Chapter 16 Using mixed modes in survey research: Evidence from six experiments in the ESS

Ana Villar and Rory Fitzgerald

Introduction

Technological advances and increasing access to telephone and the internet mean that it is now possible for people to complete surveys by phone or online rather than face-to-face with an interviewer present. As budgets tighten and the benefits of using other modes (such as speed and flexibility) become more apparent, there has been a debate about whether it is better to use alternatives to face-to-face surveys either exclusively or in combination with in-person interviews. This chapter summarises findings from research conducted under the auspices of the European Social Survey (ESS) over the past 12 years to evaluate the potential effects of mixed-mode data collection on its own cross-national survey estimates.

The ESS is an academically driven cross-national survey that has been conducted every two years since 2001 across large parts of Europe. One of the main aims of the ESS is 'to chart stability and change in social structure, conditions and attitudes in Europe' by analysing, combining, and comparing data from different countries. In order to be able to ensure meaningful comparisons of survey estimates across countries, it was decided that essential survey conditions, such as mode of data collection, should be as similar as possible across participating countries (Jowell, 1998), and that the high standards of scientific rigor would be implemented in all of them (Jowell, Roberts, Fitzgerald, & Eva, 2007).

Cross-national surveys like the ESS face challenges not usually encountered in national studies. To ensure data comparability across different countries, these surveys depend on a 'principle of equivalence' (Jowell, 1998). This principle relies on standardisation of all possible aspects of the survey process: sampling (Lynn, Japec, & Lyberg, 2006), question wording (Harkness, 2003; Harkness, Villar, & Edwards, 2010), response options and coding schema (Jowell et al., 2007), and so on. For this reason, the most ambitious multinational projects tend to require all participating countries to employ the same mode of data collection. In the case of the ESS, the exclusive mode for data collection in the main questionnaire is face-to-face interviewing.

Face-to-face interviewing has been recognised by many as the 'gold standard' among data collection methodologies. Research has shown face-to-face surveys obtain higher response rates (de Leeuw, 1992; Groves et al., 2004; Holbrook, Green, & Krosnick, 2003), largely because of the effectiveness of in-person contacts at persuading sampled individuals to take part. Face-to-face data collection can also improve data quality, partly because interviewers can help navigate the questionnaire, probe answers to ensure they are recorded correctly, and motivate respondents when they feel fatigued. In addition, coverage and familiarity with different modes of data collection varies widely across European countries, whether because of survey traditions or because of differential access to technologies (such as internet or fixed-line telephone coverage). In-person interviewing reduces these gaps and facilitates consistency of survey implementation across countries, thus making it the best-possible single-mode data collection option for a rigorous cross-national survey, especially with the ESS current hour-long questionnaire. For additional details about the goals and initial setup of the ESS and the survey climate at the time, see Jäckle, Roberts, and Lynn (2006).

However, at the time the ESS was established, there were concerns that face-to-face data collection would not be optimal in all countries. Researchers in a few countries worried that a) an interviewer infrastructure did not exist to carry out large-scale face-to-face surveys, b) costs would be higher than feasible, and/or c) participation would be lower or more biased than using a different mode of data collection, given the survey tradition in the country. These concerns were valid and important, but it was not clear from the evidence available how using different modes in different countries would affect survey estimates and data quality, especially in a cross-national context. Therefore, the Core Scientific Team (CST) of the ESS established a programme of methodological research to examine the impact of mixing data collection modes (within and between countries) on the overall quality of survey estimates. At the early stages of the research programme, the most likely alternative mode for the ESS was telephone interviewing, given that the length and complexity of the interview questioned the suitability of using postal self-completion surveys, and that large proportions of the population did not have internet access. Thus, the first studies focused on evaluating the suitability of telephone interviews in the ESS. As internet penetration increased across Europe over the life of the ESS mixed-mode research programme, however, attention to the web as an alternative mode has increased, and all studies since 2008 have included web data collection. This chapter summarises findings from the six studies carried out since 2003 under the ESS mixed-mode research programme to evaluate the effects of mixed-mode data collection on survey estimates.

The six studies

Six ESS studies on mixed-mode data collection have been conducted since 2003 (see Table 16. for an overview). Of these, three focused solely on assessing the effect of mode on measurement (studies 1, 2 and 3). The other three investigated the feasibility and practical challenges of implementing different mode designs in cross-national surveys, where all sources of survey error were potentially affected by the mode design (studies 4, 5, and 6[1]). In this section, we present an overview of the main essential survey features for each study.

Study 1: Measurement differences across four modes. The first study focused on studying measurement differences in estimates by mode, comparing face-to-face interviews, telephone interviews, self-completion paper and pencil questionnaires and web-based self-completion questionnaires. The study took place in May and June 2003 in Hungary, where a convenience sample of participants were approached on the street and invited to participate in a 'hall test'. Those agreeing to take part were randomly assigned to one of the four modes of data collection. Participants were re-interviewed in a different randomly-assigned mode[2], either 20 minutes later at the same venue or a few days later at home. Cooperation rates to the second interview for those completing at home were 73% for telephone and 48% for mailed questionnaires. All participants received the same 25 questions from the ESS and the Eurobarometer in both interviews, making it possible to examine differences in responses both between and within the same participants.

Study 2: Measurement differences: Finding the cause. The second study aimed to investigate the causes of measurement differences between telephone and face-to-face data collection observed in Study 1. The experiment was designed to isolate the effect of mode from the effect of question changes related to adaptation of the questionnaire for a different interview setting (i.e., a telephone interview). Three experimental groups were compared: 1) a group interviewed face-to-face using 44 ESS standard questions including showcards, 2) a group interviewed by telephone (fixed or mobile)

---

[1] It is worth noting that all studies intended to evaluate feasibility and compare alternative modes to the standard ESS face-to-face data collection (studies 4, 5, and 6) used fieldwork procedures as close as possible to those applied in the country for the corresponding ESS round.

[2] The experimental design was not fully crossed: those interviewed by web in the first wave of data collection were re-interviewed in the same mode.

using the same 44 questions, adapted for telephone implementation and without showcards, and 3) a group interviewed face-to-face using the 44 telephone-adapted questions without showcards.

Table 16.1. Overview of design characteristics for each study

	Study 1: Measurement differences across four modes	Study 2: Measurement differences: Finding the cause	Study 3: Measurement differences: F2F vs web
Main goal	Mode measurement differences	Mode measurement differences	Mode measurement differences
Modes	F2F vs Telephone vs Paper SAQ vs Web	F2F vs Telephone	F2F vs Web
Experimental manipulations	Within-respondent re-interview design (different modes)	Compared questions adapted for telephone (F2F & phone) to standard ESS questions (F2F)	Within-respondent re-interview design (different modes)
Sample design	Convenience, hall test	Probabilistic, limited to countries' capitals	Subsample of ESS respondents
Year	2003	2005	2010
Countries	Hungary	Hungary and Portugal	United Kingdom

	Study 4: Feasibility of ESS telephone interviews	Study 5: Feasibility of mixed-mode designs: sequential vs concurrent	Study 6: Feasibility of mixed-mode designs across countries
Main goal	Feasibility of alternative mode: telephone	Feasibility of mixed-mode designs	Feasibility of mixed-mode designs
Modes	F2F (if fixed-line at home) vs Telephone	F2F vs Web+Telephone+F2F	F2F vs Web+F2F vs Telephone +F2F
Experimental manipulations	Duration of the survey: 60 minutes vs. 45 minutes vs 30+30 minutes	Sequential vs Concurrent mixed-mode designs	Sweden: F2F+ Telephone vs Telephone +F2F UK: incentives, respondent selection method
Sample design	Probabilistic, sometimes different from ESS	Probabilistic, same as standard ESS	Probabilistic, same as standard ESS
Year	2006	2008	2012
Countries	Germany, Poland, Hungary, Cyprus, Switzerland	Netherlands	Estonia, Sweden, United Kingdom

Gallup Europe carried out the experiment using random samples from Hungary and Portugal starting in July 2005. Fieldwork was concentrated in Budapest and Lisbon, where suitable sampling frames were available that included telephone numbers *and* addresses, thereby holding error from sampling and coverage constant between modes. Respondents answered questions about their interview experience, and interviewers answered six questions to describe the respondents' engagement and effort during the interview process. The response rate was 32% in both modes.

Study 3: Measurement differences: F2F vs web. This study intended to evaluate measurement differences between face-to-face and web data collection, using a within respondent re-interview design. The goal was to evaluate how open-ended questions, complex questions and seemingly mode-sensitive questions compared across the two modes. The study was carried out in two waves: the first wave was the main stage of Round 5 of the ESS in the United Kingdom (September-November 2010), and the second was a follow-up web survey of a subsample of the same face-to-face respondents (October 2010-January 2011). The sampling design for the face-to-face stage was a clustered, stratified, 3-stage random probability design using a frame of addresses. At the end of the interview, UK ESS respondents were invited to participate later on in a web survey; the survey included a subset of 32 of the questions they had answered during the face-to-face interview. Of all respondents invited to the follow-up study[3], 43% said they were willing to be contacted. Invitations were sent about three weeks after the face-to-face interview to 927 respondents, of whom 613 completed the web survey, for a completion rate of 66% and a cumulative response rate (Callegaro & DiSogra, 2008) of 29%. Comparisons between responses given to the face-to-face and web by the same respondents can provide insights about measurement differences across the two modes. There was, however, no face-to-face control group due to limited funding availability. In addition to the implications for mode differences in measurement, this study is relevant to the current ESS pilot to establish a cross-national probability based panel as part of the SERISS project (see www.seriss.eu).

Study 4: Feasibility of ESS telephone interviews. In 2006, a study was designed to examine a) the practical challenges and feasibility of using telephone as a mode of data collection in the ESS and b) the effect of inviting respondents to participate in telephone interviews of different lengths on their willingness to participate in the survey. It was the first study to implement the full ESS interview in an alternative mode. Sample members, selected using probability sampling methods, were randomly assigned to one of three treatment groups, which varied according to the length and design of the questionnaire: A) 60 minutes; B) 45 minutes; and C) 30 minutes (and a further 30-minute interview, offered at the end of the first 30 minutes). The target population in the telephone samples were households with fixed-line telephones; for this reason, all analyses comparing the telephone and face-to-face samples in Study 4 excluded respondents from the standard ESS who did not have a fixed-line at home. Response rates varied widely across different countries and interview lengths (see section 4.1 on outcome rates). The experiment was carried out in five countries: Cyprus, Germany, Hungary, Poland and Switzerland. The selection of countries was based on pragmatic considerations and the available budget; but within those constraints, countries with divergent traditions of survey practice and facing different challenges in their data collection efforts were chosen. Countries were encouraged to use fieldwork procedures as similar as possible to the standard ESS in their countries, but some features varied due to budget constraints (e.g., incentives, pre-notification letters).

Study 5: Feasibility of mixed-mode designs: sequential vs concurrent. Study 5 also intended to investigate the feasibility of alternative mode designs, including for the first time web mode as a way to increase cost-efficiency. The study was conducted in The Netherlands in 2008, parallel to Round 4 of the standard ESS, and was the first ESS study to use within-country, across respondents mixed-mode data collection. A random sample of 2,500 addresses was selected, using the same postal sample frame and probability-based design as in the standard Dutch ESS. The sample of addresses

---

[3] Invitations were issued during the first two months of fieldwork; 2,144 ESS respondents were invited.

was matched with information about the name and telephone number of people living at the address for 70% of the sampled units using a commercial database. The sample for which a telephone number was available was then randomly allocated to one of the two mixed-mode designs: a 'concurrent' design, where respondents were contacted by telephone and offered three modes to choose from, and a 'sequential' design, where respondents first received a telephone request to complete the survey using the web (without mention of other modes if they had internet access). Those who had not completed the survey after a certain period were offered a telephone interview or, finally, a face-to-face interview. For addresses for which telephone contact was not available, the design was different: an interviewer visited the household and offered an in-person interview; those who refused were invited during the same visit to complete a web survey.

