Section V

Mixed Mode and Mixed Methods

18

The Design and Implementation of Mixed-mode Surveys

Edith D. de Leeuw[1], Z. Tuba Suzer-Gurtekin[2], and Joop J. Hox[1]

[1] *Department of Methodology and Statistics, Utrecht University, Utrecht, The Netherlands*
[2] *Survey Research Center, University of Michigan, Ann Arbor, MI, USA*

18.1 Introduction

18.1.1 Need for Mixed Mode

Mixed-mode surveys, that is, surveys that combine multiple modes of data collection, have become commonplace in the twenty-first century.[1] Within countries, researchers increasingly use mixed-mode designs in an attempt to reduce Total Survey Error (TSE), including coverage, nonresponse, and measurement error, and data collection costs [1–3]. Response rates have been declining over time both in Europe and the United States [4, 5]. At the same time survey costs have been increasing, for example, the per-household cost for the US decennial census increased by a factor of six in the period 1970–2010 [3]. Cost reduction has been the number one reason for European statistical agencies to use mixed-mode data collection and facilitating respondents and increasing response rate the second and third most often named reasons [6]. The single-mode paradigm, which implies that one data collection mode fits all respondents perfectly, is no longer tenable for all survey research purposes. For instance, sequential mixed-mode designs, where nonrespondents in one mode (e.g. mail or web) are followed up with a more expensive mode (e.g. interview), offer an attractive alternative because this design helps to maintain acceptable response rates at affordable costs [7]. For examples of sequential mixed-mode surveys, see Dillman et al. [8, chapter 11], de Leeuw [1], and Tourangeau [3].

1 Mixed mode refers to the use of two or more quantitative data collection methods in one survey. Mixed method refers to surveys that combine qualitative and quantitative methods.

Advances in Comparative Survey Methods: Multinational, Multiregional, and Multicultural Contexts (3MC), First Edition. Edited by Timothy P. Johnson, Beth-Ellen Pennell, Ineke A.L. Stoop, and Brita Dorer.

The twenty-first century also marked the rise of online surveys, which rapidly became one of the major modes of data collection, combining cost-effectiveness with fast and timely data [9]. However, a serious threat to the validity of online surveys is undercoverage. Although Internet coverage is high in Northern Europe (e.g. 95.5% for the Netherlands), it is lower in other countries [10]. Furthermore, the penetration of new technologies (e.g. Internet connections, smartphones, iPads, or tablets) differs between subgroups, the so-called "digital" divide within and across countries [11, 12], although there is evidence that the gap is diminishing over time both in the United States [13] and in Europe [12]. A combination of different survey modes in one study can reduce undercoverage of specific subgroups (e.g. elderly or lower educated). Just as the rise of telephone surveys in the twentieth century encouraged the development of mixed-mode strategies [14, 15], the popularity of web surveys in the beginning of the twenty-first century has pushed researchers further toward the use of mixed-mode strategies (e.g. Ref. [16]).

Finally, the growing need for international comparative data and the resulting increase of international surveys almost automatically brought mixed-mode surveys into focus. Countries have different survey traditions, and differences between countries pose challenges to survey design, including coverage and sampling and available data collection infrastructures. In practice this may result in a mixed-mode design across countries as standardization of the mode of data collection across countries may result in a suboptimal design (e.g. a poor sampling method) for some countries, which may threaten comparability [17].

In the literature on mixed-mode surveys [1, 18, 19, chapter 8], a distinction is being made between mixing modes for contact (e.g. a paper advance letter for a telephone or web survey) and mixing modes for response (e.g. following up the web nonrespondents with an interview). Multiple modes of contact (e.g. advance letter, reminder) do improve response rates [8, 20, 21, chapter 11] and are therefore used in many surveys around the world [3, 6]. When multiple modes of contact are used, but the data collection (response) is conducted in a single mode, this affects the response rate but not the measurement itself [1]. In this chapter we concentrate on multiple mode surveys to collect data, that is, on mixed mode in the response phase.

