and for some self-completing respondents), but tends to complicate things further in telephone interviews. By contrast, instructing interviewers to probe if respondents' answers seem unclear can in theory help telephone surveys

(and face-to-face), but it is unclear how much it is actually done in the context of a time-pressured phone interview.

Cognitive pre-testing would be one way to highlight questions that are likely to cause problems in certain modes, and research efforts could then be focused on unimode methods of item construction. Developing questions in the primary response mode and then routinely testing them in different modes would help in the effort to develop maximally portable questionnaires. But a further complication of introducing mixed modes to a time-series arises from the existing constraints of the primary mode. For example, the reliance on showcards in ESS face-to-face interviews limits the possibility of using a unimode approach to developing multi-mode questionnaires. There was general agreement among the workshop participants, therefore, that it was unlikely that all measurement error could be dealt with in this way and that a completely mode insensitive instrument was perhaps unachievable.

b. Correcting for mode effects

The alternative position is that research should be focused on developing methods of correcting for measurement error arising from the use of different modes. But how can we judge the validity of responses in different modes? One approach is to look at response accuracy (validating responses against external records), but this is not possible (or at least, very difficult) with attitudinal data. A method for dealing with cross-national variation in measurement error (which relies on meta-analysis of data from MTMM experiments) is being developed on the ESS by Willem Saris and his colleagues. A similar approach might be applied to the problem of mode differences in measurement error. What other approaches exist? Might imputation methods (such as in those used for dealing with non-response) be more appropriate than post-hoc adjustments? For example, when dealing with the problem of selection bias in mixed mode designs that allow respondents to choose their preferred mode, it could be possible to impute how respondents would have answered in a different mode using multiple imputation procedures.

The assumption underlying the second view is that data from a mixed mode ESS would be less user-friendly, to the extent that data users would need some correction factors to handle it. But this is an issue facing many surveys and is already a problem on the ESS because of cross-national differences in measurement error. Assuming that the observed differences between modes were large, it would be necessary to develop standard methods of handling them. Even so, many data users are unfamiliar with the correct methods of using data of this kind, so in a mixed mode future it would be important to spread the appropriate analytic methods.

5. Issues in face-to-face interviewing: the use of showcards

Not much is known about how showcards are used and the purpose they serve in face-to-face interviews. The possibility of using visual aids is generally seen as one of the advantages of in-person interviewing because they are assumed to reduce the cognitive burden placed on respondents (e.g. by removing the need to remember long lists of response options). Yet what do we really know about how showcards are used in the field by both interviewers and respondents? How do respondents read showcards? Do showcards play a more important role in facilitating the interaction between interviewer and respondent (e.g. contributing to rapport building by actively engaging respondents in the interview process)? More generally, how is information communicated in an interview? Certainly, some survey questions only make sense once the response categories are known. Should showcards only contain response options, or should they show questions and answers? Would it help to improve comprehension if the respondent were allowed to read the question as well as the answers? Do showcards increase or lessen the cognitive burden on respondents?

Face-to-face interviews make use of various channels of communication (e.g. oral, visual, as well as body language). But does the mix of aural and visual stimuli help or distract respondents? One way of minimising the occurrence of certain types of mode effect (e.g. response order effects) would be

to harmonise the channels of communication involved in each mode of data collection. First, however, more research is needed to develop best practice in the both the design and use of showcards in face-to-face interviews. This should include an overall review of existing work on showcards, as well as new studies to find out more about how showcards work in the field and whether there are variations in practice cross-nationally.

6. Issues in telephone interviewing

For survey designers considering telephone interviewing, it is necessary to decide whether or not to include mobile phones. Mobile phones present major challenges for sampling (e.g. due to the fact that mobile phone number structures are not fixed). So we need to develop alternative ways of selecting people for mobile phone interviews. This means we need to understand better who has mobile access, how people use mobile phones (for business or pleasure), how it is paid for and so on.

There is also a need to find out more about how practical it is to interview people on mobile phones. Are there problems with reception? Might this impact on comprehension?. What are the burdens on respondents and interviewers? How is mobile phone interviewing currently conducted? (What is protocol? And what actually happens in the field?) This is an area where more could be learned from other countries – for example, mobile phone interviewing is now routine in Finland, but how was the breakthrough achieved?

Who will conduct telephone interviews? In a mixed mode setting, it might be desirable to use the same interviewers for both face-to-face and telephone interviewing. But different skills are required in different modes, so training is important. Research should explore variations between agencies, countries and in the experience of individual interviewers.

Interview length is a crucial variable in conducting telephone interviews successfully. It should be studied from the start as a design feature, so that different solutions for conducting long interviews by phone can be tested. For example, if a long interview does present problems, can it be made shorter? Can it be broken down into chunks? If not, can different people be asked subsets of questions and can imputation methods be used to fill in the missing data?

7. Other issues raised

We need to understand better the non-response mechanisms associated with each mode. For example, non-response in self-completion surveys is often linked to variables of interest. A weakness of face-to-face interviewing is that we get greater non-response in urban populations than in rural ones. Each mode has weaknesses, and we need to be aware of what those weaknesses are.

The move to Internet surveys raises new issues that force us to reconsider how other modes work. For example, the significance of design and visual layout of web surveys has led CAPI programmers to consider ways of improving the screen layout for interviewers (particularly if CASI is to be used in the interview). Standardising screen layout for CAPI and CASI (including web-based instruments) would be an important step in attempts to minimise differences between modes. And this is particularly relevant in cross-national surveys considering the move to web-based data collection: it would not be desirable to allow different countries to use different designs for the Internet version of the questionnaire, so developing a standardised design based on best practice would be essential.