Study 6: Feasibility of mixed-mode designs across countries. In this study, countries were invited to select the ideal mixed-mode design given the country's survey environment. The Estonian experiment compared the standard ESS survey to a sample of 1,050 units selected from the same population register from which the main ESS sample was drawn. An invitation was mailed to each sample unit (named individual) asking them to participate in the ESS by completing a web survey. An interviewer visited those who did not complete the web survey after two reminders, to try to complete a face-to-face interview. Similarly, the UK tested a mixed-mode data collection design where invitations to a web survey were mailed to a random sample of UK addresses; if no response was obtained from the address after one reminder, an interviewer visited the selected address and tried to complete a face-to-face interview[4]. However, the UK lacks a sample frame of individuals, so additional procedures were implemented to try and randomly select a household member by enclosing instructions for a household member to carry out the selection. The UK experiment compared the main ESS survey of about 4,500 respondents to a sample of 2,000 households selected from the same address sample frame from which the main ESS sample was drawn.

The Swedish experiments compared the main ESS survey to two samples of 300 respondents selected from the same population register from which the main ESS sample was drawn. The proposal focused on the comparison of two mixed-mode designs: a) a *sequential, response enhancing design* where individuals who refused to make an appointment for an in-person interview would be re-contacted and offered the possibility to complete it over the telephone and b) a *sequential, cost-effective design* where sampled individuals were first offered to respond to a telephone survey, and those who refused were re-contacted by telephone and invited to make an appointment for an in-person interview.

Sampling and coverage in the ESS mixed-mode studies

When survey programmes consider new designs of data collection, this may have consequences for their sampling procedures. In this section, we consider the implications of different mode designs for sampling frames, sampling designs, and fieldwork procedures. The measurement studies in general did not involve large-scale probability-based sampling; therefore, this section focuses on the feasibility studies (studies 4 through 6).

Survey sampling involves defining a target population and deciding on an approach (the sampling design) to randomly select the final units of research from a list of units (the sample frame). As with the standard ESS, sample designs were allowed to vary cross-nationally (see Table 16. for an overview), depending on the availability and quality of sampling frames in each country (Lynn, Haeder, Gabler, & Laaksonen, 2007), but this 'flexibility assumes probability selection methods: known probabilities of selection for all population elements' (Kish, 1994, p. 173). In section 3.1, we discuss the sampling frames used in the three feasibility studies, and we evaluate challenges faced and

---

[4] To reduce costs, only one-fourth of those who did not complete the web interview were contacted for a face-to-face interview.

solutions applied in the context of conducting these studies. In section 3.2, two self-administered approaches to within-household random selection of respondents are compared, addressing one of the biggest obstacles for surveying the general population when sampling frames of individuals are not available and an interviewer is not present.

Sampling frames and sampling designs

The use of different modes for data collection can affect the choice of sample frame and sample design. For simplicity, one could choose sample frames with contact information best suited for the mode of data collection: addresses for in-person and mail surveys, telephone numbers for telephone surveys, and email addresses for online surveys. However, such sampling frames are not always available or do not have adequate quality, especially when conducting surveys among the general population. Telephone lists often suffer from under-coverage problems—excluding certain parts of the population—as well as over-coverage problems when members of the population can be linked to more than one telephone number. Email lists are hardly available. Therefore, researchers often face the decision of choosing between a high quality frame that does not contain contact information for the desired mode of data collection and a lower quality frame with the contact details.

Choosing the sampling frame with the highest quality might mean changing mode from recruitment stages to the actual data collection. For example, in Study 6 the ESS used sampling frames containing addresses to recruit respondents for an online survey in Estonia and the UK. Given that these sampling frames did not include email addresses or other online contact information for population units, the invitation mode had to be different from the mode of data collection. In both countries, it was decided that a mailed invitation was the most cost-effective approach.

Table 16.2. Overview of design characteristics for each study: sampling

	Year	Location	Sample design	Sampling frame	Realised sample size
Study 1	2003	Hungary	Convenience sample	Street recruitment	1,987 respondents. Interviews: 402 web 1,109 F2F, 1,189 Telephone
Study 2	2005	Budapest/Lisbon	Simple random sample	Database with telephones and addresses	515 F2F with showcards 518 F2F without showcards
Study 3	2010	United Kingdom	All UK ESS Round 5 respondents who were interviewed before 11 November 2010 were invited	ESS respondents	2,422 ESS 613 Web
Study 4*	2006	Cyprus, Germany, Hungary, Poland, Switzerland	Random sample (varied)	RDD, population register + matched telephone data	Cyprus: 918 ESS, 83 Telephone Germany: 2,663 ESS; 369 Telephone Hungary: 855 ESS; 252 Telephone Poland: 1,218 ESS; 339 Telephone
Study 5	2008	The Netherlands	Systematic sampling of households, last birthday method for respondents	Population register + matched telephone data	1,778 ESS 198 Web 123 Telephone
Study 6	2012	Sweden, Estonia, United Kingdom	For each country, same as in ESS Round 6	For each country, same as in ESS Round 6	Estonia: 2,380 ESS; 356 Web, 230 F2F UK: 2,286 ESS; 389 Web, 64 F2F

*ESS sample sizes in study 4 represent respondents with fixed telephone lines at home

Study 4 - Feasibility of ESS telephone interviews. Participating fieldwork agencies were instructed to use the best possible probability sample design available containing telephone contact information that would support telephone interviewing (Roberts, Eva, Lynn, & Johnson, 2009). The sampling frames that countries were using for the standard ESS did not include telephone contact information, so alternative sampling frames were used for the telephone surveys. The main sampling challenge posed by using telephone as the survey mode stemmed from the fact that available sampling frames included only households with fixed-line phones, which resulted in exclusion from the sample of households with mobile telephones only and households without telephone access. Therefore, the frames led to undercoverage of the intended ESS population (residential adults aged 15+). The proportion of households without a fixed-line varies across Europe (surpassing 5% of the general population in most countries, and rising in many), which would lead to different coverage error levels across countries, potentially jeopardising comparability and introducing bias.

Study 5 – Feasibility of mixed-mode designs: sequential vs concurrent. In Study 5, the mode design involved telephone contact to invite respondents to participate using different modes of data collection (Eva et al., 2010). The goal was to reduce fieldwork costs by not having interviewers travel to the households. Ideally, target respondents would be selected from a sampling frame containing telephone contact information. However, before interviewers called households, a pre-notification letter was to be sent, as research has shown that this is an effective way to increase response rates (De Leeuw, Callegaro, Joop, Korendijk, & Lensvelt-Mulders, 2007). Therefore, address contact information was also necessary. The only possible sampling frame with telephone and address data available were telephone directories, but using a telephone directory would have excluded members of the population who did not have fixed-line telephones and those with "unlisted" numbers. These groups comprised about 12% of the population in 1987, but the number has considerably increased since then, which may explain why only 70% of the addresses selected for Study 5 could be matched to a telephone number. Furthermore, unlisted households differ significantly from those found in telephone directories (Beukenhorst, 2012). Therefore, a sampling frame of addresses was chosen for Study 4. The research team selected 2,500 address from a list of postal addresses (Martin & Lynn, 2011). For the addresses where no telephone could be matched, contact with the household took place in person. This increased both complexity of fieldwork management and the costs, reducing the overall benefits of the design. The last-stage of the sampling procedure where the target respondent was identified matched the standard ESS sampling procedure in the Netherlands, so no additional challenges were met in relation to this procedure. However, this study shows the complex relationship between different error sources, which multiplies exponentially when adding modes of data collection and countries.

Study 6 – Feasibility of mixed-mode designs across countries. Sampling in the ESS mixed-mode designs followed the same design procedures that were agreed for the main stage in Round 6 (Villar, Winstone, Prestage, & Lynn, 2014). In Estonia, there were no differences across mode designs in how the sampling design was implemented: a random sample of individuals was drawn from the same population register used for the main ESS, with explicit stratification by region and implicit stratification by age and gender. Instead of an interviewer visiting the address available from the sampling frame, a letter was addressed to the named respondent with an invitation to take part in a web survey. The challenge was that a considerable proportion of the population had a different address from the one in the register. Addresses are easier to update when an interviewer visits the address than by just sending a letter, which suggests that a web survey alone could lead to higher nonresponse rates than a face-to-face survey amongst a rather specific group of the population due to noncontact. This raises the question of whether those for whom address information is incorrect differ systematically from those with correct information, creating a potential for nonresponse bias. In Sweden, the sampling frame, sampling design and first contact attempts were conducted identically in all mode designs, given that telephone contact is used in the standard Swedish ESS to make appointments for face-to-face interviews. The design was a simple random sample without

replacement using a population register as sampling frame, which includes all individuals living in Sweden together with their contact information, including address and telephone number. Finally, in the UK, the sample design for the selection of addresses was identical across all mode designs: a stratified, clustered, random sample, drawn using the Postcode Address File (PAF). However, selection of the final respondent in households with more than one individual aged 15 or older differed. In the standard ESS, selection of the final respondent was done by the interviewer using a Kish grid (Kish, 1949), whereas in the mixed-mode sample this step was left to the target household themselves. Section 3.2 discusses in detail the implications of these approaches.

Within-household random selection of individuals

Some ESS countries have access to population registers that provide information about individuals and contact information, whereas others have to resort to using samples of addresses or area probability sampling. When sampling frames that contain addresses are used, contact with the final respondent cannot take place until the selection of a person within the household has been carried out; sometimes, the sampled address contains more than one household and two selection procedures are necessary. Face-to-face data collection is especially useful to obtain a random sample in countries using sampling frames of addresses or areas, because one can train interviewers to implement these random selection methods. When interviewers are not involved, selection of the individual has to rely on the person in the household that receives the instructions. Research suggests that compliance from these individuals is less than perfect, because people are sometimes unable and other times unwilling to follow these instructions (Olson & Smyth, 2014; Olson, Stange, & Smyth, 2014). Therefore, in many countries, a change in mode of data collection can affect whether the final sample is actually a random sample of the population.

In Study 6 (*Feasibility of mixed-mode designs across countries)*, recruitment of UK respondents started by sending an initial invitation to 'the resident'. Two approaches were considered to complete the sampling frame and carry out within-household selection of the respondent: a) an interviewer-based within household selection procedure. Sampled addresses would be sent a letter with a description of the study and a free toll number to call if they were interested in participating in the online study. During the call, the interviewer would ask the household informant to describe the household composition, apply random selection procedures to identify the final sample unit, and ask the selected person to complete the web survey. b) A respondent-based selection procedure. Sampled addresses would be sent a letter with a description of the study, an explanation of why random selection is important and how to apply the procedure. In the end, the respondent-based approach was chosen, because we expected that very few household informants would actually call the toll free number to go through the selection procedure. Two respondent-based selection procedures were tested: a) the last/next birthday method, where the eligible household member with the last (or next) birthday is invited to participate; this procedure has been described as 'quasi-random' (Gaziano, 2005), and b) the household roster method consisted on a few steps: first, a household informant followed the survey link provided in the letter; second, the informant was asked to list all household members 15 or older; third, a built-in algorithm carried out the random selection of one of the household members and this was communicated to the informant; last, if the selected member was someone other than the person completing the roster, the informant was asked to pass on the letter to that member. The household roster method is more intrusive than the birthday method, but it is truly random if instructions are followed correctly.