18.1.2 Mixed-mode Studies Across and Within Countries

A mixed-mode approach across countries, that is, different countries use different modes of data collection, is often inevitable due to differences in survey traditions and climate, literacy, availability of registers and sampling frames, and differences in Internet and telephone penetration. Therefore an *across-country mixed-mode* approach is regularly encountered in international comparative surveys [22]. For example, at the start of the International Social Survey Programme (ISSP), the required mode was a self-administered survey, but when more countries eventually joined in this was no longer tenable, and face-to-face

interviews were allowed for low literacy countries [23]. Pew Research Center conducts its international global attitudes and trends surveys via telephone or face-to-face interviews, depending on the country, as Gallup does in its World Poll. Sometimes changes in political climate or the onset of armed conflicts make it necessary to change data collection methods almost overnight as was the case in Iraq, where Gallup had to replace face-to-face interviews with telephone interviews from call centers in countries outside Iraq.

Although across-country mixed mode is often necessary and does have advantages in solving sampling and coverage challenges between countries, it also introduces the potential danger of differential mode measurement effects between countries. Survey methodologists and statisticians have long been aware of the need for measurement equivalence in comparative international surveys; for a general overview see Harkness et al. [24]. For special programs with a larger budget that can spend relatively large amounts on field work (e.g. the European Social Survey: ESS), it is still feasible to avoid mode effects and stay within a single-mode framework, but the high fieldwork costs sometimes force international surveys to adopt a mixed-mode approach; a good example is the Generation and Gender Program (GGP), where the limited funding for each contributing country forces individual countries to opt for a mode change.

Within-country mixed-mode designs, that is, in one country for a particular survey multiple data collection modes are used, are becoming increasingly common, and single-mode studies are now fewer than ever [25]. A prime example of a long-standing, successful within-country mixed-mode survey is the American Community Survey. In international comparative surveys the situation can become complex as different countries may not only use different data collection modes but also use different mixes of data collection modes. From a relatively simple mail-interview mix across countries, the ISSP now has developed a complex within- and across-countries mixed-mode design with many variations between countries (see Chapter 19). Another illustrative example is described by Blanke and Luiten [6], who provide an overview of data collection modes used for the Labor Force Survey (LFS) in European countries. For the first wave of the LFS, many countries use a single-mode design, but the countries differ in mode used. For instance, Poland uses paper and pencil (PAPI), Denmark uses computer-assisted telephone interviewing (CATI), and Cyprus uses computer-assisted personal interviewing (CAPI). Other countries use a mixed-mode approach, such as CATI–CAPI (e.g. UK) or computer-assisted web interview (CAWI), CATI, CAPI (the Netherlands). For later waves, the situation becomes even more complex, as some countries keep the same mode (e.g. PAPI in Poland), while other countries switch mode (e.g. the Netherlands and Cyprus to single CATI and Denmark to a CAWI–CATI mix). The most popular mixed-mode approach in the second and later waves is a CATI–CAPI combination.

Finally, data collection methods for a particular survey may change over time, leading to *across time mixed-mode* data collection, as the second wave of

the LFS above illustrates. Sometimes a mode change over time has external causes and is not the preferred choice of the researcher (e.g. Gallup's mode change in Iraq). Another example is the Netherlands Kinship Panel Survey (NKPS), which provides the Dutch data for the GGP, where due to budget cuts the data collection changed from CAPI to a sequential mixed-mode design, starting with the cost-effective web mode and using the more expensive CATI and CAPI modes for nonresponse follow-up [26].

