

ESSPrep WP6

Assessing the cost-effectiveness of different modes for ESS data collection

Draft Research Proposal

Caroline Roberts, April 2008

1. Introduction

This document outlines a proposal for a feasibility study, involving a trial wave of data collection for the European Social Survey (ESS) using a sequential mixed mode design. This study forms part of the ESS 'Preparatory Phase' (ESSPrep), a programme of work funded under the European Commission's 7th Framework Programme designed to prepare the project for its transition into an upgraded and sustainable infrastructure for the future. For the most part, ESSPrep will focus on the financial, legal and governance work required for an upgraded ESS; in this context, the present research activity is an exception. In seeking to transform the existing ad-hoc funding arrangements that the survey has relied upon until now, and to secure its continuity as a lasting infrastructure, the aim of this workpackage is to assess the cost effectiveness of the existing single-mode data collection strategy (face-to-face interviews in all participating countries) alongside a mixed-mode alternative. The findings of this exercise will feed into decisions as to whether the current fieldwork arrangements could be made more flexible in future rounds of the survey, without seriously compromising the ESS's stringent quality thresholds.

This is a working document, to be developed in collaboration with members of the project team at City University (UK), ESADE (Spain) and Katholieke Universiteit Leuven (Belgium), advisers in the research, including Peter Lynn at the University of Essex (UK) and members of the ESS Methods Group, and the survey organisation who will undertake the fieldwork for this study (who will also become partners in the ESSPrep consortium). The document outlines key elements of the proposed research design and discusses the various decisions that need to be taken in order to finalise this design. The present draft will form the basis of a full description of the final agreed research design, to be submitted to the EC as the first deliverable from this workpackage at the end of June 2008.

2. Background

The work proposed here builds upon an existing programme of methodological research exploring the feasibility of mixing modes of data collection on the ESS, funded by the FP6 ESSi grant and the STREP funding for rounds 1 and 2 of the ESS. This research has involved a number of different elements, including an extensive review of the literature relating to mixed mode data collection (Roberts, 2007); a 'mapping exercise', designed to evaluate the demand for alternative modes on the ESS and the capacity in different countries for conducting fieldwork in alternative or multiple modes (Roberts, Eva and Widdop, 2008); questionnaire development work aimed at minimising the likelihood of mode effects in alternatives to face-to-face

interviewing (see Heerwegh, 2007 and Roberts, 2008); as well as three phases of experimental research¹ designed to assess the sensitivity of the ESS questionnaire to mode effects (in face-to-face and telephone interviews, and paper and web self-completion forms) and to test the feasibility of conducting the full ESS interview by telephone. These studies have been focused on data quality concerns and practical considerations that might pose a barrier to adopting a mixed mode data collection design on a survey like the ESS. They have not enabled us to assess the potential value of a mixed mode data collection strategy.

Mixed mode survey fieldwork – in the form we are considering here, where different members of the same sample answer the same questions in different modes² – has been gaining popularity in recent years. It can take several forms, including one where respondents are offered their own choice of participation mode – referred to as ‘concurrent’ mixed mode data collection (Dillman, 2000), and one where respondents are offered modes according to a hierarchy that varies by costs – referred to as ‘sequential’ mixed mode data collection. The considerable attraction to survey designers of either approach is that they offer the potential to reduce (or at least, better manage) survey costs – for example, by collecting data from a larger proportion of the sample using more cost-effective (self-administered) modes, and reserving the more expensive (interviewer-administered) modes for sample members that are either unable to participate in a self-administered mode, harder to make contact with or harder to persuade to participate. At the same time, despite the danger of increasing the amount of measurement error in the data, combining modes offers the potential to reduce the likelihood of data being affected by nonresponse errors, by providing access to a larger and possibly more representative sample of the population (on the assumption that different modes are differentially favoured by different subgroups of the population).

Attempts to empirically test these claims in favour of using concurrent and sequential mixed mode designs have produced mixed results, however (de Leeuw, 2005). A cursory review of the literature suggests that offering sample members their own choice of mode has little impact on overall response rates and offers no significant advantages in terms of costs. By contrast, sequential mixed mode designs seem to be reasonably effective in terms of obtaining higher response rates, and there is evidence that despite the elevated set-up costs associated with mounting a survey in multiple modes, overall data collection costs can be controlled more easily by directing cooperative respondents towards the cheaper modes. Given that in certain participating countries high fieldwork costs and falling response rates pose significant challenges for the future viability of face-to-face interviewing on the ESS, there appears to be considerable value in considering mixed mode fieldwork (and particularly sequential designs) as a potential solution. Partly because the available evidence is inconclusive, and partly because many of the factors that influence the success of any data collection endeavour are likely to be survey (and country) specific, we feel a full feasibility study testing a mixed mode design alongside the

¹ See Roberts (2008) for a summary of these studies and Jäckle, Roberts and Lynn (2006) for a full report of phase 2. Preliminary reports on phases 1 and 3 are also available on request.

² An alternative mixed mode design under consideration on the ESS is one where a single mode is used in each country, but where national teams are permitted to select the most suitable mode for that country. Coverage problems associated with the main alternatives to face-to-face, however, place severe limitations on this as a viable alternative for the ESS.

standard face-to-face approach is now necessary if the ESS is going to take further steps towards a mixed mode future.

3. Summary of research aims

- i. One of the most critical issues facing an upgraded ESS is whether or not it should permit the use of data collection modes other than face-to-face interviewing in future waves of the project. In this context, the key objective of this workpackage is to build on existing ESS research in this area by conducting a feasibility study, the results of which will inform future decisions.
- ii. The results of the work will help to determine the advantages and disadvantages (in terms of cost and quality considerations) of two alternative data collection strategies: the standard single mode approach by face-to-face interviewing and a sequential mixed-mode design.
- iii. The key objective will be to develop recommendations on the viability or otherwise of using other modes at all and, in particular, of a mixed mode approach.