After fieldwork was completed, we evaluated both approaches. The goals were to find out whether household members who received the invitation letter complied with the instructions, and to test the impact of the two different methods on response rates. To assess compliance, we asked respondents to provide the month of birth of all household members, and we asked those who completed a household roster to provide their name or alias at the end of the survey so that the promised incentive

could be sent to them. Compliance with the birthday method was evaluated by checking whether the respondent had the birthday closest to the month of interview in the household, whereas compliance with the household roster method was evaluated by comparing the name provided at the end of the survey to the name or initials of the respondent selected from the household roster. Compliance tests proved to be challenging. In many instances, compliance could not be assessed because the necessary information had not been provided (e.g., month of birth for one of the household members, or the initials at the end of the survey) or because it could not be determined with certainty (e.g., two household members were born in the sample month, or two household members had the same initials). The birthday method led to noncompliance levels of 23% of the total sample. About a third of households only had one adult (therefore, no selection was necessary), in 17% cases it was not possible to determine compliance, and 27% cases were identified as compliant. Unfortunately, we could not appropriately evaluate compliance of the household roster method. Of those households with more than one adult, more than half did not provide either name or initials. Of the remaining respondents (n = 93), two thirds provided names that matched the selected respondent and one third provided a different name, but that may just signify that the name in the first instance was just a pseudonym or fake name.

We found no evidence of an association between the within-household selection method and response rates (39% with the birthday method, 37% with random selection; $\chi^2 (2) = 3.37$; $p = 0.19$).

Summary

Sampling in the existing cross-national face-to-face survey is already rather complex, with a mixture of samples of individuals, samples of households or addresses, and area probability sampling (including random routes procedures) all being used. However, all these sampling frames are followed by visits to the address where, if necessary, an interviewer can complete the sampling process by applying procedures to randomly select a respondent from among all household members. Based on these ESS experimental studies, it seems that cross-national surveys switching to mixed-mode designs may need to either use new frames (and accept lower coverage in most cases), or use the existing frames and allow households to do their own selection. This could not only jeopardise quality or lead to reduced cost savings, but it may introduce even greater differences in sampling design across countries, potentially reducing comparability of the collected data.

Survey participation: response rates, nonresponse bias, and sample composition in the ESS mid-mode studies

One of the expected advantages of adopting a mixed-mode data collection design rests on improving survey participation, either because more potential respondents end up taking part or because the final make-up of the realised sample of respondents is more similar to the target population than the respondents to the alternative single-mode design would have been. Different mechanisms for these differences in sample composition can be hypothesised (De Leeuw, 2005; Martin, 2011):

a) Sample units and target respondents might be easier to contact via some modes. In many European countries, letters and telephone calls are generally easier (or more tempting) to disregard nowadays than a personal visit to an address, due to the volume of contact attempts received through those channels of communication. Perhaps the same could be said of e-mail and pop on-liner survey requests. At the same time, in certain populations a personal visit may not be appropriate or well received.

b) Sample units may also be more likely to agree to participate depending on the mode of data collection. Some may prefer to complete the survey at their own time and pace or without an interviewer present and thus be more likely to agree to complete a self-administered questionnaire but others may prefer to have the personal interaction with the

interviewer. Many may find it easier to refuse (ignore) a mailed or telephone request than an in-person request by an interviewer.

c) Some modes may be more accessible to certain populations than others. For instance, self-administered modes may not be appropriate for populations with low literacy levels (De Leeuw, 2005), and on-line surveys might not be accessible to the elderly or those on low incomes modes relying on telecommunications may not reach the entire target population in war zones.

Therefore, depending on the target population and the single mode that we are comparing the mixed-mode design to, survey participation may increase or improve when allowing respondents to complete the survey in an alternative mode. Given baseline levels of response rates, web and telephone surveys have the greatest potential for improved survey participation, and mail and face-to-face surveys have the least potential because they generally achieve high rates for that mode of completion, following well-established response maximisation strategies. In fact, there is no empirical evidence suggesting that mixed-mode designs can reduce nonresponse error with respect to face-to-face surveys (Dillman, Smyth, & Christian, 2009; Lynn, Uhrig, & Burton, 2010; Martin, 2011). Surveys with high response rates around the world tend to be conducted face-to-face, like the Labour Force Survey and the European Quality of Life Survey in many European countries, the National Health Interview Survey in the United States, and the British Crime Survey in the UK. Mandatory mixed-mode surveys can achieve response rates as high as face-to-face surveys (e.g., the American Community Survey in the United States or the Labour Force Survey in some European countries), but these are exceptional cases because, unlike most surveys, they are not voluntary. Therefore, we have no grounds to expect more people will agree to participate in the ESS if mixed-mode data collection is allowed. However, if response rates were not damaged substantially, measurement differences between modes were minimal, nonresponse bias did not increase and costs were reduced, mixed-mode data collection might be preferable on the grounds of cost-effectiveness. Therefore, it was still worth testing whether mixed-mode designs could be effectively implemented in the ESS to improve cost-efficiency. In addition, response rates to face-to-face surveys might see substantial droops in the future and this become prohibitively expensive in certain countries. This is what has made assessing the possible implications of introducing mixed-mode data collection so critically important. Four of the studies in the ESS mixed-mode programme of research (studies 2, 3, 4, and 6) evaluated the impact of mixed-modes on survey participation; we summarise their findings in this section.

Outcome rates

Survey participation can be measured using different outcome rates. Final response rates are a general way to assess and compare survey participation across studies; in addition, analysing different nonresponse components (e.g., refusals, noncontacts) provides insight into the mechanisms that explain the differences in survey participation across modes. We start this section discussing overall response rates and then focus on noncontact and refusal rates, comparing outcomes from the face-to-face standard main stage ESS to each alternative mode design. All outcome rates below assume that cases where eligibility could not be determined during fieldwork would have been eligible for participation[5] in the survey (yielding rates similar to the American Association for Public Opinion Research Response Rate 1).

Response rates

---

[5] Eligible sample units in the ESS are 'all persons aged 15 and over (no upper age limit) resident within private households in each country, regardless of their nationality, citizenship or language' (for further details on how eligibility of sample units is determined in the ESS see section 10.4 in the ESS Round 8 Survey Specification, http://www.europeansocialsurvey.org/docs/round8/ESS8_project_specification.pdf).

The percentage of all selected sample units that participated in the survey was lower in all mixed-mode designs compared to the main stage face-to-face ESS. Some of these differences were rather striking, like for most countries in Study 4 (*Feasibility of ESS telephone interviews*), where response rates were two to three times higher for the face-to-face survey than for the alternative telephone design (see Figure 16.). Even though these differences might in part be attributed to variation in the invitation protocol between the two mode designs (e.g., sample units assigned to the alternative mode received lower incentives or did not receive a pre-notification letter), it is unlikely that the large differences in response are entirely (or even mainly) due to those design differences.

The mixed-mode data collection designs of studies 5 (*Feasibility of mixed-mode designs: sequential vs concurrent*) and 6 (*Feasibility of mixed-mode designs across countries*) did better than the single alternative mode from Study 4 (*Feasibility of ESS telephone interviews*), but response rates were still lower than for the face-to-face standard ESS sample. In Study 5, the response rate in the standard ESS in The Netherlands was 52%, but was somewhat lower for the two alternative designs: 43% for the sequential design, 44% for the concurrent design (see Figure 16.), even though the recruitment process was carried out by interviewers, either by telephone or in person (Eva et al., 2010).

In the UK, the web + face-to-face design implemented in Study 6 resulted in a response rate loss of 17 percentage points compared to the standard ESS (55%, see Figure 16.)[6], in spite of the additional monetary incentive offered to web respondents conditional on completion (£15/£35 pounds, depending on the experimental condition).
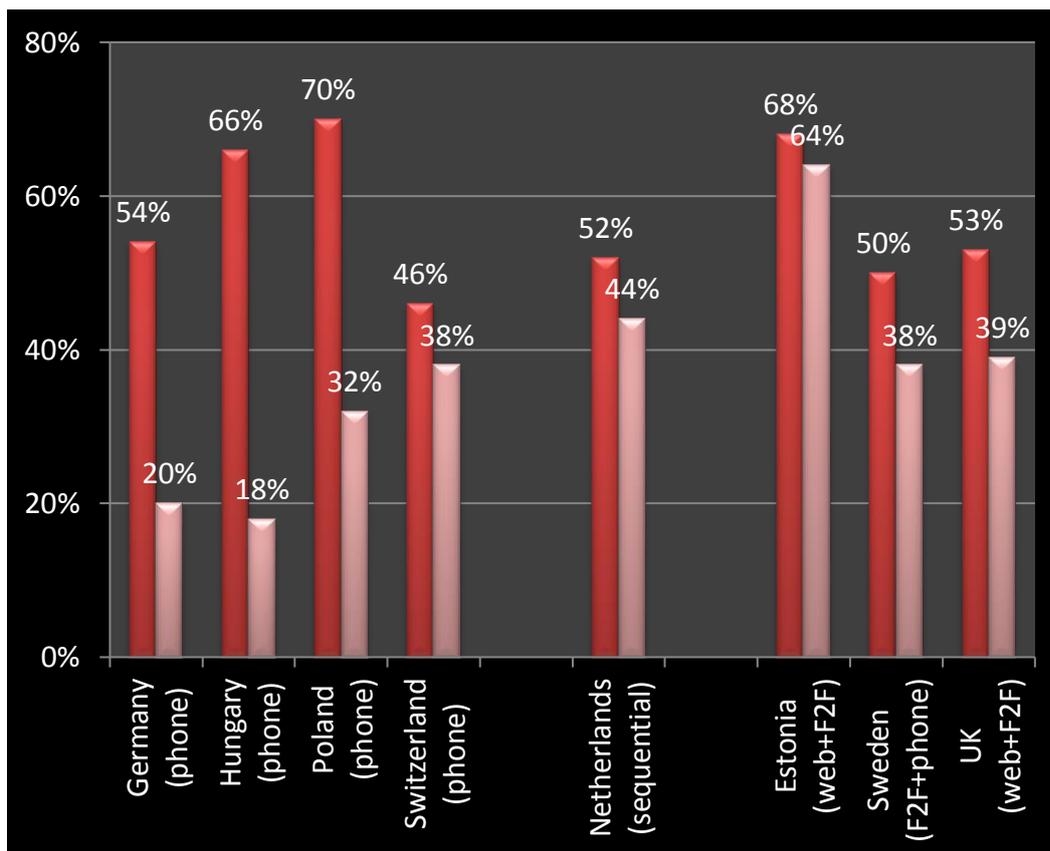
Figure 16.1 Response Rates (RR1) for Studies 3, 4, and 6

---

[6] Note that the mixed-mode experiment was conducted by a different fieldwork agency (NatCen) than the standard ESS (IPSOS). NatCen had conducted the ESS before and since, but achieved a substantially lower response rate for the standard ESS in Round 7 than IPSOS achieved in Round 6.

The loss was much lower in Estonia, where the mixed-mode design achieved a response rate of 64%, only 4 percentage points lower than the standard ESS (68%, see Figure 16.). This is particularly striking as respondents completed the almost 1-hour long ESS survey, which suggests the length of the questionnaire on-line is not a barrier for web implementation, in Estonia at least. The response rate in the Swedish response-enhancing design (38%) was much lower than in the standard ESS design (50%). However, compared to the mixed-mode sample, additional resources (extra incentives and visits) were allocated to the standard ESS sample to try to achieve the highest possible response rate; therefore, the lower response rates may be partly due to this difference.

Overall, there were no apparent advantages from offering additional modes of data collection to respondents in terms of response rates. Response rates were low in all mixed-mode designs compared to the face-to-face design, although the Estonian experiment rates were similar to the face-to-face rate. However, the Estonian mixed-mode survey benefitted from a large internet penetration rate and a sampling frame with named individuals to whom a mailed invitation could be sent. It is doubtful whether these findings could be replicated in countries without these features.

Refusal rates

Results were mixed within and across studies when it came to refusal rates. Respondents were more likely to refuse to participate in the telephone survey than in the standard face-to-face ESS in Germany, Hungary, and to a lesser extent in Poland (Study 4, see Figure 16.). However, they were less likely to refuse telephone invitations compared to face-to-face invitations in Switzerland (Study 4) and Sweden (Study 6), both countries where telephone surveys are generally more prevalent than in other ESS countries. Similarly, the web + face-to-face designs led to higher refusal rates in Estonia and UK (Study 6) but not in the Netherlands (Study 5), where refusal rates for the face-to-face sample were quite high (43%) compared to only (34%) in the sequential mixed-mode approach. This suggests that, in some countries, offering additional modes of data collection can persuade sampled individuals who would otherwise have actually refused to take part.