In the examples above researchers did not have much choice and were forced to change. But mixed-mode designs and mode changes are often implemented for sound methodological reasons. Deliberate mode changes over time often occur in specific longitudinal or panel designs, where the first mode used for panel formation is the interview, preferably face-to-face, while a less expensive mode is used in later data collection periods. Typical examples in the United States are the Current Population Survey (CPS) and the National Crime Victimization Survey (NCVS). This approach is also the gold standard in building and maintaining probability-based online panels. As there is no adequate sampling frame of email addresses for the general population, a probability sample is drawn using regular sampling frames (e.g. addresses), and interviews are used to recruit members for the online panel. An early example is the Dutch LISS panel, but this procedure is now used in several countries [27]. Similar procedures are used in the European LFS where the first wave uses more intensive, higher cost procedures to ensure good response rates [6]. For an overview and typology of mixed-mode surveys and their implications; see de Leeuw [1], de Leeuw et al. [2], and Dillman et al. [19, p. 307].

18.2 Consequences of Mixed-mode Design

18.2.1 Trade-off Between Costs and Errors

In high quality survey research, be it within one country or across countries, the goal is to minimize the TSE within reasonable time and at affordable cost [17]. Using a mixed-mode design implies a careful trade-off among coverage, nonresponse, measurement errors, and cost. How well do mixed-mode designs perform? Empirical evidence is still limited to within-countries mixed-mode designs, but does show promise. A review is provided below.

18.2.2 Reducing Coverage and Nonresponse Error

The risk of undercoverage and the resulting *coverage error* is one of the main reasons to combine online surveys with a second mode. The type and implementation of the mixed-mode design is critical. For instance, a sequential web–paper mail survey raised both response and improved coverage for the

general population in the United States [28]; a concurrent mixed-mode approach, where a mode choice was offered, did, however, not show these advantages [29]. In general, offering respondents the choice between mail and web in a concurrent mixed-mode survey does not improve response. A meta-analysis, summarizing 19 experimental comparisons between a concurrent mail–web choice and single-mode mail surveys, showed that giving respondents a choice lowered response rates with an average of 3.8% points [30]. Although giving a choice seems respondent friendly, it also puts a higher burden on the respondent and has negative consequences for the decision process, causing some potential respondents to postpone their reaction and finally to do nothing. However, concurrent mixed-mode surveys can be effectively used either by providing a sample member with their preferred mode based on answers in an earlier survey [31] or by targeting. The latter approach was used in a concurrent mixed-mode design, where high response propensity respondents were invited to a web survey and low propensity respondents to a mail survey. All nonrespondents were followed up with CATI; the researchers were able to maintain the previously achieved single-mode response rate and showed that the mixed-mode approach resulted in a better representativeness on key variables [32]. Furthermore, sequential mixed-mode surveys, in which nonrespondents are followed up with a more expensive method, do improve response rates. The American Community Survey is a prime example [33]. The sequential approach goes back to the seminal work by Hochstim [34] and Siemiatycki [35], who showed that data collection strategies, beginning with the least expensive modes (e.g. mail) and following up with more expensive interview modes, provided response rates as high as single-mode face-to-face interviews for only half the costs. Sequential mixed-mode surveys do increase response both for the general population [36] and for different racial and ethnic groups [37, 38].

A study among Dutch immigrants [38] showed that although a sequential mixed-mode (CAWI, CATI, CAPI) approach did raise the response rate compared with a single-mode CAPI survey, the single-mode CAPI still resulted in the best representation of immigrants on sociodemographic variables. It also showed that different groups of respondents tend to prefer different modes (younger and second-generation ethnic minorities participate more in web, older female respondents more in CATI, and in general older and first-generation immigrants more in CAPI/CATI). In Germany, a sequential follow-up of an online survey with a mailed paper survey resulted not only in a higher response rate but also in a better representation [39], confirming findings from the United States [28]. A slightly different outcome was reported in a Dutch study [40] that investigated three sequential mixed-mode strategies, CAWI–CAPI, CATI–CAPI, and paper mail–CAPI. For both the mailed paper survey and the telephone (CATI) survey, the follow-up face-to-face mode increased the representativeness on sociodemographic variables compared with the

single-mode face-to-face survey. However, the representativeness of the single-mode web was already at the level of single-mode CAPI and could not be increased any further by a CAPI follow-up. The discrepancy in results regarding CAWI surveys between the Netherlands versus the United States and Germany could be explained by the extremely high Internet penetration in the Netherlands (95.5%) compared with Germany (88.4%) and the United States (87.9%) [41]. Finally, a recent meta-analysis [42] reports that mixed-mode surveys have higher degrees of representativeness than single-mode surveys.