4. Timetable for the research

The mixed mode study, which will be conducted as an extension of the ESS round 4, has been planned around several phases:

Phase 1 – February – June 2008 (Months 1-5)

- *Select participating country*
- *Develop research design*
- *Preparation of questionnaires*

The immediate priority is to decide where to carry out the research. This decision will be based on a careful evaluation of how best to maximise the available budget in different locations. The rules for reimbursement vary depending on the type of organisation – private- vs. public-funded – and we are limited in our choice of survey houses to those responsible for conducting the face-to-face fieldwork for ESS round 4. We have also limited the range of options to a small number of countries where the benefits to be gained by switching to mixed mode data collection are likely to be greatest (e.g. in terms of a reduction in fieldwork costs). Once decided, the fieldwork house selected will become a partner in ESSPrep, but their role will be limited to this workpackage.

During this stage, we will also undertake a review of previous mixed mode surveys and consult as appropriate about the most viable study design. To this end, we have assembled a small group of technical advisors who will assist in the development of the research design. We shall also be consulting the ESS Methods Group and working closely in collaboration with the survey house.

As soon as the round 4 ESS source questionnaires are finalised (mid-May), work can begin on adapting them and other fieldwork materials for use in a mixed mode context. This work will be building on our existing research.

Phase 2 – July – August 2008 (Months 6 & 7)

- *Finalising questionnaires and other field materials*
- *Fieldwork preparations*

This phase of the study will consist of preparations for fieldwork by the survey agency, including translating the source questionnaire and translating adaptations to the source questionnaires required for mixed mode administration, CAPI/CATI and CAWI programming, and printing of self-completion questionnaires, etc.

Phase 3 – September 2008 – January 2009 (Months 8-12)

- *Survey fieldwork*

Our preference would be to time the fieldwork to coincide with the face-to-face data collection for ESS round 4, the aim being to control as far as possible for influences on our variables of interest (notably, response rates, and components of cost such as the number of contact attempts needed to achieve a completed interview) other than the use of single versus mixed modes. If the fieldwork period is not the same for each survey, then the data collection strategy will be confounded with the timing of the survey, which may exert its own influence on response rates.

Phase 4 – February – December 2009 (Months 13-22)

- *Data processing and delivery to ESS team*
- *Data analysis*
- *Report writing*

The results of the research will feed directly into the Methods Group, Scientific Advisory Board, Funders' Forum and ultimately the full ESSprep Consortium in time for decisions in principle about future rounds to be made and its possible consequences for budgeting to be taken into account. The final report, due in December 2009 (month 22), will be timed specifically to meet this aim.

Data analysis will focus on three main issues (described in greater detail below), including an assessment of the success of the mixed mode design in terms of response rates and the representativeness of the achieved samples (compared with the standard face-to-face approach); an assessment of data quality across the different modes based primarily on the results of built-in MTMM experiments (see below and the appendix); and an evaluation of the cost-efficiency of the two approaches. The analysis will be shared across the three CCT partners (KUL, ESADE and City, respectively).

5. Overview of research design

The basic sequential mixed mode survey design involves inviting sample members to participate in the survey via the most cost-efficient modes first. If the main four modes are to be used, this entails an initial invitation to sample members to participate in a web or mail survey. Non-respondents to this initial contact attempt are then invited to take part in the survey by an interviewer, either on the telephone or in person. The number of steps involved depends on the nature of the survey, the survey population and the information available on the sample frame. For example, if no telephone numbers are available, or if telephone numbers are not available for all sample members, then this step in the sequence will either have to be eliminated or used only for a subset of sample members. Similarly, if a high proportion of the population has internet access, then it makes sense to offer this mode before all others. If, however, only a small proportion of the population has internet access, then it may be better to offer either a web or a postal questionnaire at the first step in the sequence. For these reasons, the precise details of the research design to be used here will need to be agreed in collaboration with the participating survey organisation. This will allow us to tailor the design to the specific requirements of that country, informed by the experiences of the fieldwork agency with mixed mode data collection.

6. Key measures

In order to assess the ‘success’ of the mixed mode data collection design compared with the standard face-to-face survey, we need to agree on a number of indicators with which to measure success by. The following (by no means exhaustive) list of questions identifies some possibilities:

- 1) How do response rates from the mixed mode survey compare with response rates from the round 4 face-to-face survey?
- 2) How does the socio-demographic composition of the samples compare across the two surveys? Do we reach a more representative sample using a mixed mode design? If so, is there evidence of a corresponding reduction in non-response bias?
- 3) How comparable are our estimates taken from the sample as a whole? Do differential mode effects reduce the overall equivalence of estimates across the two surveys?
- 4) How comparable are the estimates obtained in each of the different modes individually?
- 5) How cost effective is the mixed mode fieldwork design, compared with the standard face-to-face approach?

Measuring response rates and sample representativeness

Sequential mixed mode designs have been credited with achieving higher response rates than single mode studies, and helping to reduce the likelihood of nonresponse bias by providing access to a more representative sample of the population. For this reason, one of the most important indicators of the success of our mixed mode design compared with the regular face-to-face survey design will be the overall response rates achieved and measures of the representativeness of the sample together with an

assessment of the degree of non-response bias associated with the achieved sample. Assessments of this kind have been conducted at each round of the ESS by KUL and more recently, in collaboration with SCP, ZUMA and the University of Ljubljana as part of the ESSi Joint Research Activity on 'Improving representativeness and Response'. The present study will require a replication of some of this work, including

- 1) detailed analysis of contact data to compute response rates based on final case dispositions;
- 2) analysis of the composition of achieved samples overall, in each of the individual modes, as well as in comparison with recent population statistics for the participating country;
- 3) comparisons of key estimates that have been shown to be affected by non-response bias in previous rounds of the survey

The most challenging of these tasks is likely to be the calculation of response rates based on final case dispositions because of the added complexity of recording the sequence of contact attempts and the outcomes of those attempts in a mixed mode context. Because of the centrality of this measure in assessing the overall success of the research design, a critical part of the preparations for this study will involve the development of a method for recording the outcome of contact attempts in each mode that meets the requirements already set in the face-to-face ESS. Some progress has been made on this with respect to telephone mode. When we recently ran a trial ESS by telephone in five countries, we developed a list of outcome codes for interviewers to use when recording the outcome of contact attempts. Some organisations experienced difficulties using the call recording procedures specified by the research team (e.g. interviewers found the list of outcome codes overly long; there were problems with integrating a new set of call outcome codes into existing CATI programs, etc.), so some fine-tuning to the procedures used in that study may be necessary – in particular, it will be necessary to work closely with the selected survey organisation to make sure that they can accommodate our requirements into their existing systems. For the two self-completion modes, we will need to develop new procedures for recoding the outcome of our contact attempts (again, in collaboration with the selected fieldwork agency).