The lower refusal rates in Switzerland may be related to the survey tradition in this country. Since the early 1980s, high-quality surveys in Switzerland have been carried out by telephone (Ernst Staehli, 2012). Therefore, Swiss target respondents may have assigned less importance and legitimacy to face-to-face requests than they did to the telephone request. Unfortunately, it is not possible to test this hypothesis with the existing data.
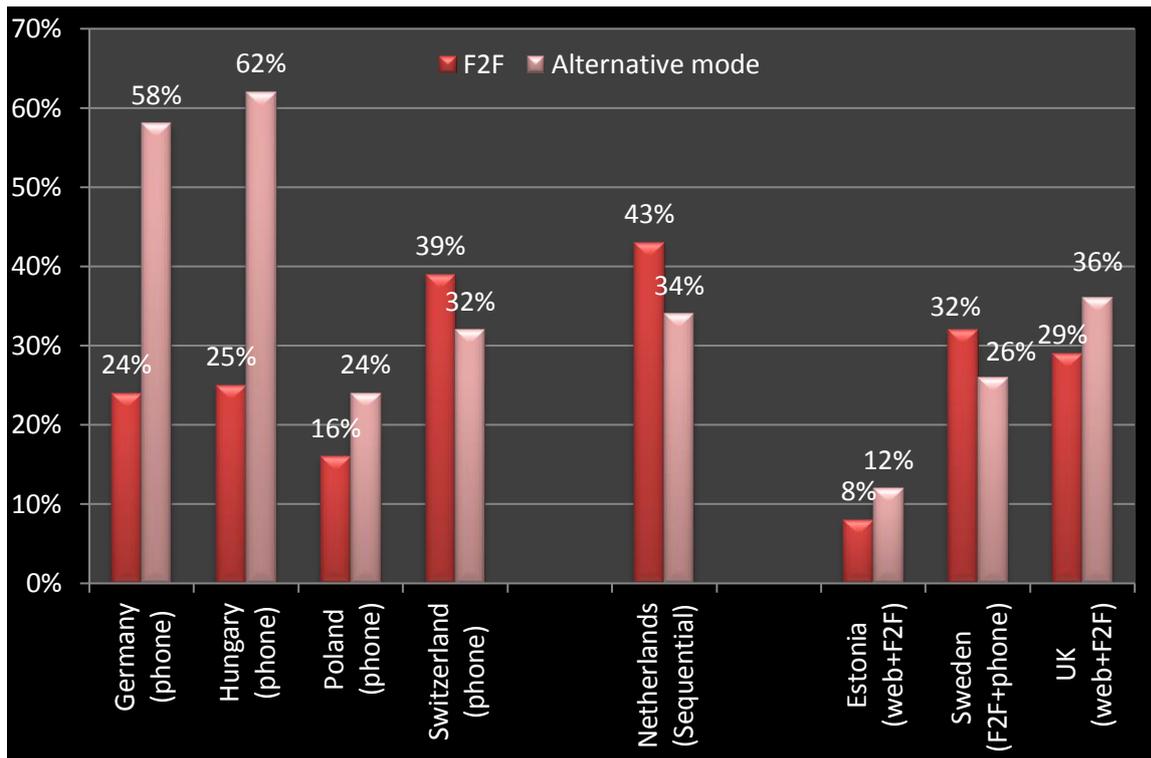
Figure 16.2. Refusal Rates for Studies 4, 5, and 6
Noncontact rates

In most countries and studies, differences in contact rates across modes were a lot smaller than refusal rates, with the exception of Germany and Poland (Study 4), where reaching sample units was considerably more difficult when contact was made via telephone than in the standard ESS (see Figure 16.). Contact rates were also slightly lower for the alternative mode design in Hungary (Study 4), the Netherlands (Study 5) and Estonia (Study 6), but no differences were observed in Switzerland (Study 4) or Sweden (Study 6). Therefore, among the countries that participated in these studies, increased refusal to participate played a larger role in the lower response rates than lack of contact with respondents.
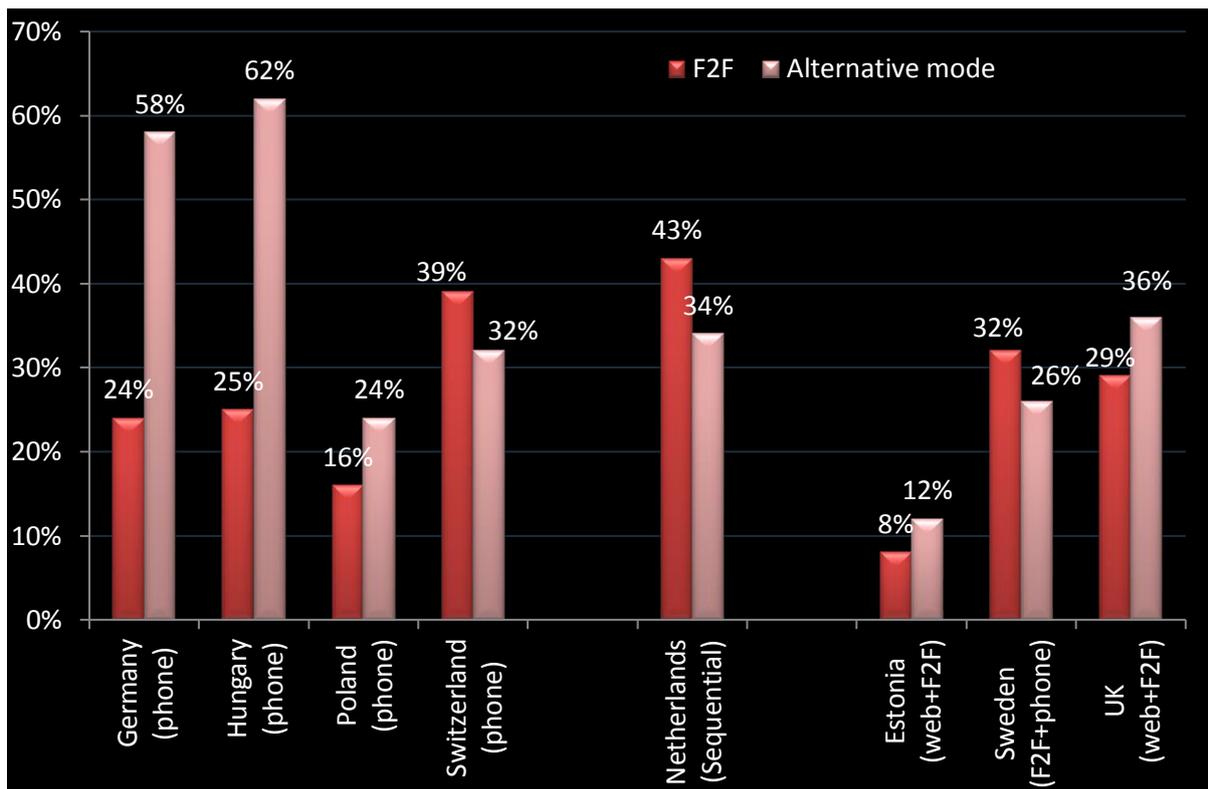
Figure 16.3. Household Contact Rates for Studies 4, 5, and 6

Sociodemographic composition of realised samples and nonresponse bias

Outcome rates can help a researcher to understand why certain sample units are not part of the final, realised sample. However, nonresponse bias analysis is necessary to assess the extent to which nonresponse affects survey estimates. To assess nonresponse bias, one would ideally compare respondents to nonrespondents with respect to the variables of interest. However, this information is typically unavailable. For the ESS studies, we used an alternative approach where we compared the sample of respondents obtained with each mode design to estimates from the population with respect to a few socio-demographic variables, providing an assessment of the extent of nonresponse bias for those variables[7].

In Study 2 (*Measurement differences: Finding the cause*), the face-to-face and telephone samples were compared with respect to five variables: age, gender, education, having a paid job, and occupation (Jäckle et al., 2006). Telephone respondents included significantly fewer males (33% vs. 40%), fewer manual workers (25% vs. 36%) and fewer individuals with lower education (48% vs. 55%) than face-to-face respondents did. No differences were found for the age or in the proportion of respondents having a paid job. Population estimates for Lisbon and Budapest were not available, but on the basis that most populations in the world have a nearly even divide in the proportion of males and females, it could be argued that the gender proportion in the telephone sample is probably more biased than in the face-to-face sample.

---

[7] This comparison is not a 'pure' measure of nonresponse bias, given that the sampling frame and sampling design can also affect the final response composition, and that the socio-demographic variables were measured using different modes and somewhat different wording. Based on the rigorous sampling procedures followed in the ESS, and on the simplicity of some of the demographic questions we use, we assume here that most of the differences between the realised sample and the population estimates are due to nonresponse effects.

In Study 4 (*Feasibility of ESS telephone interviews*), we compared face-to-face samples to telephone samples in six socio-demographic variables: gender, age, education, being in paid work, occupation, and area of residence (Roberts et al., 2009). Countries differed on how similar the realised sample in the telephone survey was to the realised sample in the standard ESS survey. Respondents to the German telephone survey were less likely to be in paid work than respondents to the German face-to-face survey (42% vs. 53%), but the opposite was found in Hungary (65% vs. 42%). Years in education and paid work differed significantly in almost all countries. In Germany, Poland, and Hungary, the telephone sample showed higher educational attainment than the face-to-face survey. The Hungarian samples showed the largest number of differences, for five of the six variables. The Swiss samples only differed significantly in percentage of the sample that was in paid work, but the difference was striking: 34% in the telephone survey compared to 62% in the face-to-face survey. Of particular concern here are the cross-national differences in how mode impacts the sample composition.

In Study 5 (*Feasibility of mixed-mode designs: sequential vs concurrent*), samples were compared with respect to age, gender, urbanisation, size of household, occupation and educational level (Eva et al., 2010). In addition, data were compared to population estimates from Statistics Netherlands (Centraal Bureau voor de Statistiek, 2008) in the first four of those variables. There were few differences between the composition of the face-to-face realised sample and the mixed-mode alternatives: the mixed-mode samples included fewer women with low educational level (20.4%) than the face-to-face sample (23.4%), and more women of middle educational level, 21.4% in the sequential deign compared to 17.1% in the face-to-face sample. However, these differences were rather modest.

Differences were larger when comparing the ESS standard and experimental samples to the population estimates. All ESS samples under-represented males under 45, females over 75, single person households, and people living in urbanised areas. The face-to-face sample was closer to the population values with respect to age and gender, and the concurrent sample differed most from the population estimates for all variables.

Comparing the face-to-face sample to those respondents from the sequential design that completed the web survey (the 'web sample') can provide a simulation of what would happen if the ESS were implemented as a web survey. In terms of sample composition, the web sample included fewer respondents over 65, especially among females (4.6% compared to 12.3% in the face-to-face sample), fewer men with higher education, fewer respondents from highly urbanised locations, fewer 1-person households and fewer retired respondents. Therefore, the web sample exacerbated the nonresponse bias observed in the face-to-face sample as compared to the population estimates rather than reducing it.

Study 6 – Feasibility of mixed-mode designs across countries. Comparing the profiles of the responding samples in each mode design with external population data, we found broadly similar distributions between the 2011 Census data in Estonia and the UK and the ESS samples, both from the face-to-face and mixed-mode designs (Villar et al., 2014). However, the mixed-mode sample showed greater bias relative to Census statistics than the face-to-face sample. In the UK, all samples attracted fewer respondents from the youngest and oldest age groups and more people looking after the home or family compared to the UK census data, but the difference was larger in the mixed-mode sample. In Estonia, only the mixed-mode sample under-represented people in paid work and those born outside Estonia, whereas the CAPI-only face-to-face sample was not problematic.