18.2.3 Reducing Cost

Costs are an important consideration for introducing mixed-mode data collection, and this was one of the most important reasons mentioned by the European National Statistical Institutes (NSI) for mixed-mode surveys [6]. Most countries indicated that using a mixed-mode strategy indeed decreased fieldwork costs, but some reported that costs remained about the same or even became higher. Of course, much depend on the initial single-mode data collection and on the specific mix of modes used. Face-to-face is the most expensive mode, about 5–10 times higher than telephone, and telephone is about 2–5 times more expensive than paper mail, all with the same number of respondents. Paper mail surveys are usually more expensive than web surveys because of higher processing costs, but this depends on the sample size, because mail expenses increase with the number of respondents (flexible costs vs. fixed development costs). Finally, the cost of mixed-mode surveys is typically a weighted average of the data collection costs and the developmental costs for each mode [43]. This is well illustrated by the following European examples from NSI [6]. In the Netherlands, fieldwork costs decreased by more than 50% when the data collection for the LFS changed from the very expensive CAPI single mode to a sequential CAWI–CATI–CAPI mode. However, in Italy where an inexpensive paper questionnaire was replaced by a CATI/CAPI mixed-mode design, the costs increased. Furthermore, the order in which modes are sequenced affects the total cost estimate. For instance, a paper mail survey followed by a CAWI survey hardly raises the response and therefore does not justify the costs of implementing a CAWI survey, while a CAWI first followed by a mail survey sequence is far more efficient. The latter not only increases response rates considerably but also brings in different types of respondents [8, pp. 429–432].

18.2.4 Measurement Error

Mixed-mode surveys have advantages in reducing coverage error and nonresponse error at affordable cost. However, a combination of different survey modes in one study may lead to different kinds of measurement errors due to mode effects [3, 18, chapter 19]. It is important to note that mode effects have two major components (i): mode (self) selection effects (i.e. different modes are

associated with different noncoverage and nonresponse errors) and (ii) mode measurement effects (i.e. different modes produce different observation errors). Therefore, a mode effect is the net effect of the nonobservation and measurement error differences by mode [44]. In general, mode selection effects, where different types of respondents self-select in different modes, are desirable effects when studying a single population, as they have the potential to reduce nonresponse and coverage error. If, for instance, in a within-country sequential mixed-mode design, no mode selection effects occur meaning that the same type of respondents reacts to the different modes, researchers may as well stay with the least expensive mode (e.g. web) and do not have to implement a more costly and logistically more complicated sequential mixed-mode design (e.g. CAWI–CATI–CAPI), as the latter would then just bring in more of the same respondents. Of course, when different modes are used in different countries in an across-country mixed-mode design, mode selection effects may result in differential errors of nonobservation between the countries. However this is a situation 3MC researchers are already well aware of, and it is usually accommodated by weighting (e.g. in nonresponse adjustment); see Chapter 19.

Potential mode measurement effects are more cumbersome as they threaten the comparability of the data and the validity of comparisons between groups or countries. Survey modes differ on several dimensions, for instance, interviewer-administered versus self-administered questionnaires and information transmission and communication (e.g. aurally vs. visually). For an overview and discussion, see de Leeuw [1] and Couper [45]. To understand mode measurement effects, it is very useful to distinguish (i) mode-inherent factors, (ii) context-specific factors, and (iii) implementation-specific factors [18]. Mode-inherent factors are given and cannot be negated by survey design (e.g. interviewer involvement, such as absence of interviewers in self-administered modes). The long history of studies into mode effects indeed shows a dichotomy between modes with and without interviewers, where self-administered forms, be it on paper or online, perform better with sensitive questions. CAWI and mailed paper surveys result in less social desirability and more openness in answers than face-to-face and telephone interviews. For meta-analyses and overviews, see de Leeuw [46], de Leeuw and Hox [47], and Tourangeau et al. [44].