Assessing data quality

In attempting to assess the quality of the data collected from each of the two survey designs, we will have two immediate concerns: (1) the extent to which overall estimates from the mixed mode survey are comparable with estimates from the face-to-face survey; and (2) the equivalence of measures in each of the four modes. To address the first of these concerns, we can compare response distributions on key measures (controlling for any sample differences, if necessary). This exercise has a number of limitations, however. Aside from the difficulty of making comparisons where there may be both observed and unobserved differences in the samples (only the former of which we can control for), differences in estimates obtained in each of the individual modes may either exaggerate differences between the two surveys, or conceal them (for example, where mode effects in two modes cancel each other out when the data are pooled). It will be important to take these limitations into consideration when drawing conclusions from our comparisons.

With respect to assessing the equivalence of measures in different modes of data collection our opportunities for analysis will be similarly restricted. Because the study will not involve the random allocation of cooperative respondents to each of the four modes, but rather the self-selection of sample members to what is effectively their choice of mode, our achieved samples for each mode are likely to vary systematically on a number of different variables. These variables may be correlated with variables of interest in the survey, and they may also influence the respondents' predisposition to respond to survey questions in biased ways in interaction with characteristics of the data collection mode. In other words, if we are interested in whether or not estimates are affected by mode effects, we cannot simply compare responses given across each of the four modes, because differential measurement errors will be confounded with nonresponse errors.

The situation is further complicated by the fact that at different stages of the sequence of modes, sample members will have varying response propensities that are conditional on having been previously contacted in the preceding mode(s) in the sequence (and the sequence itself may vary depending on factors such as whether or not a telephone number is available for the sample member). This may further confound the structure of errors in the data – for example, respondents cooperating later in the sequence may be more 'reluctant' than early responders, and this differential motivation may interact with characteristics of the mode to influence the respondent's propensity to adopt certain response sets when answering the questionnaire.

The complexity of these confounded influences on data quality means that we cannot use the proposed research design to draw conclusions about the presence or absence of mode effects in the data. However, some attempt to assess the equivalence of a limited number of ESS measures is possible through the use of the Multi-Trait Multi-Method experiments built into the supplementary questionnaire. The appendix contains a detailed proposal developed by Willem Saris, to include such an experiment in the present study. The research design involves the replication of a number of experiments already tested on the face-to-face survey (and the inclusion of one new one), to allow the evaluation of the quality of ESS measures (i.e. their reliability and validity) across different modes of data collection. The appendix contains full details of the traits and methods to be assessed in the experiment (along with proposed question wording). Note that according to this proposal, the experiment will require a minimum achieved sample of 200 cases in each mode to ensure sufficient precision in estimation.

Assessing the cost effectiveness of different data collection approaches

Although we hope to learn something about the effect on data quality of using a sequential mixed mode fieldwork design for ESS fieldwork (e.g. by looking at response rates and assessing the quality of measures via an MTMM experiment), the primary aim of this study is to assess the cost effectiveness of using such an approach to data collection, compared to the existing single-mode, face-to-face approach. For this purpose, it seems unlikely to be sufficient to simply compare the financial cost of data collection using the two approaches. In fact, focusing purely on financial costs is likely to be problematic for a number of reasons. Firstly, because fieldwork is to be undertaken by the same agency that will conduct the round 4 face-to-face fieldwork,

around the same time as round 4 data collection, there are likely to be considerable savings for the agency in terms of the overall set-up costs of the survey. Thus, any representation of the total costs of the study is unlikely to accurately reflect the costs of setting up either survey from scratch. Secondly, the survey organisation responsible for data collection is unlikely to be willing to disclose full details of the actual costs of the various elements of the study, and even if they were willing to do so, would be unlikely to approve the publication of those details in any output from this study. Thirdly, even with a detailed breakdown of the financial cost of the exercise, the information we will gain will have only limited value in terms of drawing conclusions about the relative costs of different data collection strategies in different countries (or indeed, by different survey organisations within the same country).

For these reasons, we need to develop a list of alternative measures to serve as indicators of ‘cost effectiveness’- portable statistics, which will allow us to draw generalisable conclusions from our research, particularly for the purposes of estimating the likely costs of the same (or a similar) data collection design in another country. The following are examples of the kinds of paradata we should collect in order to be able to draw conclusions about cost efficiency. They focus mainly on measuring the ‘effort’ required to achieve a complete interview (or a completed SAQ by web or mail), but it will also be important to take other factors into account – such as the added costs involved in developing a sampling design that will generate a sufficient number of cases across each mode; the costs of developing a web-based instrument; printing, postage and data entry costs associated with the paper self-completion mode, and so on. By generating a detailed catalogue of the various parameters of the survey’s cost, it will be possible draw conclusions about the likely costs of conducting the survey in a similar way in a different country. We invite suggestions for additions to the list of examples given below:

- Number of attempts to contact sample unit in each mode
- Mode of contact attempt
- Timing of contact attempt
- Number of days since first contact attempt (i.e. overall time to complete case)
- Response, non-contact and refusal rates at different stages of the sequence
- % sample members sent reminder mailings
- % of sample members for which a contact telephone number is available, etc.

7. Sample design

We are proposing to use the same sample design as is currently used on the ESS and hope to achieve around 1500 interviews in total. In order to have sufficient power for the MTMM experiment, we need to achieve a minimum of 200 cases per mode (see appendix). At present we are considering carrying out the study in countries where we can use a frame of individuals, so we do not have to address the question of respondent selection within the household (which becomes problematic with invitations to participate in web and mail surveys).