Summary

In line with previous research, evidence from these four studies suggests that mixed-mode data collection designs are unlikely to increase response rates or reduce nonresponse bias in survey

estimates compared to data collected face-to-face. In fact, overall they suggest that response rates would decrease and nonresponse bias would probably increase. Participation was lower for the mixed-mode designs mostly because of higher refusal rates rather than because of lower noncontact rates, which suggests that respondents can be contacted using different channels but may still be unwilling to take part in an hour-long survey over the telephone or on the web, or find it more difficult to refuse an in-person request from a persuasive interviewer. When there were differences in the composition of the realised sample, the sociodemographic profile of the face-to-face survey design was usually closer to the population estimates than the mixed-mode design.

Measurement error in the ESS mixed-mode studies

Multiple studies have found that collecting data in different modes can lead to measurement differences (e.g., Christian, Dillman, & Smyth, 2006; de Leeuw, 1992); the same respondent answering the same question may give different answers depending on whether the question is asked in person, over the telephone, on a web browser or on paper. In fact, in-person surveys sometimes switch to self-administered modes for sections asking sensitive questions (Couper, Tourangeau, & Smith, 1998; Tourangeau & Smith, 1996), on the grounds that reporting will be more accurate that way. However, it is less clear whether data collected from respondents who answer in different modes should be combined to compute survey estimates. If modes can attract respondents with different characteristics (e.g., a survey conducted via SMS may attract younger people compared to a face-to-face survey), any observed differences between those groups could be due to measurement differences related to mode, rather than to actual differences related to the variable of interest. This could in turn affect interpretation of survey findings. Similarly, if countries in a cross-national survey collect data using different modes, differences in survey estimates for each country could be related to actual differences in those estimates or to differences in how respondents answered to the specific mode of data collection. Therefore, the question of measurement differences across modes is of the utmost importance for cross-national research.

Multiple factors can help explain measurement differences across data collection modes (De Leeuw, 2005; Tourangeau, Rips, & Rasinski, 2000). From the start, the ESS mixed-mode research programme paid special attention to measurement effects to better understand these mechanisms. Three of the ESS studies investigated the effect of mode on estimates by designing experiments that allowed disentangling these effects from the other sources of error like nonresponse, coverage or processing. Observed differences in estimates can thus be attributed to the effect of mode on measurement. In the feasibility studies, observed effects can be due to differences in the composition of the realised sample or to differences in measurement. However, applying different approaches, one can try to account for selection effects and gain insight by comparing survey estimates across modes in those studies. In addition, measurement effects may also be assessed by observing respondent behaviour across groups of items. For example, item nonresponse, acquiescence (the tendency to agree with survey statements), and nondifferentiation of response options have often been used to assess measurement effects (e.g., Schuman & Presser, 1981). In this section, we will evaluate the effects of mode on univariate estimates, on multivariate estimates, on response behaviours and on data quality. Survey errors are item- and statistic- specific (Groves, 1989): mode effects might be found in univariate estimates of one variable (such as averages and percentages) and yet leave estimates of other variables or multivariate estimates including that variable unaffected. For a survey programme like the ESS, an important concern was the effect that switching mode designs would have across different questions, estimates, and multivariate analyses that data users may want to carry out. Furthermore, it is important to evaluate how mode effects could affect the interpretation of changes observed across time.

Effect of mode on univariate estimates

Effect of mode on univariate estimates: measurement studies

In two of the studies (1 and 3), respondents answered the same set of questions in two different modes. In both studies, the percentage of respondents who gave the same answer on both occasions was relatively low. In study 1, the percentage of identical answers ranged from 52% for the income question to 76% for self-reported belonging to a discriminated group. In Study 3, the average matching percentage across all attitudinal questions was 52%, ranging from 28% in a question about the effect of immigrants on cultural life of a country, to 70% matching for questions about perception of different types of crime. Behavioural questions led to higher levels of matching, ranging from 38% to 98%. Attitudinal questions (which are prevalent in the ESS) showed lower average consistency than behavioural questions, in line with previous research. Nonetheless, some of the differences between attitudinal and behavioural questions could be related to other formal differences between the questions like the number of response options. For example, 4 out of the 11 attitudinal questions had 11-point answer scales. It would be reasonable to expect that the larger the number of response options, the easier it will be for respondents to provide a different answer on two occasions, partly because the meaning of the two scale points may become closer. Questions with more response options indeed exhibited lower consistency across the two modes, and this was true both for attitudinal and behavioural questions (Jäckle et al., 2006).

Study 2 *(Measurement differences: Finding the cause)* investigated the mechanism behind measurement differences between telephone and face-to-face data collection. Comparing response distributions of 33 items across the three experimental groups (telephone, face-to-face with showcards, and face-to-face without showcards), 17 of the 33 items were affected by mode of interviewing after controlling for differences in sample composition due to age, gender, educational qualification and occupation. In 12 cases, telephone estimates were different to face-to-face estimates, but no differences were observed between the two face-to-face versions (with vs. without showcards), implying that mode differences were not driven by the questionnaire adaptation for telephone mode related to the sensory channel through which the response options were conveyed, but to other mechanisms related to mode (Jäckle et al., 2006). In all but one case (household income), the observed mode effects were comparatively small (no greater than one standard error) and would be unlikely to affect the conclusions of analysts using mixed-mode data.

Effect of mode design on univariate estimates: feasibility studies

For each feasibility study, large numbers of items were compared across mode designs. The goal was to estimate the proportion of ESS questions that would be affected if mixed-mode designs were allowed. The proportion of significant differences varied across studies[8], but in all cases and mode designs it was higher than what would be expected by chance alone. Out of the 43 wellbeing variables analysed in study 4, 31 (72%) showed statistically significant differences across the two samples, and a further 4 were marginally significant. This finding was replicated when each country was analysed separately, although fewer of the differences were significant given the smaller sample sizes. In the feasibility study conducted in the Netherlands (study 5), we compared estimates for 230 items as measured using the standard ESS face-to-face approach to the mixed-mode sequential design and the mixed-mode concurrent design. The concurrent mixed-mode sample featured differing distributions in 25 items (11%), whereas in the sequential mixed-mode sample 38 items (17%) had

---

[8] In each study, an appropriate regression model was fitted for each item, with the survey item as the dependent variable and a dummy variable 'survey design' as the independent variable, where 'survey design' was coded '1' for cases from the mixed-mode sample, and '0' for cases from the face-to-face survey. The regression model differed depending on the measurement level of the item: for questions with 6-point-scales, 11-point questions or ratio-scale measurement, we used ordinary least squares regression; for questions with 4- or 5-point scales and other ordinal scales, we used ordinal logit regression; for dichotomous items, we used logistic regression; and for nominal items, multinomial regression. Data were weighted, using weights inversely proportional to selection probability, and the 'svy setup' command in STATA was used to compute appropriate standard errors in each regression model.

distributions significantly different from the face-to-face sample. Finally, in study 6 we compared distributions of 128 questions across mode designs. Mode effects were present in numerous questions for all countries and across all mode comparisons. Effects were larger when comparing web and face-to-face in Estonia (63%) and the UK (83%) than when comparing telephone to face-to-face data in Sweden (32%, see Figure 16.). This may be partly due to the smaller Swedish sample, but differences between face-to-face and the alternative mode (telephone) were smaller overall in Sweden than in the other countries.

Significant differences were more common when comparing respondents to two single different modes (e.g., web vs face-to-face) than when comparing mixed-mode designs to face-to-face (see Figure 16.). Differences were also largest for estimates of personal wellbeing, followed by questions from the core ESS questionnaire about political attitudes, political participation and attitudes towards immigrants. These are some of the most frequently used items on the ESS according to ESS user statistics (http://www.europeansocialsurvey.org /about/user_statistics.html) and bibliographic studies (Malnar, 2015).
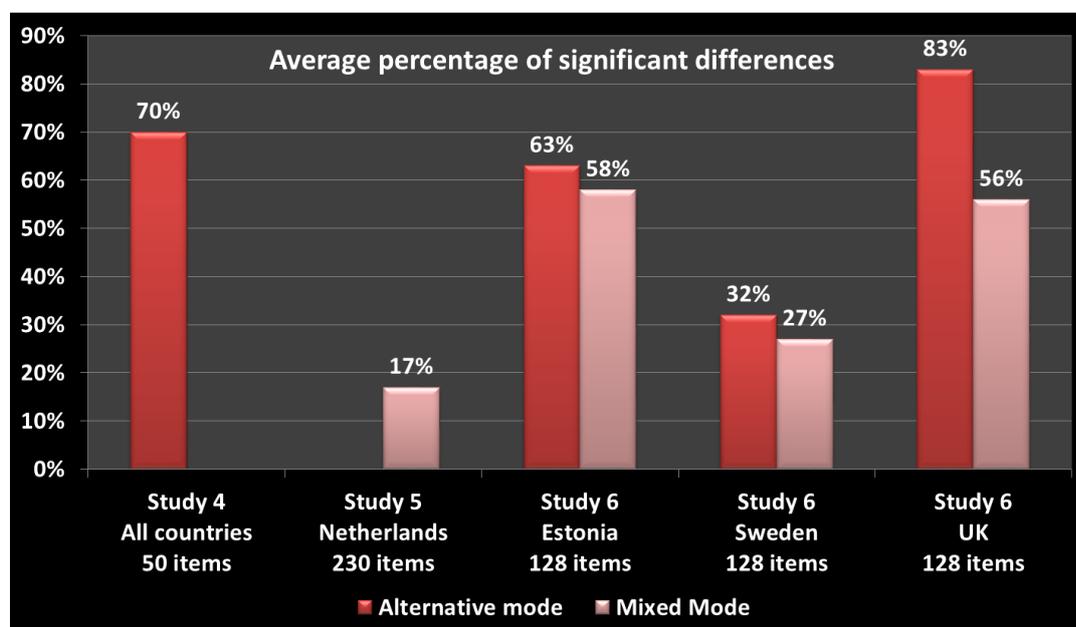
Figure 16.4. Percentage of variables showing significant differences across modes

To estimate the size of the differences, the average absolute difference between mode designs was computed for different sections of the questionnaire and different types of questions in study 6 (see c)[9]. Differences between the face-to-face data and the alternative mode designs were moderate but non-ignorable, ranging from 2.1 to 8.3 percentage points for categorical variables. The smallest difference was found when comparing estimates on questions about wellbeing in Sweden, in striking contrast to the large differences in these questions between web and face-to-face in the UK and Estonia (see row 7 in Table 16.). Coincidentally, these comparisons also reflected the smallest and largest differences for the quantitative questions.

---

[9] For questions with 6-point, 11-point, or ratio answer scales, the difference between the means was compared; for ordinal and categorical questions, the absolute difference for the category with the highest proportion of respondents (modal category) was computed.

Table 16.3. Average absolute error by section, mode, and country

	Estonia		UK		Sweden	
	F2F vs web	F2F vs MM	F2F vs web	F2F vs MM	F2F vs phone	F2F vs MM
Section A (5 quantitative)	0.3	0.3	0.3	0.3	0.4	0.2
Section B (20 quantitative)	0.3	0.3	0.6	0.6	0.3	0.1
Section B (15 qualitative)	6.5%	3.0%	5.9%	5.2%	4.4%	2.2%
Section C (6 quantitative)	0.2	0.2	0.6	0.5	0.2	0.2
Section C (13 qualitative)	5.0%	2.8%	3.7%	2.9%	4.1%	2.5%
Section D (24 quantitative)	0.4	0.2	0.8	0.7	0.1	0.1
Section D (15 qualitative)	6.8%	4.4%	8.3%	5.7%	4.1%	2.1%
Section E (30 quantitative)	0.2	0.2	0.5	0.3	0.2	0.1
Average absolute error quantitative questions	0.3	0.2	0.6	0.5	0.2	0.2
Average absolute error qualitative questions	6.1%	4.4%	8.3%	5.7%	4.1%	2.3%

Effect of mode on multivariate estimates

Effect of mode design on multivariate estimates: feasibility studies

A number of statistical models involving political attitudes were tested in Study 5 (Martin & Lynn, 2011). In all the models tested, some model parameters differed across mode designs: when looking at political interest, just two out of sixteen examined variables showed a significant interaction effect with data collection design. However, when political interest is used as a predictor variable, there were significant differences in several models. The study suggested caution when comparing models from surveys carried out with different data collection designs (Martin & Lynn, 2011).