Context-specific characteristics depend on the social and cultural aspects of the population for which the survey is implemented. Familiarity with the medium used for surveying and its associated expectations play an important role here. Context-specific characteristics may change over time and differ between countries, and contrary to mode-inherent factors, context-specific characteristics can be countered by clever survey design. For example, in international surveys it is advocated to add context-specific (meta) information to the core questionnaire before translation into multiple languages, as differences in sociocultural context affect how a respondent perceives the meaning of a question [48]; see Harkness [49, p. 63] for examples.

Implementation-specific characteristics depend on the way a mode is implemented in a specific survey. These are characteristics that a researcher can control and exploit to achieve an optimal survey design. A prime example is questionnaire design; the way a questionnaire is designed and implemented may differ over modes and in single-mode surveys, because of specific mode traditions on how questions are structured and presented; for overviews, see Dillman and Edwards [25] and Dillman and Christian [50].

When different question formats for the same question are used in different modes, this has the consequence that respondents are presented with a different stimulus in each mode, which may lead to unwanted question wording and mode effects. To avoid this Dillman [51] proposed a unified (uni-) mode design, in which the same or closely similar question structures, question wordings (including response options), and visual design are integrated [25].

18.3 Designing for Mixed Mode

18.3.1 Mode Effect or Design Effect

To minimize potential mode measurement effects, researchers should not focus on one mode only, but from the onset design and implement the questionnaires from a mixed-mode perspective. Mode-inherent factors (e.g. interviewer presence vs. absence) cannot be changed, but a clever design will reduce both context-specific and implementation-specific mode measurement effects. This idea is not totally new to researchers in the field of cross-national and cross-cultural studies, who have a long tradition of research in questionnaire design and measurement equivalence, for example, Harkness et al. [24, 52].

18.3.2 Unified Mode Design

Dillman's unified (uni-) mode design is now current best practice in mixed-mode design. The basic principle is writing and presenting questions in the same or closely similar way to ensure that respondents in all modes receive a common mental stimulus. This includes using equivalent question structure and wording across modes. This is easier for some mixed-mode designs than for others. For example, web and paper mail surveys are both self-administered, and both use visual channels of communication; for a good example of a unified mode design to achieve equivalent measurements in a mail–web survey, see the Lewiston and Clarkston Quality of Life Survey [25].

In the above description of unified mode design, the words "the same or closely similar" were used. A unified mode design does not imply that the same

text is rigidly implemented in each mode; sometimes one should deviate for a good reason. The aim of unified mode design is to present the same perceived stimulus in each mode since using the same wording rigidly may lead to unnatural situations. Some simple examples are the following: in order for instructions to make sense, these must be adapted to the mode used, for instance, "click next to continue" is not necessary in an interview situation, while a text stating "I will now read out a list" is unnecessary online as the respondents already can see the list. Furthermore, in computer-assisted data collection methods, one should use the intelligence of the system to avoid mistakes (e.g. in routing) and thereby enhancing the data quality, even if these checks are not possible for paper versions of the questionnaire.