8. Questionnaire design considerations

In order to field the ESS as a mixed mode survey, we need to develop questionnaires that will work effectively in all data collection modes, and at the same time, minimise the likelihood of mode effects (i.e. differential measurement error across the modes). The ESS mixed mode research has already begun to explore how best to achieve this aim, but there we still have much to learn. This study will provide further opportunities to extend the work we have done so far, so some consideration should be given to questionnaire design issues at the planning stage. The following summarises what we have done so far, and what we will need to do in preparation for fieldwork:

- *Face-to-face questionnaire* – the research we have done to date has been built around the assumption that if mixed mode data collection were introduced on the ESS, no changes would be made to the existing face-to-face questionnaire in order to preserve the continuity of measures in this mode. For this reason, we will use the standard ESS Round 4 face-to-face questionnaire for the purposes of this study. The only exception will be the supplementary questionnaire which will be slightly different from that used on the main survey (see appendix).
- *Telephone questionnaire* – in ESS round 3 we conducted a trial ESS by telephone, experimenting with different versions of the questionnaire (which varied by length and structure). For this research we adapted the full face-to-face questionnaire to make it more suitable for telephone administration. This involved, for example, deleting references to showcards and inserting instructions to interviewers to read out response categories, as well as more substantial changes to questions that rely on more complex showcards in face-to-face mode (notably, a number of socio-demographic measures such as main activity, household income, marital status, and so on). In the present survey, we can use the same source questionnaires for telephone as we used in round 3. The new round 4 rotating modules will need to be adapted following the same principles as for the rest of the questionnaire.
- *Paper self-administered questionnaire (SAQ)* – as part of the ESSi research activity on mixed mode data collection, we will be developing a ‘prototype’ version of the ESS source questionnaire suitable for self-administration. This work is also necessary for the present survey. The design and adaptation of the questionnaire will involve several elements, including eliminating references to showcards, interviewer instructions and rewriting questions to make them suitable for self-administration; making decisions about how to maximise the equivalence of the SAQ with the other modes – e.g. concerning Don’t Know response options that are not explicitly offered in face-to-face mode even though a DK response will still be recorded by the interviewer; and decisions about design and layout in order to enhance the usability of the questionnaire and minimise errors by the respondent (e.g. by ensuring that routing/skip patterns are easy to follow). The process of adapting the questionnaire should draw on the literature in the area of designing paper and web questionnaires and questionnaires for mixed mode surveys (e.g. Dillman, Gertseva and Mahon-Haft, 2005; Dillman, 2007).
- *Web questionnaire* – once we have adapted the source questionnaire for self-completion, the same questionnaire can form the basis of the web

survey instrument. As part of the ESSi workpackage, we have already developed a prototype web version of the source questionnaire, which involved making decisions about how to enhance comparability with the face-to-face instrument (see Heerwegh, 2007). We should draw on this work to develop specifications for the CAWI programmers in the participating survey organisation, as well as when developing the paper SAQ. We should work closely with the developers to ensure the final design meets our specifications.

Two further issues relating to questionnaire design are worth considering at this stage:

- 1) Whether or not to split the questionnaire into two or more parts –the ESS face-to-face interview lasts around 1 hour, which may prove to be especially burdensome for respondents in the other modes (there is also a danger that if the first contact attempt involves sending sample members the paper SAQ, they will be put off by the length of the form). In the round 3 telephone survey, we experimented with splitting the questionnaire into 2 parts. Preliminary findings suggest this had a negative impact on both response rates and data quality. However, we have no experience of fielding the full questionnaire in a self-completion mode, where respondent burden will be especially heavy and may lead to break-offs or a deterioration in data quality. Anybody wishing to try filling in the ESS source questionnaire in web mode can visit: https://www.kuleuven.be/dvz/survey/026/login_EN.php. It takes around 30-40 minutes to complete!
- 2) Whether or not to use the mixed mode survey as an opportunity to test different versions of questions that seem likely to generate mode effects. Our opportunities for doing this are likely to be limited given the lack of randomisation in the survey design (although, see appendix for details of the proposed MTMM experiment).

9. Specifications for fieldwork procedures

In addition to preparing the questionnaires, we need to make decisions about each of the following:

- Whether to offer incentives and how to administer them in each of the different modes
- How to monitor and record the contact procedure in each mode (i.e. develop lists of contact outcome codes for each mode, as well as a list of final disposition codes that captures all possible outcomes across the different modes)
- The maximum number of contact attempts at different steps in the sequence of modes and procedures for switching modes (i.e. how to decide when to switch from one mode to another)
- How to manage refusals, how to decide when a refusal is soft/hard in different modes and whether a refusing case should be re-contacted in a different mode
- How to combine data from different modes (i.e. develop a data protocol for the interview and contact data).

10. Milestones and Deliverables

We have specified the following milestones and deliverables in the Description of Work (where Month 1 is February 2008):

- Research proposal for project team (Month 5)
- Questionnaires and fieldwork documents (Month 5³)
- Final report (Month 22)

Milestone no.	Delivery month	Comments
1	4	Research design agreed for mode survey
2	7	Source questionnaires and field materials for mixed mode survey delivered to field agency.
4	9	Start of fieldwork
8	14	Data processing complete. Data file to project team