In study 6, bivariate correlations were compared between 30 items about perception of democracy. Out of 435 correlations, we considered how many were significantly different in the two mode designs (see Table 16.). As with univariate estimates, differences were smaller when comparing the mixed-mode sample to the face-to-face than when comparing the web sample to the face-to-face. In Estonia, 13.1% of the correlations were significantly different, whereas in the UK it was almost a third (32.3%). Larger differences in correlations ($p < .001$) represented only 2.5% of the 30 variables in Estonia but 9.7% in the UK. Overall, substantive conclusions in Estonia would be relatively close, but could lead to different relationships across the 30 variables studied and may cause problems when comparing statistical models in these countries. In the UK, depending on the variables studied, the substantive conclusions might be impacted.

Table 16.4. Significance of the difference in correlations across mode designs

	Estonia	UK

	(n = 435)	(n = 435)
	%	%
p < .001	2.53	9.66
.001 ≤ *p* < .01	2.76	10.34
.01 ≤ *p* < .05	7.82	12.18
.05 ≤ *p* < .10	5.75	7.59
p ≥ .10	81.15	60.23
Total	100.00	100.00

Finally, using studies 5 and 6, data quality for each mode design was compared using sets of questions measuring various concepts (Revilla, 2010, 2013). Overall, there were small differences across the three mode designs in terms of reliability and validity, and those differences did not show any one mode having consistently higher quality. The largest difference was found in quality of the political trust concept, where the face-to-face survey showed higher validity than the mixed-mode designs.

Effect of mode on response behaviours: item nonresponse, acquiescence, non-differentiation, and other response behaviours

In this section, we review findings from the measurement studies about the impact of mode on how respondents answer questions. A variety of response behaviours have been used as measures of data quality: item nonresponse, response styles such as acquiescence, and tendencies to select the same response option regardless of the real answer (e.g., extreme response style, nondifferentiation).

Effect of mode on response behaviours: measurement studies

Item nonresponse. In Study 1 *(Measurement differences across four modes)*, significant differences in item non-response were found for half of the 38 questions analysed. Item non-response rates were highest for self-completion (paper) mode in 12 questions. For the remaining 7 questions, item non-response was significantly higher among telephone respondents compared to other modes. In Study 2 *(Measurement differences: Finding the cause)*, item nonresponse rates were similar across the three groups; however, respondents interviewed face-to-face without showcards had higher overall item nonresponse rates, and significantly higher item nonresponse to the income question, with almost double the proportion of refusals than the telephone group. In Study 3, item nonresponse was again low in both modes, but selecting don't know, refusal, or no answer to the 35 analysed questions was more likely on the web than face-to-face, even though the web survey was much shorter than the face-to-face interview.

Non-differentiation, extreme response style, acquiescence, and order effects. In Study 1, *(Measurement differences across four modes),* for three out of five questions, there were significant mode differences in the tendency to select answer scale endpoints (extreme response style), but there was no obvious pattern: sometimes web mode obtained the highest proportion of extreme responses, sometimes face-to-face did. No differences were observed when comparing acquiescence response style across modes.

In Study 2 *(Measurement differences: Finding the cause),* telephone respondents were just as likely as face-to-face respondents to acquiesce, repeat the same response option across questions, and select the

first or last options. This suggests that showcards had little effect on data quality. Similarly, there was no evidence of increased non-differentiation or increased acquiescence among telephone respondents compared to the face-to-face respondents who were not given showcards. However, when comparing face-to-face interviews with showcards to telephone interviews, non-differentiation is significantly higher in the latter; potentially, not having the interviewer be physically present and not having showcards has a cumulative effect leading to higher non-differentiation (Jäckle et al., 2006).

Effect of mode on response behaviours: feasibility studies

Item nonresponse. Across all feasibility studies, item nonresponse was generally low in all modes, as it was in the standard ESS conducted face-to-face. Study 4 (*Feasibility of ESS telephone interviews*) found that item nonresponse was significantly higher in the standard face-to-face survey than in the telephone survey, although this difference was small. In Study 5 (*Feasibility of mixed-mode designs: sequential vs concurrent*), item nonresponse was very low in all groups and all modes, but slightly higher in face-to-face (4.3 'don't knows' and 4.7 refusals per interview on average) than telephone (2.8 'don't knows' and 2.9 refusals per interview on average) and web (0.1 'don't knows' and 0.2 refusals per interview on average). In Study 6 (*Feasibility of mixed-mode designs across countries*), there was higher item nonresponse in the mixed-mode design than in the standard ESS but showed opposite directions in the UK (higher for web respondents) and Estonia (higher for face-to-face respondents). The differences across countries in efficiency of the different mode designs underline how mixing modes in different ways could be problematic for cross-national comparisons. Furthermore, changes over time within a country could also cause difficulties.

Non-differentiation, acquiescence, response option positional preferences. In Study 6 (*Feasibility of mixed-mode designs across countries*), the proportion of respondents who selected the same response option in sets of items that shared the same answer scale was, in general, similar or lower in the overall mixed-mode design than in the standard ESS. When looking at respondents by mode, there was reassuringly no evidence of higher straight-lining among web respondents.

Mode, interview length and interview experience

In Study 2, interviews conducted by telephone were significantly (although slightly) shorter than those conducted in person (15.3 vs 17.5 minutes). In Study 4, actual duration was about 8 minutes shorter by telephone than face-to-face than when it was over the telephone for all countries but Poland, where interestingly it was 4 minutes longer.

Regardless of telephone surveys being seemingly shorter than face-to-face surveys, previous research (e.g., de Leeuw, 1992; Holbrook et al., 2003) and common sense suggested that an invitation to carry out the hour-long ESS survey would yield lower response rates than a face-to-face request for several reasons (see Groves, 1990 for a review of the possible mechanisms behind this difference). Therefore, Study 4 tested the feasibility of conducting the ESS over the telephone and whether shorter versions (45 and 30 minutes long) of the survey could help improve the expected low response rates. Response rates differed depending on the duration communicated during the invitation to participate: cooperation was higher for the 45-minute interview than for the one-hour long interview, increasing response rates by 5 percentage points in Germany and Poland, 4 in Hungary, and 2 in Switzerland. When respondents were first invited to a 30-minute interview followed by an additional 30-minute interview, response rates were highest for the 30-minute interview, but were significantly lower if we consider the total number of respondents who completed both 30-minute interviews. In addition, the 30-minute interview led to a response rate lower than the 45-minute interview in Poland (50.5% vs 55.6%). More importantly, response rates for the standard ESS face-to-face surveys were higher than for any of the three versions of the telephone survey in all countries. Even when considering only the

30-minute interview, cooperation rates were better in the face-to-face design than the telephone design in all countries except Switzerland.

In Study 2, 69% of respondents when asked whether they would be willing to continue the interview for longer, telephone respondents were more likely to be willing (69%) than face-to-face respondents (60%). It is not clear whether that willingness would turn into higher or lower response rates if the interview indeed continued, especially when less than half of the respondents to the 30-minute interview in Study 4 completed the second part of the survey.

A number of questions were asked in Study 2 to evaluate respondents perception of the survey experience. Face-to-face respondents were more likely than telephone respondents to report that they had felt uneasy about answering certain topics. However, respondents who felt uneasy were also more likely to report having answered truthfully. Based on answers recorded by the interviewer, telephone respondents were less likely to ask for clarification, less likely to be reluctant to answer, more likely to answer questions to the best of their ability and more likely to have understood the questions. It is unclear how accurate these interviewer perceptions were, especially when telephone interviewers had considerably fewer cues to base their assessments on.

Summary

This section highlights that mode does indeed have some impact on how respondents answer questions. It is not always clear which mode is best in terms of quality, but the differences between modes have the potential to damage comparability. Mode-related differences in a country's estimates across time could confound comparisons within and across countries. The ESS, as well as other cross-national projects, has built a time series using face-to-face data collection, and this single-mode design is likely to continue in some countries where other mode designs would not be optimal (or even possible). This has implications for the usage of ESS data, which are publicly and freely available. Advanced data users familiar with the ESS may quickly note and try and account for these changes, but few reliable methods are available to allow for mode-difference correction (Vandenplas, Loosveldt, Beullens, & Vannieuwenhuyze, 2014). Data users less acquainted with the ESS procedures and history may miss the switch in mode design altogether, and thus proceed analysing the data as if no mode effects were present.

The possibility of shortening the ESS questionnaire to facilitate completion in other modes has been discussed during the mixed-mode methodology programme. Offering a 45-minute questionnaire rather than 60 minutes could be beneficial for the telephone mode whilst administering two 30-minute sections would be damaging to the response rate. However, contrary to expectation, the 60-minute web survey did not appear to be a problem in Estonia in Study 6. The large web participation levels in Estonia and moderate participation levels in the UK suggest that long surveys can be conducted using self-administered modes. Nonetheless, the one-hour standard ESS face-to-face survey gets the highest cooperation compared to any single mode and to most mixed-mode designs.

Costs in the ESS mixed-mode studies

The ESS studies have shown that switching to mixed-mode data collection is unlikely to produce increases in response rates or improvements in data quality. In fact, reductions in response rates seem likely and additional mode effects could be introduced with each new mode permitted. However, one of the key advantages of a switch to mixed-modes is the possibility of substantially reducing *fieldwork* costs. Alternative modes may lead to reduced efforts to engage respondents, where fewer contact attempts are necessary to obtain a response from sampled units. In addition, certain modes with lower cost per respondent contact attempt could lead to overall lower cost per respondent. At the same time, these potential savings need to be weighed against any potential increased *fixed setup costs,* which, in a cross-national context, include central coordination costs as well as country-level costs. This

section summarises the evidence gathered in the ESS research programme regarding the effect of mixed-mode designs on overall costs.

Like many publically funded cross-national social survey programme's, the ESS has a rather complex funding structure. There are two distinct budget streams, one to cover each country's own national survey and coordination costs, and one to cover central coordination and dissemination costs. Central costs were historically met by the European Commission but have been funded collectively by participating countries since the ESS was established as a European Research Infrastructure Consortium (ESS ERIC) in 2013. Currently, it is not possible to transfer funds between these two different funding sources: costs for design of questionnaire and other fieldwork instruments, central quality control and data archiving and dissemination are met from the central budget, whilst all translation and data collection costs are met nationally. Consequently, different researchers are in direct control of each budget, and overarching changes to the design of the survey become more complex than they would be for a single budget study like that in many single nation survey projects.

In a complex budgetary context like the ESS's, mixed-mode data collection would likely lead to cost savings at the national level, where data collection agencies would reduce the amount they charge to national funders. However, one would expect higher costs for development of questionnaire instruments, quality control, fieldwork monitoring and data processing in multiple modes. Some of this investment might reduce fieldwork savings. Thus, central ESS costs could increase, as source instruments and other materials (such as instructions and invitation letters) would have to be designed and tested for multiple modes. If it proves feasible in future, the production of correction factors for mode differences would also need to be funded. The ESS mixed-mode methodology programme was not designed to simulate all of these different interactions but it is important to remember them when assessing the available evidence.