18.3.2.1 Unified Mode Design for Interview Self-administered Mixes

Interviews and self-administered questionnaires are more disparate than online and paper mail questionnaires are. The latter two are both self-administered, visual, may use graphical language, and are self-paced. However, they lack interviewer assistance and interviewer observations. Furthermore, interviews are sequential in nature, and questions are posed one after another, while in self-administered forms respondents can see more than one question at a time. These differences have led questionnaire designers to implement different questionnaires for these modes, leading to unwanted and unnecessary mode measurement effects in mixed-mode studies. A prime example is the usage of "do not know." In interview surveys a "do not know" answer in general is not offered explicitly, but the interviewer is generally instructed to accept it after a friendly interviewer probe. In web surveys, initially most questions were programmed to be mandatory, and each time a respondent tried to skip a question, a red error message appeared, stating that one must answer. This had the disadvantage that it could irritate respondents and lead to breakoffs [11]. To avoid this, web designers now tend to allow skips or include "do not know" options with questions; the latter sometimes is visually separated from the substantive response categories. Understandably, this easy way out for the respondents leads to a higher percentage of missing data in web than in interview surveys. However, innovative design can use the intelligence of the web to emulate interviewer probes. Wine et al. [53] pioneered the implementation of a friendly interviewer-type probe, based on the actual words spoken by telephone interviewers in a CAWI survey. De Leeuw et al. [54] experimentally investigated the effectiveness of this approach. Not offering "do not know" explicitly, but allowing respondents to skip questions, resulted in the lowest amount of missing information in both telephone and CAWI surveys. Following up with a polite probe when skips occurred further reduced the number of missing items in both modes (telephone and online), without influencing the respondents' evaluations of the survey negatively.

A major difference between (face-to-face and telephone) interviews and CAWI surveys is that in interviews the questions are posed sequentially one by one, while in online questionnaires they are often presented together in a grid (matrix) format. Grids have advantages, such as no need to press "next" resulting in lower respondent burden, but grids also have disadvantages, such as evoking satisficing behavior and the danger of triggering context effects [11]. Furthermore, as more and more respondents use their smartphones to go online, grid questions pose an extra challenge [55]. Presenting questions one by one on the screen would avoid many problems but has the disadvantage of increasing the respondent burden and the overall interview duration. A potential solution is the auto-advance carrousel. The format of an auto-advance carrousel question is (i) question body, (ii) response categories, and (iii) navigation/overview bar. When an answer is selected by the respondent, after a couple of seconds the answer is cleared, and the wording of the next question appears. The wording with instructions and the list of response categories stay in place, and the navigation bar provides information where the respondent is in the series of questions; in addition, by clicking on the navigation bar, the respondent can navigate and go back when desired (for an example see Figure 18.1). Roberts et al. [56] investigated the efficacy of the auto-advance format in an online panel and experimentally compared it with standard grid formats of various lengths. The auto-advance format resulted in fewer break-offs and better data quality (e.g. less straightlining and other response tendencies). Respondents also evaluated the auto-advance questions as easier to complete, more enjoyable, less irritating, and less strenuous than the standard grid. In a mode comparison in Slovenia, the auto-advance format also performed well and reduced mode measurement differences [18].

A final example, illustrating the importance and usefulness of graphical language can be found in Christian et al. [57]. They describe how the use of

Figure 18.1 Example carrousel question in Dutch. First question in a series of five (see navigation bar). Previous and Next buttons are disabled (gray). Seven-point response scale (totally agree – totally disagree).

symbols and adapted box sizes improves the quality of responses to a CAWI question on important dates to the level of interviewer-administered data. In a standard telephone interview, the question, "When did you start attending Washington State University?" led to only 13% of the respondents reporting the desired month and year. Instead, most respondents gave comments like "last spring semester," "Fall 2002," or "this is my first semester." As in any good telephone interview, these responses were followed up by the interviewers to get the desired response format. In the web survey, a standard write-in format was used (i.e. Date Started [ ] / [ ] (MM/YYYY), as two digits for month and four digits for year were requested). In the initial web survey only 45% of the respondents answered correctly the question of what month and year they started school, using this standard format. As web surveys are self-administered, no interviewer assistance or follow-up is available, and all help and information should be communicated using the visual channel. In a series of experiments, visual design manipulations were tested to improve the accuracy of the answers through the web. By decreasing the size of the month box relative to the year box (thereby graphically indicating that month should be typed in fewer digits than year), using more precise language of symbols (not a combination of MM/YYYY, but separating these and separating the boxes), and placing symbols in natural reading order ahead of the appropriate response boxes (i.e. MM[ ] YYYY[ ]), the percentage of people responding in the desired format increased from 45 to 96%. These results clearly illustrate how different approaches to the question (telephone and web survey) can lead to the same result but through different mechanisms. In the telephone survey, the interviewer served as an intelligent system that could easily convert the answer to the desired format and if necessary ask for more information. In the web survey, the emphasis was on graphical language, size of answer space, and labeling in order to engage respondents to answer in the desired format and avoid error messages.