Caroline Roberts – April 2008

11. References

- de Leeuw, E. D. (2005). To Mix or Not to Mix Data Collection Modes in Surveys. *Journal of Official Statistics*, 21(2), 233-255.
- Dillman, D. A. (2000). *Mail and internet surveys: The tailored design method* (2nd ed.). New York: John Wiley Co.
- Dillman, D. A. 2007. *Mail and internet Surveys: The tailored design method with internet, visual and mixed mode guide* (Second Edition). New York: John Wiley.
- Dillman, D. A., Gertseva, A. and Mahon-Haft, T. (2005). Achieving usability in establishment surveys through the application of visual design principles. *Journal of Official Statistics*, Vol. 21, No. 2, pp.183-214.
- Heerwegh, D. (2007). *JRA1 Task 2 Prototype web questionnaire*. Deliverable prepared for ESSi JRA1, available on request from the Centre for Comparative Social Surveys, City University, London.
- Jäckle, A., Roberts, C. E., and Lynn, P. (2006). Telephone versus Face-to-Face Interviewing: Mode Effects on Data Quality and Likely Causes (Report on Phase II of the ESS-Gallup Mixed Mode Methodology Project). *ISER Working Paper*, 2006-41. Colchester: University of Essex.
<http://www.iser.essex.ac.uk/pubs/workpaps/pdf/2006-41.pdf>
- Roberts, C. E. (2007). Mixing Modes of Data Collection in Surveys. *ESRC National Centre for Research Methods Review Papers*, NCRM/008.
<http://www.ncrm.ac.uk/publications/methodsreview/MethodsReviewPaperNCRM-008.pdf>

³ This is likely to be delayed as the round 4 source questionnaires will not be available until the end of May.

- Roberts, C.E.(2008). *Designing equivalent questionnaires for a mixed mode European Social Survey: Report on the findings of ESS mode experiments.* . Deliverable prepared for ESSi JRA1, available on request from the Centre for Comparative Social Surveys, City University, London.
- Roberts, C. E., Eva, G., and Widdop, S. 2008. *Assessing the Demand and Capacity for Mixing Modes of Data Collection on the European Social Survey: Final Report of the Mapping Exercise.* Deliverable prepared for ESSi JRA1, available on request from the Centre for Comparative Social Surveys, City University, London.

12. Appendix

Proposal for the evaluation the quality of the mixed mode data collection

In order to start the discussion about the criteria to evaluate the quality of the mixed mode data collection I make here a proposal for the evaluation of the quality of the questions. I restrict me to this topics even though I recognize that other aspects should also be evaluated such as the costs, the number of refusals, response rates, item nonresponse, response distributions between the mixed mode data and the face to face data. According to me measurement in the same period is necessary for all these criteria. I leave it to others to develop the evaluation of these criteria. I will concentrate on the evaluation of the quality of some questions and concepts. This can be done with MTMM experiments which have been used before as I will indicate in the next section. If we agree on this point then the main task is to select the topics for the supplementary questionnaire to be evaluated. This will be discussed in the second section. In section 3 I will propose the different sets of measures.

1. The Use of MTMM experiments

In order to determine the quality of single questions and concepts MTMM experiments have been used in the ESS. In each round 6 experiments have been done. Based on these experiments the quality of the questions has been studied (Saris and Gallhofer 2003, 2005, 2007). Recently several studies have been finished which evaluate the quality of the concepts proposed for the Core questionnaire of the ESS. The quality of the measurement of the two measures for Political efficacy have been evaluated by Saris 2006). The measurement of media use has been evaluated by Coromina and Saris (2008), The quality of the value scales ahs been studied by Davidov el al. (2002), Davidov (2007) and Knoppen and Saris (2008). Other concepts are at this moment evaluated.

These evaluations include the comparison of the quality of the questions and concepts across countries. The comparison of the quality of questions and concepts across modes of data collection can be done in exactly the same way. If in each mode the same MTMM experiments are done, one can compare the quality of the measures across modes of data collection. This requires that the sample sizes in all modes should be larger than 200 in order to have enough precision in estimation. If the latter requirement is satisfied it only means that in all modes of data collection, besides the main questionnaire, also the supplementary questionnaire has to be provided to the respondents. We suggest to use for this purpose the same design as we normally do i.e. the Splitt ballot MTMM design which reduces the response burden for the respondents considerably where 2 methods concern different forms of the response scale while the third one is the mode of data collection.

2. The choice of the topics for the experiments

In the past many experiments have already been done. In these experiments some important effects of method variations have been detected. However experiments with the mode of data collection were rather limited except between face to face in the

main questionnaire and self completion in the supplementary questionnaire. It would be important to see if the same effects are found for the different modes of data collection.

2.1 Use of batteries of statements

It has been found that batteries with statements and agree/disagree scales were much worse in the face to face mode than in the self completion mode. This was found for the socio – political orientations asked in the core questionnaire. We think that it would be good to repeat this experiment given that now different forms of self completion questionnaire and personal interviews will be used.

2.2. Open or closed questions of frequencies or amounts

It was found that open questions with respect to amounts or frequencies were much less good than closed question specifying numeric categories. This was found for media use and division of home work. Since we do not know if this result generalizes to all different modes of data collection it would be good to test this. We suggest to do this for media use because that is the concept that is the most evaluated so far.

2.3/4. Number of categories

We have found that the number of categories has a strong effect on the data quality. This became for example clear in the correlations between political trust and social trust. Since then the favourite scale in the ESS is the 11 point scale. Question is if that is still the case if we use different modes of data collection and if the qualities in the different modes are the same. For this we should repeat a test for the topics “social trust” and “political trust” because for the latter topic the results were different because the stimuli were shorter and the labels of the scales as well and the scale was unipolar.

2.5. Trait specific scales versus agree disagree items

It has been shown for several items that trait specific scales are much better than agree disagree scales. A very convincing case was received for the different political efficacy scales. Unfortunately the items that remained for political efficacy are not best possible (Saris 2007). I suggest to repair the instrument in this study but this requires also an increase of the core questionnaire once with 5 questions (see below).

2.6. Test of the left right scale

The core contains the standard left right scale which has been used so far in most election studies. However we do not know how good this scale is and if the responses on this scale can compared. There are arguments made that in German there is a social desirability bias in the direction of the left because of the Nazi past while in the East European countries a bias to the right can be expected because of their communist past (Weber 2007). This is typical a sensitive topic that should be studied moving from interviewer administered interviews to self completion data collection.

3. The concretization of the experiments

Having mentioned the different experiments which we think would be useful, the plans are worked out in more detail in this section. Although we will use in the mixed mode experiments only two forms of the questions we add a third form of the questions in this overview because we will make an argument to do the same

experiments extended with one extra question form in the main research of the 4th round (see next section).