Unfortunately, there are many other difficulties of isolating costs differences by mode from the work conducted in the ESS mixed-mode methodology programme. As Villar and colleagues (2014) commented:

"Comparing the actual costs incurred to carry out the survey in each mode is fraught with difficulties. Some of the costs in a split-run experimental design are inevitably shared between the mode designs and any allocation of costs to each design will be arbitrary. Furthermore, the costs will be shaped by a variety of contextual factors that could differ greatly in a different situation (market buoyancy, relevant experience of the survey organisation and the interviewers, etc). And then there is the additional difficulty of comparing costs cross-nationally (page 3)".

It is therefore helpful that general information about data collection costs was collated during the mixed-mode methodology programme. In all ESS countries, face-to-face fieldwork was the most expensive mode (see Figure 16.). Information was collected by asking national statistical institutes in ESS countries to provide information about the relative costs of different data collection modes (Roberts, Eva, & Widdop, 2008). In addition, countries reported that face-to-face fieldwork costs were increasing, even simply to maintain the same response rates. Often in the most expensive fieldwork countries for ESS fieldwork, the difference between face-to-face and the next most expensive mode is particularly large (see

Table 16. 5). In Germany, for example, reported costs for web interviewing were almost 6 times cheaper than those for face-to-face interviewing, whereas in Switzerland they were almost 20 times cheaper (Roberts et al., 2008).

Table 16.5. Relative costs of fieldwork using different modes.

Country	F2F	Tel.	Postal	Web
ESS Field Agencies				
(~=National Statistics Institute)				
Austria~	100	30	25	-
Belgium	100	70	-	25
Belgium~	100	80	83	62
Cyprus	100	50	-	-
Cyprus~	100	80	50	40
Denmark	100	35	25	25
Finland~	100	50	35	-
Germany	100	43	20	15
Hungary	100	80	60	50
Iceland	100	40	25	15
Ireland	100	65	50	*
Italy	100	60	70	50
Norway~	100	50	25	20
Poland	100	-	40	-
Portugal	100	75	50	-
Romania~	100	60	-	-
Russia	100	63	-	-
Slovakia	100	85	35	65
Slovenia	100	30	25	20
Spain	100	60-70	30	30
Sweden	100	60	30	-
Switzerland	100	50	-	-
Switzerland~	100	40	20	5
Ukraine	100	60	50	40

Note: Question phrasing was: 'To help give us an idea of the relative costs of fieldwork using different modes of data collection, please estimate the average cost of conducting a survey of a random probability sample of the population using the modes listed below. (Assume 1,000 achieved interviews and a 20 minute questionnaire).You do not need to give the actual cost estimate. Simply describe the relative costs of modes b, c, and d (below) as a percentage of the cost of mode a (a survey using face-to-face interviewing). Please enter your answers below.

Fieldwork Costs and Efforts

The mapping exercise carried out before Study 4 (see Figure 16.), showed that in countries where face-to-face was highest, the gap with respect to the next most expensive mode also tended to be larger—thus, the potential for cost-savings using a different mode was larger. Conversely, in countries where face-to-face fieldwork was not as expensive, the difference in cost by mode was smaller (Roberts, Eva and Widdop, 2008). As a result, it was expected that those countries where data collection was the most expensive would be more interested in switching to an alternative, cheaper, mode. Indeed, countries such as Netherlands and Switzerland have been the most eager to see mixed-mode data collection on the ESS. However, there has not been such pressure from countries where there is no sampling frame of individuals that would allow mixed-mode data collection without sending an interviewer to do initial respondent selection. For instance, there has been no real pressure from Germany and the UK to switch modes thus far despite face-to-face data collection costs being particularly high in these nations.
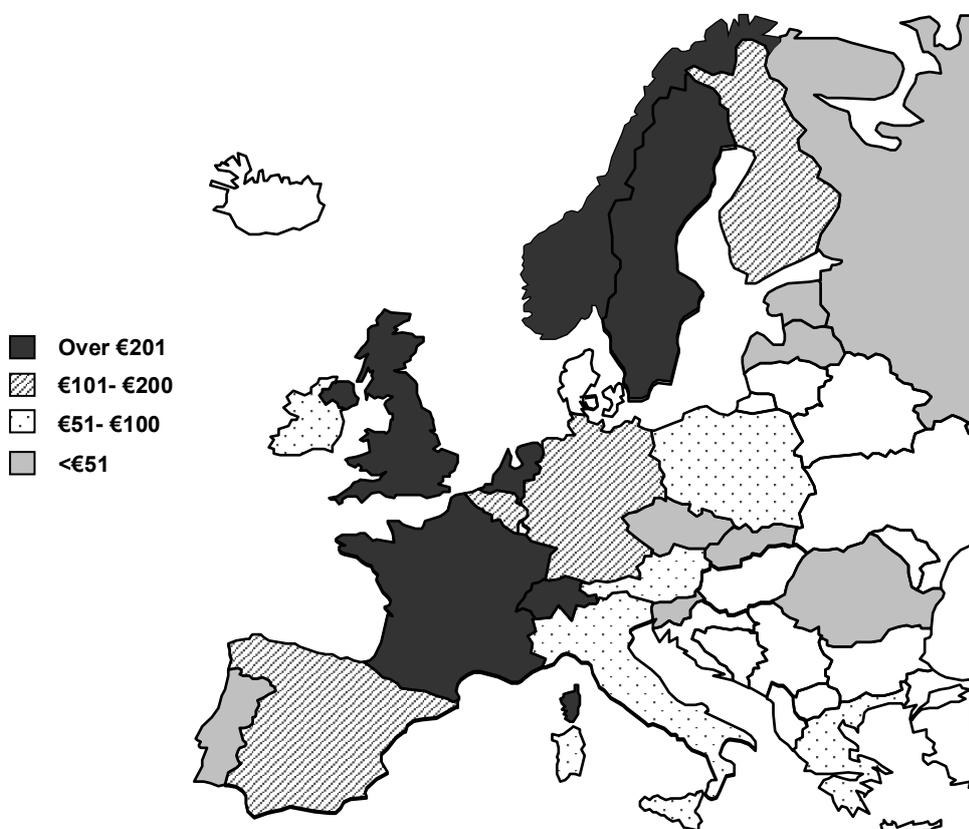
Figure 16.5. Mean costs per successful interview (based on ESS Round 3 2006 figures). Three studies in the ESS mixed-mode methodology programme directly addressed the costs issue: Study 4 (*Feasibility of ESS telephone interviews*), Study 5 (*Feasibility of mixed-mode designs: sequential vs concurrent*) and Study 6 (*Feasibility of mixed-mode designs across countries*). Survey costs depend on a range of factors. In the ESS, various elements that would affect cost are fixed, such as length of questionnaire or sample size. Other factors will differ by country—such as the size of the country, the population density, mode coverage and available sampling frame—and by survey agency, such as interviewer experience and training, and the available software.

The cost findings from these studies demonstrate that mixed-mode or alternative mode data collection have the potential to save costs. The telephone studies suggest that cost savings might be extensive in

countries where face-to-face fieldwork is most expensive (like the UK and Switzerland) but non-existent in countries like Poland, where there would be no savings at all (Roberts et al., 2009). However, as we approach the end of the ESS mixed-mode methodology programme it appears that telephone is largely seen as a mode of the past rather than the future. Therefore, the real focus now is on whether web surveys or mixed-mode web and face-to-face surveys can save money.

The evidence from the ESS studies in the programme suggested that a web + face-to-face mixed-mode design has the potential to save field data collection costs without harming the survey response rate in Estonia. However, the equivalent results from the UK suggest that it may prove impossible to realise this potential in all or probably most ESS countries in the near future since extensive face-to-face activity would still be required to achieve an acceptable response rate. The differences between Estonia and the UK in this respect are really striking. In Estonia, effectively the same response rate as the main single-mode face-to-face ESS was achieved at a substantially lower cost. In the UK, a significantly lower response rate was achieved despite extensive field efforts that resulted in very little, if any, cost savings(Villar et al., 2014). It seems quite likely that such variability in experiences with mixed-mode data collection would also be observed between other ESS countries. The experiment in the Netherlands is somewhere between the two, showing that substantial cost savings are possible but once again at the cost of a substantial drop in the response rate (Eva et al., 2010).

For a cross-national survey like the ESS, there are the additional development and coordination costs that would be incurred at a central level if more modes were allowed. A switch to mixed-mode interviewing would save money in some countries and in a small group these savings could be rather substantial. However, a switch would only be possible in some countries, would only save money in some countries where a switch was possible, would lead to a loss of quality in most countries where a switch took place, would undermine cross-nationally and longitudinal comparability and would increase the central costs.

Implementation challenges in the ESS drawing on the mixed-mode studies

The ESS, like many cross-national social surveys, has a complex organisational and implementation structure. On the one hand, it can be seen as one survey in multiple countries in the sense that each country contracts and carries out their own fieldwork. On the other hand, the ESS is more like a single survey across countries. The project is a compendium of national surveys conducted under a common framework, where the CST specifies how the survey should be conducted, sets targets, provides key documents like the questionnaire and contact forms, and supports and monitors the work of ESS national teams. The CST harmonises the datasets then makes the data and documentation available and provides supporting tools such as data analysis platforms and weights. As noted earlier, the ESS also operates an input harmonised model of data collection trying to keep all essential survey conditions the same and organising necessary variation in a way that is equivalent across countries.

Introducing mixed-mode data collection would require designing future ESS questions in a uni-mode manner, so that the same exact question can be used in all modes. Questions from previous rounds would have to be re-drafted and adapted to allow collection in new modes. Taking into account more than one mode during the question-design process would increase the intricacy of an already complex process where potential language and cultural effects are being considered. In addition, the existing core questionnaire would have to be reviewed and parts of it rewritten to make it suitable for use in, say, a web survey. This would mean that the time series that the ESS has been creating since 2002 would have to be disrupted, with the extent varying between countries and different questionnaire items.

Further decisions would likely ensue from a change to mixed-mode. If web data collection were allowed, the ESS would need to decide how to handle web survey programing efficiently. A central programing tool would be most convenient, but may have implications related to data protection

issues associated with collecting data across borders. For efficient management of data collected with different modes, such web survey tool should preferably be compatible with the CAPI system used for face-to-face data collection. Therefore, a decision to select a central web tool could entail all countries adopting the same CAPI system, whereas current procedures allow countries to use their preferred system. Even though this unifying approach may indeed provide advantages, the CST would need to make an investment to evaluate, select, and prepare training materials for the tool.

The ESS also harmonises and makes data from interviews and data about interviewer efforts to contact respondents publicly available. Additional steps would be needed to prepare and process data files from different modes. Survey agencies and national teams involved with some of the experimental studies in the ESS mixed-mode programme reported that the additional complexity caused difficulties and required additional resources. Problems with sample management were experienced on at least one of the ESS experiments, highlighting the overall issue that mixed-mode data collection increases the likelihood of implementation error simply because it involves more instruments.

Whereas none of the issues identified in this section are insurmountable and some would be helpful innovations in their own right, they would require additional coordination time and resources for the ESS Core Scientific Team, who would need to carefully evaluate and plan for such a change.

Strengths and Limitations

The ESS mixed-mode methodology programme has been one of the most extensive exercises looking at the potential impact of introducing such a change on a single survey and certainly for a cross-national survey. That said, experiments were only conducted in ten countries across the years, and data collection designs differed across these. Therefore, the results may not be generalisable to the whole of the ESS. In addition, implementation errors were identified in some of the studies that may impact the conclusions drawn. The individual reports for each of the studies, cited throughout this chapter, provide more details on these issues.