18.3.3 Lessons Learned

When comparing groups or countries, measurement equivalence is of the utmost importance. Mode measurement effects (i.e. mode-inherent, context-specific, and implementation-specific effects) threaten this equivalence. Mode-inherent effects cannot be counteracted by clever design. An example is that answers to self-administered forms (CAWI and paper) are more honest and open when sensitive questions are asked. The only way to counteract mode effects in mixed-mode (interview-online mix or interview-paper mail mix) surveys with sensitive questions is to incorporate for all respondents in face-to-face and telephone interviews a specific section for the sensitive questions in a self-administered format (e.g. hand over a paper questionnaire, switch to ACASI or interactive voice response (IVR)). Context-specific and implementation-specific

mode measurement effects can be negated by good design, and researchers should pursue equivalence both in questionnaire design and implementation. Although this is far from easy, a unified mode design helps in accomplishing this goal. It is important to realize that unified mode design is *not* designing down to the lowest common denominator; for instance, when mixing web and paper, one should always use web technology to help online respondents navigate and avoid mistakes. Unified mode design does require an excellent understanding of the single modes in the mix, of mode differences, graphical design, and of implementation procedures in all modes, as the above examples show.

Also, whenever one is designing a mixed-mode survey, one should ideally also collect auxiliary information. From nonresponse studies [58, 59], we know the importance of auxiliary information for the investigation of nonresponse bias and adjustment. Also, for mixed-mode studies where mode selection effects and mode measurement effects are confounded, we need auxiliary data. This is needed in the analysis phase, where we first have to estimate the mode selection and mode measurement effects and then adjust. Estimating and separating mode selection and mode measurement effects is essential for adjustment in a later phase [3, 18, 60].

18.4 Auxiliary Data for Assessing and Adjusting Mode Effects

If different modes in a mixed-mode design result in different observed responses, the observed difference between the survey modes is the total effect of selection and measurement, and as a consequence the substantive differences in response and measurement are confounded (see Figure 18.2). An example: in a health survey, a web–telephone mix is used to investigate risk behavior such as drinking alcohol. We do know that respondents to web surveys are in general younger; we also know that self-administered modes such as CAWI result in more openness and less social desirability. If web respondents report more drinking than telephone respondents, we do not know if this is a real difference due to self-selection of younger respondents in the CAWI mode and a relationship between drinking and age or if it is caused by a mode measurement effect due to lesser social desirability and more openness in the web. Therefore in the analysis phase, before adjusting for the (unwanted) mode measurement effects, statisticians have to differentiate between mode selection and mode measurement effects. This requires additional (auxiliary) data. There are several approaches to estimating and adjusting for mode effects, and each approach uses different types of auxiliary information; cf. Hox et al. [60, chapter 19]. In the design phase of the study, it is necessary to plan the desired analyses and specify the auxiliary information needed.