3.1 Use of batteries of statements

To compare the quality of batteries with statements and agree/disagree scales with normal questions in different modes of data collection, we suggested to use the items for the socio – political orientations asked in the core questionnaire. Below we specify the different forms for the different modes.

This form for the main questionnaire

CARD 16 Using this card, please say to what extent you agree or disagree with each of the following statements. **READ OUT EACH STATEMENT AND CODE IN GRID**

	Agree strongly	Agree	Neither agree nor disagree	Disagree	Disagree strongly	(Don't know)
B43 The less that government intervenes in the economy, the better it is for [country]	1	2	3	4	5	8
B44 The government should take measures to reduce differences in income levels	1	2	3	4	5	8
B45 Employees need strong trade unions to protect their working conditions and wages	1	2	3	4	5	8

First alternative form

Please indicate to what extent you agree or disagree with each of the following statements.

HS16 "The less that government intervenes in the economy, the better it is for [country]"
Please tick one box.

Agree strongly 1

Agree 2

Neither agree nor disagree 3

Disagree 4

Disagree strongly 5

HS17 “The government should take measures to reduce differences in income levels”.
Please tick one box.

Agree strongly 1

Agree 2

Neither agree nor disagree 3

Disagree 4

Disagree strongly 5

HS18 “Employees need strong trade unions to protect their working conditions and wages”.
Please tick one box.

Agree strongly 1

Agree 2

Neither agree nor disagree 3

Disagree 4

Disagree strongly 5

Second alternative form

HS34 Is it generally good for [country] if government intervenes less in the economy? **Please tick one box.**

Definitely 1

Probably 2

Not sure either way 3

Probably not 4

Definitely not 5

HS35 Should the government take measures to reduce differences in income levels ? **Please tick one box.**

Definitely 1

Probably 2

Not sure either way 3

Probably not 4

Definitely not 5

HS36 Do employees need strong trade unions to protect their working conditions and wages ? **Please tick one box.**

Definitely 1

Probably 2

Not sure either way 3

Probably not 4

Definitely not 5

3.2. Open or closed questions of frequencies or amounts

It was found that open questions with respect to amounts or frequencies were much worse than closed questions specified with numeric categories. This was found for media use and division of home work. We suggest to do this experiment for media use because that is the concept that is the most evaluated so far and we have not complete data for all countries about these variables.

The form for the main questionnaire

A1 CARD 1 On an average weekday, how much time, in total, do you spend watching television? Please use this card to answer.

No time at all	00	GO TO A3
Less than ½ hour	01	
½ hour to 1 hour	02	
More than 1 hour, up to 1½ hours	03	
More than 1½ hours, up to 2 hours	04	ASK A2
More than 2 hours, up to 2½ hours	05	
More than 2½ hours, up to 3 hours	06	
More than 3 hours	07	
Don't know	88	

A2 STILL CARD 1 And again on an average weekday, how much of your time watching television is spent watching **news** or programmes about **politics and current affairs**? Still use this card.

No time at all	00
Less than ½ hour	01
½ hour to 1 hour	02
More than 1 hour, up to 1½ hours	03
More than 1½ hours, up to 2 hours	04
More than 2 hours, up to 2½ hours	05
More than 2½ hours, up to 3 hours	06
More than 3 hours	07
(Don't know)	88

CARD 1 On an average weekday, how much time, in total, do you spend listening to the radio? Use the same card.

- No time at all
- Less than ½ hour
- ½ hour to 1 hour
- More than 1 hour, up to 1½ hours
- More than 1½ hours, up to 2 hours
- More than 2 hours, up to 2½ hours
- More than 2½ hours, up to 3 hours
- More than 3 hours
- (Don't know)

00	GO TO A5
01	
02	
03	
04	
05	ASK A4
06	
07	
88	

A4 STILL CARD 1 And again on an average weekday, how much of your time listening to the radio is spent listening to **news** or programmes about **politics and current affairs**? Still use this card.

- No time at all 00
- Less than ½ hour 01
- ½ hour to 1 hour 02
- More than 1 hour, up to 1½ hours 03
- More than 1½ hours, up to 2 hours 04
- More than 2 hours, up to 2½ hours 05
- More than 2½ hours, up to 3 hours 06
- More than 3 hours 07
- (Don't know) 88

ASK ALL

A5 STILL CARD 1 On an average weekday, how much time, in total, do you spend reading the newspapers? Use this card again

No time at all	00	GO TO A7
Less than ½ hour	01	
½ hour to 1 hour	02	
More than 1 hour, up to 1½ hours	03	
More than 1½ hours, up to 2 hours	04	
More than 2 hours, up to 2½ hours	05	ASK A6
More than 2½ hours, up to 3 hours	06	
More than 3 hours	07	
(Don't know)	88	

A6 STILL CARD 1 And how much of this time is spent reading about **politics and current affairs**? Still use this card.

No time at all	00
Less than ½ hour	01
½ hour to 1 hour	02
More than 1 hour, up to 1½ hours	03
More than 1½ hours, up to 2 hours	04
More than 2 hours, up to 2½ hours	05
More than 2½ hours, up to 3 hours	06
More than 3 hours	07
(Don't know)	88

First alternative form

HS1 On an average weekday, how much time, in total, do you spend watching television?

WRITE IN HOURS: **AND MINUTES:**

HS2 On an average weekday, how much time, in total, do you spend listening to the radio?

WRITE IN HOURS: **AND MINUTES:**

HS3 On an average weekday, how much time, in total, do you spend reading the newspapers?

WRITE IN HOURS: **AND MINUTES:**

Second alternative form

HS19 On an average weekday, how much time, in total, do you spend watching television?
Please tick one box.

No time at all ₀₁

Very little time ₀₂

A little time ₀₃

Some time ₀₄

Quite a lot of time ₀₅

A lot of time ₀₆

A great deal of time ₀₇

HS20 On an average weekday, how much time, in total, do you spend listening to the radio?
Please tick one box.