Conclusion

The face-to-face standard ESS mode design fared better than all the alternative designs we tested in terms of survey participation and net sample representativeness. So whilst it remains the most expensive mode, face-to-face is still the highest quality mode in these respects. The findings relating to data quality are more complex due to the lack of external record data or population benchmarks for most variables, and to the difficulty of disentangling 'selection effects' from 'measurement effects' (Vannieuwenhuyze & Loosveldt, 2013; Vannieuwenhuyze, Loosveldt, & Molenberghs, 2010, 2014). However, numerous differences between mixed-mode and the standard ESS face-to-face survey were observed, and these differences remained even after controlling for demographics. Therefore, using mixed-mode designs in some countries while others continue using a face-to-face design would introduce alternative explanations to differences not present in the current data collection model. Furthermore, if the proportion of respondents participating in each mode varies over time—which would be likely unless carefully controlled—cross-time comparisons could also be jeopardised, further complicating comparability of the ESS datasets.

In recent years, there has been increasing evidence of differential interviewer effects across countries leading to questions about the gold standard status of face-to-face interviewing (Beullens & Loosveldt, 2014). In addition, whilst there were many differences between the mixed-mode designs and the standard ESS face-to-face survey, initial analyses of effect size suggest that these were relatively small (less than 5 percentage points), although some larger differences were noted in questions about personal wellbeing and income. At the same time, some of the larger differences may be reducible using careful question design (Dillman, 2009) although this would require disrupting the time series for repeat items. In addition one of the key challenges remains that there is no cost

effective way to correct for measurement differences arising from data collection mode (Vandenplas et al., 2014).

Like with all survey design decisions, it is necessary to accept some trade-offs between different sources of survey error. As the underlying goal of the ESS is to keep measurement as similar as possible across countries and rounds, countries using a single mode different from face-to-face would potentially jeopardise comparability to a larger extent than countries using mixed-mode designs that include face-to-face data collection. As recently as 2011, no country appeared ready to switch to a single alternative mode of data collection than face-to-face (Martin, 2011) and there is no indication that has changed.

In the current context of the ESS, the limited cost saving available in terms of fieldwork itself cannot justify a switch to mixed-mode data collection that in most countries and on many key quality dimensions would reduce quality. Until the time comes when more countries have access to named person samples, more people are willing to take part in surveys without being persuaded by an interviewer, and cost effective mechanism are developed to allow mode differences to be corrected, a high quality cross-national survey like the ESS would be unwise to make such a switch. In fact, on the advice of the ESS ERIC Methods Advisory Board the ESS Core Scientific Team has rejected such a switch for the near future. Instead, preparations are focused on the possibility of recruiting ESS face-to-face respondents to a probability-based web panel, which could overcome many of the serious problems mixed-mode data collection poses.

References

Beukenhorst, D. (2012). The Netherlands. In *Telephone surveys in Europe: Research and Practice* (1st ed., pp. 17–24). Berlin: Springer-Verlag.

Beullens, K., & Loosveldt, G. (2014). Interviewer Effects on Latent Constructs in Survey Research. *Journal of Survey Statistics and Methodology*, *2*(4), 433–458. http://doi.org/10.1093/jssam/smu019

Callegaro, M., & DiSogra, C. (2008). Computing response metrics for online panels. *Public Opinion Quarterly*, *72*(5), 1008–1032.

Christian, L. M., Dillman, D. A., & Smyth, J. D. (2006). The effects of mode and format on answers to scalar questions in telephone and Web surveys. In *2nd International Conference on Telephone Survey Methodology*. Miami, Florida.

Couper, M. P., Tourangeau, R., & Smith, T. W. (1998). Collecting sensitive information with different modes of data collection. In M. P. Couper, R. P. Baker, J. Bethlehem, C. Z. F. Clark, J. Martin, W. L. Nicholls II, & J. M. O'Reilly (Eds.), *Computer assisted survey information collection* (pp. 431–454). New York: Wiley.

De Leeuw, E. (2005). To mix or not to mix? Data collection modes in surveys. *Journal of Official Statistics*, *21*(2), 1–23.

De Leeuw, E., Callegaro, M., Joop, H., Korendijk, E., & Lensvelt-Mulders, G. (2007). The influence of advance letters on response in telephone surveys: A meta-analysis. *Public Opinion Quarterly*, *71*(3), 413–43.

de Leeuw, E. D. (1992). *Data quality in mail, telephone and face to face surveys.* ERIC. Retrieved from http://eric.ed.gov/?id=ED374136

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). *Internet, mail, and mixed mode surveys: The tailored design method*. Hoboken, NJ: John Wiley & Sons.

Ernst Staehli, M. (2012). Switzerland. In *Telephone surveys in Europe: Research and practice* (1st ed., pp. 25–36). Berlin, Heidelberg: Springer-Verlag.

Eva, G., Loosveldt, G., Lynn, P., Martin, P., Revilla, M., Saris, W. E., & Vannieuwenhuyze, J. T. A. (2010). *ESSPrep WP6 – Mixed mode experiment. Deliverable 21 Final mode report* (No. 21).

Gaziano, C. (2005). Comparative analysis of within-household respondent selection techniques. *Public Opinion Quarterly*, *69*(1), 124–157.

Groves, R. M. (1989). *Survey errors and survey costs*. New York: Wiley-Interscience.

Groves, R. M. (1990). Theories and methods of telephone surveys. *Annual Review of Sociology*, *16*, 221–240.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2004). *Survey methodology* (1st ed.). John Wiley & Sons.

Harkness, J. A. (2003). Questionnaire translation. In J. A. Harkness, F. J. Van de Vijver, & P. P. Mohler (Eds.), *Cross-cultural survey methods*. Hoboken, NJ: John Wiley & Sons.

Harkness, J. A., Villar, A., & Edwards, B. (2010). Translation, adaptation, and design. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. Lyberg, P. P. Mohler, … T. Smith (Eds.), *Survey methods in multinational, multiregional, and multicultural contexts* (pp. 115–140). Hoboken, NJ: John Wiley & Sons.

Holbrook, A. L., Green, M. C., & Krosnick, J. A. (2003). Telephone versus face-to-face interviewing of national probability samples with long questionnaires: Comparisons of respondent satisficing and social desirability response bias. *Public Opinion Quarterly*, *67*(1), 79–125. http://doi.org/10.1086/346010

Jäckle, A., Roberts, C., & 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 Series No. 2006-41). Institute for Social and Economic Research. Retrieved from https://ideas.repec.org/p/ese/iserwp/2006-41.html

Jowell, R. (1998). How comparative is comparative research? *American Behavioral Scientist*, *42*(2), 168–177. http://doi.org/10.1177/0002764298042002004

Jowell, R., Roberts, C., Fitzgerald, R., & Eva, G. (2007). *Measuring attitudes cross-nationally: Lessons from the European Social Survey*. SAGE.

Kish, L. (1949). A procedure for objective respondent selection within the household. *Journal of the American Statistical Association*, *44*(247), 380–387. http://doi.org/10.1080/01621459.1949.10483314

Kish, L. (1994). Multipopulation survey designs: Five types with seven shared aspects. *International Statistical Review / Revue Internationale de Statistique*, *62*(2), 167–186. http://doi.org/10.2307/1403507

Lynn, P., Haeder, S., Gabler, S., & Laaksonen, S. (2007). Methods for achieving equivalence of samples in cross-national surveys: the European social survey experience. *Journal of Official Statistics*, *23*(1), 107–124.

Lynn, P., Japec, L., & Lyberg, L. (2006). What's so special about cross-national surveys? *Conducting Cross-National and Cross-Cultural Surveys*, 7.

Lynn, P., Uhrig, N., & Burton, J. (2010). *Lessons from a randomised experiment with mixed mode designs for a household panel survey*. Working paper, Colchester, UK. Retrieved from http://research.understandingsociety.org.uk/publica tions/working-paper/2010-03

Malnar, B. (2015). *Science Communication and Monitoring. Task 9.1: Promoting ESS to user communities* (Deliverable).

Martin, P. (2011). *What makes a good mix? Chances and challenges of mixed mode data collection in the ESS*. Working Paper, London. Retrieved from https://www.city.ac.uk/__data/assets/pdf_file/0015/125133/CCSS-Working-Paper-No-02.pdf

Martin, P., & Lynn, P. (2011). *The effects of mixed mode survey designs on simple and complex analyses*. Working Paper, Centre for Comparative Social Surveys, London.

Olson, K., & Smyth, J. D. (2014). Accuracy of Within-household Selection in Web and Mail Surveys of the General Population. *Field Methods*, *26*(1), 56–69. http://doi.org/10.1177/1525822X13507865

Olson, K., Stange, M., & Smyth, J. (2014). Assessing Within-Household Selection Methods in Household Mail Surveys. *Public Opinion Quarterly*. http://doi.org/10.1093/poq/nfu022

Revilla, M. (2010). Quality in unimode and mixed-mode designs: a multitrait-multimethod approach. *Survey Research Methods*, *4*(3), 151–164.

Revilla, M. (2013). Measurement invariance and quality of composite scores in a face-to-face and a web survey. *Survey Research Methods*, *7*, 17–28.

Roberts, C., Eva, G., Lynn, P., & Johnson, J. (2009). *Measuring the effect of interview length on response propensity and response quality in a telephone survey. Final report of the ESS CATI experiment* (Unpublished).

Roberts, C., Eva, G., & Widdop, S. (2008). *Assessing the demand and capacity for mixing modes of data collection on the European Social Survey: final report on the mapping exercise*. London: Centre for Comparative Social Surveys.

Schuman, H., & Presser, S. (1981). *Question & answers in attitude surveys. Experiment of question form, wording, and context*. San Diego, CA: Academic Press.

Tourangeau, R., Rips, L. J., & Rasinski, K. A. (2000). *The Psychology of Survey Response*. Cambridge: Cambridge University Press.

Tourangeau, R., & Smith, T. (1996). Asking sensitive questions. The impact of data collection mode, question format and question context. *Public*, *60*(2), 275–304.

Vandenplas, C., Loosveldt, G., Beullens, K., & Vannieuwenhuyze, J. T. A. (2014). *Protocols and tools for correcting mode differences in the data* (Deliverable No. 7.6).

Vannieuwenhuyze, J. T. A., & Loosveldt, G. (2013). Evaluating relative mode effects in mixed-mode surveys: Three methods to disentangle selection and measurement effects. *Sociological Methods & Research*, *42*(1), 82–104. http://doi.org/10.1177/0049124112464868

Vannieuwenhuyze, J. T. A., Loosveldt, G., & Molenberghs, G. (2010). A method for evaluating mode effects in mixed-mode surveys. *Public Opinion Quarterly*, *74*(5), 1027–1045. http://doi.org/10.1093/poq/nfq059

Vannieuwenhuyze, J. T. A., Loosveldt, G., & Molenberghs, G. (2014). Evaluating mode effects in mixed-mode survey data using covariate adjustment models. *Journal of Official Statistics*, *30*(1), 1–21.

Villar, A., Winstone, L., Prestage, Y., & Lynn, P. (2014). *Deliverable 7.4 Feasibility Report - costs, response rates, design effects & quality of alternative data collection* (Unpublished ESS DACE Deliverable No. 7.4). London: City University London.

Appendix 1. Research teams for each study

Study 1 researchers	Roger Jowell	
	Caroline Bryson	
	Ruth O'Shea	
Study 2 researchers	Caroline Roberts	
	Annette Jäckle	
	Peter Lynn	
Study 3 researchers	Rory Fitzgerald	Gideon Skinner
	Roger Jowell	Ana Villar
	Peter Martin	Lizzy Winstone
Study 4 researchers	Gillian Eva	
	Roger Jowell	
	Caroline Roberts	
Study 5 researchers	Gillian Eva	Peter Lynn
	Dirk Heerwegh	Peter Martin
	Roger Jowell	Melanie Revilla
	Willem Saris	Geert Loosveldt
Study 6 researchers	Mare Ainsaar	Hideko Matsuo
	Rory Fitzgerald	Allison Park
	Mikael Hjerm	Yvette Prestage
	Alun Humphrey	Melanie Revilla
	Roger Jowell	Jorre Vannieuwenhuyze
	Kaur Lumiste	Ana Villar
	Peter Martin	Lizzy Winstone