Developing resources for methodological research

In discussing the future research objectives of the ESS with respect to the problem of mixed modes, questions were raised about the current infrastructure for methodological experiments on the survey.

ESS questionnaires are piloted in two countries in each round of the survey. At the main stage of data collection, a supplementary questionnaire is included with the main questionnaire, either to be administered by the interviewer immediately after the main interview, or to be self-completed on paper by the respondent at a later stage. The supplementary questionnaire is currently used for MTMM experiments, designed to measure the reliability and validity of questions by repeating them in different forms, with different response scales. Thus both the pilots and the supplementary questionnaire might provide opportunities in future for methodological research relating to mode. Similarly, verification interviews might provide a further resource for methodological work.

A number of recommendations were made about making resources available on the ESS (as on other time-series) for the purposes of methodological research. One approach would be to declare a permanent test pool in each participating country that could be used as a resource for testing any new idea or proposed change in the survey's design. But alongside this, it would be important to establish a decision protocol for making changes to the core of the questionnaire – i.e. rules about how any proposed changes to the survey's methodology would be implemented – because of the appropriate resistance to change in time-series. For example, one such rule might be to test the impact of any proposed change over the course of two rounds of the survey before it could finally be approved. It is especially important to establish rules of this kind early on, to be prepared for issues as they arise.

Ideas for ESS mode experiments/ studies

In the second half of the workshop, the participants discussed more specific ideas as to how the ESS should use future funds for methodological research on mixed mode data collection. The following is a list of the different suggestions made:

Developing a portrait of mode capacity in different countries

A priority for the ESS is to find out more about survey practice in each of the countries participating in the survey. A 'survey of survey practice' (as in the IALS study) would help to assess the situation in each country and provide invaluable information about e.g. mode preferences, cross-national differences in survey culture, variations in survey costs, motivations for mode mixing, existing infrastructure for collecting data using different modes, and so on.

Investigating the role of showcards in face-to-face interviewing

Because of the potential barrier they would pose to equivalence if the ESS were to allow telephone interviewing, it is essential to build understanding of how showcards are used in face-to-face interviews. The following suggestions for research were made:

- Conduct an overall review of work on showcards and bring field directors together to talk about how showcards are used to find out about potential house effects as well as crossnational variations;
- use para-data from CAPI interviews to find out more about showcard use (e.g. ask interviewers to code whether respondent looked at the showcard (yes or no); record response latencies);
- do an experiment in which half the showcards include questions and responses and the other half just include responses to find out more about how showcards aid comprehension;
- look at the differences between different subgroups on showcard effects (e.g. response order effects) – e.g. those with high and low levels of education; reluctant vs. co-operative respondents;
- find out more about the added value of showcards - e.g. by taking the showcard out of the interview for a few items or for half the respondents in order to assess the impact on response.

Comparing data quality across modes

- For countries using CAPI, it might be possible to test a subset of questions using a web-CASI instrument on the interviewers' laptops (i.e. to make some of the interview like an internet survey). Similarly, switching to CASI would provide a method of testing the extent of social desirability bias in the current ESS face-to-face interviews.
- To begin to explore the problem of interview length, it might be valuable to run a small field experiment to see how people manage when asked to participate in a full ESS interview by telephone. This could include cell phone interviews to see if they differ from fixed line phone interviews.
- Follow-up interviews could be carried out with non-respondents to the main stage face-to-face survey using different modes to explore the impact of offering alternative modes on response rates (e.g. start with the primary face-to-face mode, then move to telephone then web, and/ or postal).
- An alternative would be to do a follow-up of respondents in an alternative mode to test for mode effects within respondents (making it possible to control for selection biases).
- To find out more about social desirability bias across modes it would be useful to include additional questions on the ESS questionnaire that assess the social desirability connotations of different question topics (e.g. respondents' unease about certain topics; asking people what 'most people' would think about x, etc.). It would be necessary to ask the same question in each mode, however, because mode of interview might bias response.

Words of caution

Finally, the workshop discussion gave rise to a number of 'words of caution' for the ESS's CCT, as well as for other survey management teams facing decisions about changes to data collection methodology:

- Agree on the ideal goal for possibly switching data collection mode(s) in the future. Is it to improve response rates? Is it to reduce costs? Is it to accommodate cultural variations in survey practice?
- Be prepared for the impact such a change might have on the time-series.
- Do not repeat the mistakes of other surveys and move to other modes without adequate preparation. For example, ensure that there is some analytic tool for proxy indicators of mode effects. Resist the urge to rush in without proper controls and plan to make the transition over time – e.g. by changing just a subsample to start with.
- Maximise the variation between countries involved in any methodological research based on assumptions about likely cultural differences.
- Be prepared for considerable variation between countries in their experience of and knowledge about different methods and methodological issues (e.g. such as imputation, measurement error, etc.). So make participating countries aware of the issues and motivate them to become interested in what is involved and involve field agencies in the research, as learning more about mixing modes might be relevant to their own development programmes.

References

- Dillman, D. A. (2000). *Mail and internet surveys: The tailored design method* (2nd ed.). New York: John Wiley Co.
- Holbrook, A. L., Green, M. C., & Krosnick, J. A. (2003). Telephone vs. Face-to-Face Interviewing of National Probability Samples With Long Questionnaires: Comparisons of Respondent Satisficing and Social Desirability Response Bias. *Public Opinion Quarterly*, 67, 79-125.
- ONS. (2000). *Measuring adult literacy: the International Adult Literacy Survey in the European Context*. London: Office for National Statistics.

Caroline Roberts - January 2005