- No time at all 01
- Very little time 02
- A little time 03
- Some time 04
- Quite a lot of time 05
- A lot of time 06
- A great deal of time 07

HS21 On an average weekday, how much time, in total, do you spend reading the newspapers?
Please tick one box.

- No time at all 01
- Very little time 02
- A little time 03
- Some time 04
- Quite a lot of time 05
- A lot of time 06
- A great deal of time 07

3.3/4. Number of categories

We have found that the number of categories has a strong effect on the data quality. Question is if that is still the case if we use different modes of data collection like telephone interviewing without show cards. We suggest to repeat a test with the social trust and political efficacy because for these two topics the results were different because the stimuli were shorter and the labels of the scales as well and for the later topic the scale is unipolar in stead of bipolar.

3.3.The form for the main questionnaire

A8 CARD 3: Using this card, generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people? Please tell me on a score of 0 to 10, where 0 means you can't be too careful and 10 means that most people can be trusted.

<i>You can't be too careful</i>											<i>Most people can be trusted</i>	<i>(Don't know)</i>
00	01	02	03	04	05	06	07	08	09	10	88	

A9 CARD 4: Using this card, do you think that most people would try to take advantage of you if they got the chance, or would they try to be fair?

<i>Most people would try to take advantage of me</i>											<i>Most people would try to be fair</i>	<i>(Don't know)</i>
00	01	02	03	04	05	06	07	08	09	10	88	

A10 CARD 5: Would you say that most of the time people try to be helpful or that they are mostly looking out for themselves? Please use this card.

<i>People mostly look out for themselves</i>											<i>People mostly try to be helpful</i>	<i>(Don't know)</i>
00	01	02	03	04	05	06	07	08	09	10	88	

First Alternative form

HS10 Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people? Please tick the box that is closest to your opinion, where 0 means you can't be too careful and 5 means that most people can be trusted.

You can't be
too careful

Most people
can be trusted

0

1

2

3

4

5

HS11 Do you think that most people would try to take advantage of you if they got the chance, or would they try to be fair? **Please tick one box.**

Most people
would try to take
advantage of me

Most people
would try
to be fair

0

1

2

3

4

5

HS12 Would you say that most of the time people try to be helpful or that they are mostly looking out for themselves? **Please tick one box.**

People mostly
look out for
themselves

People mostly
try to be
helpful

0

1

2

3

4

5

Second Alternative form

HS28 Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people? **Please tick one box.**

You can't be too careful 1

Most people can be trusted 2

HS29 Do you think that most people would try to take advantage of you if they got the chance, or would they try to be fair? Please tick one box.

Most people would try to take advantage of me 1

Most people would try to be fair 2

HS30 Would you say that most of the time people try to be helpful or that they are mostly looking out for themselves? **Please tick one box.**

People mostly look out for themselves 1

People mostly try to be helpful 2

3.4. Political Trust

Form of the main questionnaire

CARD 11: Using this card, please tell me on a score of 0-10 how much you personally trust each of the institutions I read out. 0 means you do not trust an institution at all, and 10 means you have complete trust. Firstly...**READ OUT**

	No trust At all												Complete (Don't trust know)
B7 ... [country]'s parliament?	00	01	02	03	04	05	06	07	08	09	10	88	
B8 ... the legal system?	00	01	02	03	04	05	06	07	08	09	10	88	
B9 ... the police?	00	01	02	03	04	05	06	07	08	09	10	88	
B10 ... politicians?	00	01	02	03	04	05	06	07	08	09	10	88	
B11 ... the European Parliament?	00	01	02	03	04	05	06	07	08	09	10	88	
... the United Nations?	00	01	02	03	04	05	06	07	08	09	10	88	

First alternative form

Please indicate on a score of 0 to 10 how much you personally trust each of the institutions below. 0 means you do not trust an institution at all, and 10 means you have complete trust

Please tick the box that is closest to your opinion.

		No trust					Complete trust
		at all					
HS13	[Country]'s parliament	0 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
HS14	The legal system	0 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
HS15	The police	0 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>

Second Alternative form

HS31 Please say on a scale of 0 to 10 how much you trust **[country]'s parliament**. If you have no trust at all give a score of 0. If you have complete trust, give a score of 10. The more you trust the parliament, the higher the score should be.

Your score:

HS32 Please say on a scale of 0 to 10 how much you trust the **legal system**. If you have no trust at all give a score of 0. If you have complete trust, give a score of 10. The more you trust the legal system, the higher the score should be.

Your score:

HS33 Please say on a scale of 0 to 10 how much you trust the **police**. If you have no trust at all give a score of 0. If you have complete trust, give a score of 10. The more you trust the legal system, the higher the score should be.

Your score:

3.5. Trait specific scales versus agree-disagree items

It has been shown for several items that trait specific scales are much better than agree disagree scales. A very convincing case was received for the two different political efficacy scales. Unfortunately the items that remained for political efficacy are not the best possible (Saris 2005). I suggest to repair the instrument in this study but this requires for once an increase of the core questionnaire with 2 question (see below: B1 and B3).

Form for the main questionnaire

B1 CARD 7 Do you think that you are able to participate in political activities?
Please use this card.

Definitely not	1
Probably not	2
Not sure either way	3
Probably	4
Definitely	5
(Don't know)	8

B2 CARD 6 How often does politics seem so complicated that you can't really understand what is going on?
Please use this card.

Never	1
Seldom	2
Occasionally	3
Regularly	4
Frequently	5
(Don't know)	8

B3 CARD 7 Do you think that you could take an active role in a group involved with political issues?
Please use this card.

Definitely not	1
Probably not	2
Not sure either way	3
Probably	4
Definitely	5
(Don't know)	8

B4 CARD 8 How difficult or easy do you find it to make your mind up about political issues? Please use this card.

- Very difficult 1
- Difficult 2
- Neither difficult nor easy 3
- Easy 4
- Very easy 5
- (Don't know) 8

First Alternative form

Please indicate to what extent you agree or disagree with each of the following statements.

HS4 "I am able to participate in political activities."

Please tick one box.

- Disagree strongly 1
- Disagree 2
- Neither disagree nor agree 3
- Agree 4
- Agree strongly 5

HS5 "Sometimes politics seems so complicated that I can't really understand what is going on?."

Please tick one box.

- Disagree strongly 1
- Disagree 2
- Neither disagree nor agree 3
- Agree 4
- Agree strongly 5

HS6 "I think I could take an active role in a group involved with political issues ."

Please tick one box.

Disagree strongly 1

Disagree 2

Neither disagree nor agree 3

Agree 4

Agree strongly 5

HS7 "I find it easy to make my mind up about political issues ."

Please tick one box.

Disagree strongly 1

Disagree 2

Neither disagree nor agree 3

Agree 4

Agree strongly 5

Second alternative form

Please indicate to what extent you agree or disagree with each of the following statements.

HS21 "I am able to participate in political activities."

Please tick one box.

Agree strongly 1

Agree 2

Neither agree nor disagree 3

Disagree 4

Disagree strongly 5

HS22 “Sometimes politics seems so complicated that I can’t really understand what is going on”

Please tick one box.

Agree strongly 1

Agree 2

Neither agree nor disagree 3

Disagree 4

Disagree strongly 5

HS23 “I think I could take an active role in a group involved with political issues”

Please tick one box.

Agree strongly 1

Agree 2

Neither agree nor disagree 3

Disagree 4

Disagree strongly 5

HS24 “I find it easy to make my mind up about political issues”

Please tick one box.

Agree strongly 1

Agree 2

Neither agree nor disagree 3

Disagree 4

Disagree strongly 5

3.6. Test for social desirability

Because the left right scale commonly used may be affected by a social desirability error and this can easily be determined in a mode experiment we suggest to test these effects for the left right scale using two extra questions in the main questionnaire.

The form for the main questionnaire

B28 CARD 12: In politics people sometimes talk of “left” and “right”.
Using this card, where would you place yourself on this scale,
where 0 means the left and 10 means the right?

Left												Right	(Don't know)
00	01	02	03	04	05	06	07	08	09	10	88		

B29 CARD 12: Using the same card, where would you place your most preferred party on this scale, where 0 means the left and 10 means the right?

Left												Right	(Don't know)
00	01	02	03	04	05	06	07	08	09	10	88		

B30 CARD 12: Using the same card, where would you place the party which you most dislike on this scale, where 0 means the left and 10 means the right?

Left												Right	(Don't know)
00	01	02	03	04	05	06	07	08	09	10	88		

The first alternative form

B28 CARD 13: In politics people sometimes talk of “left” and “right”.
Using this card, where would you place yourself on this scale,

Extreme left	1
Rather Left	2
Left	3
Some what left	4
central left	5
Neither left nor right	6
Central right	7
Some what right	8
Right	9
Rather right	10
Extreme right	11

B29 CARD 12: Using the same card, where would you place your most preferred party on this scale ?

Extreme left	1
Rather Left	2
Left	3
Some what left	4
central left	5
Neither left nor right	6
Central right	7
Some what right	8
Right	9
Rather right	10
Extreme right	11

B30 CARD 12: Using the same card, where would you place the party which you most dislike on this scale ?

Extreme left	1
Rather Left	2
Left	3
Some what left	4
central left	5
Neither left nor right	6
Central right	7
Some what right	8
Right	9
Rather right	10
Extreme right	11

The second alternative form

B28 CARD 12: In politics people sometimes talk of “left” and “right”. Using this card, where would you place yourself on this scale, where 0 means “Extreme left” and 10 means “Extreme right” ?

Extreme Left										Extreme Right		(Don't know)
00	01	02	03	04	05	06	07	08	09	10	88	

B29 CARD 12: Using the same card, where would you place your most preferred party on this scale, where 0 means “Extreme left” and 10 means “Extreme right” ?

Extreme Left										Extreme Right		(Don't know)
00	01	02	03	04	05	06	07	08	09	10	88	

B30 CARD 12: Using the same card, where would you place the party which you most dislike on this scale where 0 means “Extreme left” and 10 means “Extreme right” ?

Extreme Left										Extreme Right		(Don't know)
00	01	02	03	04	05	06	07	08	09	10	88	

4. Summary

This is a first proposal for the evaluation of the quality of the different instruments in the different modes. In each mode the main questionnaire contains the form used in the ESS in the different rounds. The supplementary questionnaire in each mode should contain the set of first alternative forms. This means that the supplementary questionnaire contains 19 questions. They should be collected in the same way as the main questionnaire. If the sample sizes in the different modes are sufficiently large (>200) then the MTMM model for these topics can be estimated. This has been tested.

The result of this analysis will be that I can say for all of these topics if the quality of the questions in the different modes is the same or not. Besides one can say whether the concepts in the different modes have the same interpretation. This will be done by testing for cognitive equivalence (Saris and Gallhofer 2007). Furthermore it is possible to determine whether the effects of the different choices are equally large in the different modes.

In the last section of this proposal I have also specified a second alternative form because I think that it makes sense to do the same experiments in the 4th round of the ESS. There the normal split ballot MTMM design will be used with three groups so that all people get only 12 extra questions in the supplementary questionnaire. This design with 3 different forms is more

robust with respect to estimation than the design in the mode experiment with only 2 forms. However, if the estimation in the mode experiment gives problems, the experiment of the main study can help to empirically identify the estimates of all parameters.

Besides this technical advantage, another advantage is that we have not for all new countries the quality of the different questions. To repeat these experiments will make our data more complete and help us to generate composite scores for the different concepts for all countries as has been done for some countries so far in papers of Saris (2005), Coromina and Saris (2007) and Knoppen and Saris (2008).

Two topics have not been studied before. Political efficacy has been evaluated before but this evaluation has lead to the conclusion that the wrong items have been chosen for the core questionnaire (Saris 2005). For the left right scale no analysis has been done so far. So the suggested test is a new experiment.

Another argument for the repetition of some of these experiments is that we can see if the quality estimates and the cognitive equivalence of the concepts are stable through time. Without this repetition this is not possible.