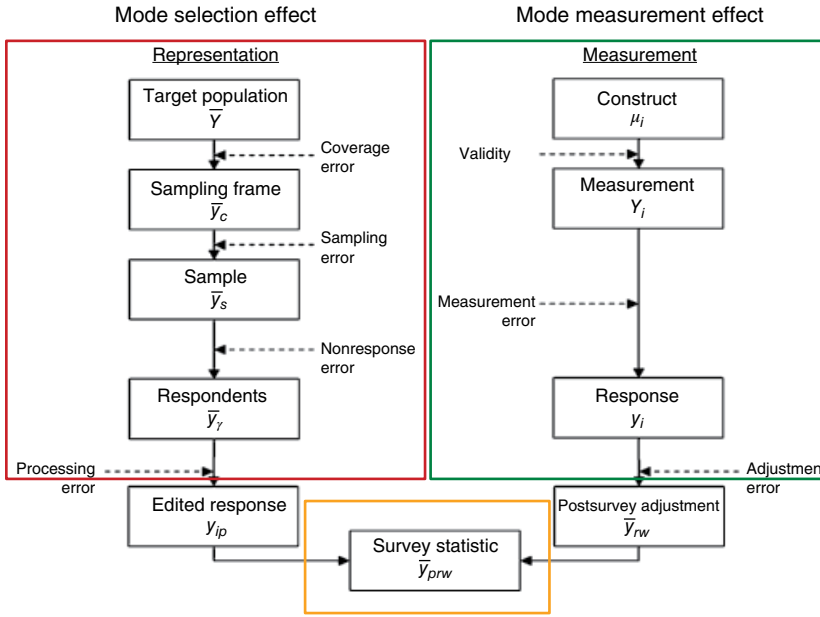

Figure 18.2 Confounding of wanted mode selection and unwanted mode measurement effects in estimating the survey statistic of interest. Adapted and extended Figure 2.2 [61].

Two general strategies can be applied: the first uses available data, the second explicitly collects additional data for an optimal adjustment.

18.4.1 Use Available Data for Estimating and Adjusting for Mode Effects

18.4.1.1 Demographic Variables

It is not always necessary to explicitly collect new data, and instead available data may be used. In rare cases validating information on important variables is available, for instance, from registers and record checks. This can then subsequently be used as a benchmark. A more general approach is the use of demographic variables. Under the mode insensitivity assumption, that is, the assumption that the demographic variables are unaffected by mode measurement effects, these can be used to separate selection and measurement effects [62]. A second important assumption is that the auxiliary demographic variables are sufficient to explain all selection effects.

This approach is well known from the literature on nonresponse adjustment [59, 63]. Provided that these two strong and somewhat unrealistic assumptions hold, controlling for the demographic variables eliminates selection effects. For instance, by using the demographic variables as predictors to estimate the

propensity of a respondent to (self) select a specific mode, conditional on this propensity, the remaining mode effects are assumed to be measurement effects. However, in general demographic variables are not very strong predictors of mode self-selection and thus underestimate selection bias [60, 62].

18.4.1.2 Reference Survey

Existing single-mode data from a comparative sample, a so-called reference survey, may be used to disentangle selection and measurement effects. This has the advantage that usually more variables are available to use in the analyses. The reference survey should be a single-mode survey, and again two important assumptions should be met. The first one is that the measurement error in the single-mode reference survey and in the comparable mode in the mixed-mode survey are equal (measurement equivalence). The second assumption is that the single-mode sample and the mixed-mode sample represent the same population. For an example with data from the ESS, see Vannieuwenhuyze et al. [64]. However, in every day survey practice, it may be hard to find an existing reference survey that meets all assumptions, and it is better to collect new data.

18.4.1.3 Longitudinal Data as Reference

A special case is longitudinal surveys, where single-mode data from previous measurements are used as reference data; for an example, see Cernat [65]. In longitudinal surveys, researchers often do not wish to disturb the comparability with previous time points (e.g. time series), and when a mode switch in later waves becomes necessary or a mixed-mode design is planned for the later waves, the single mode of the earlier waves is then used as benchmark. In this design, strong auxiliary information is available from the single-mode reference wave, since in addition to demographic data, also information on all variables of interest is available from a single-mode source. Since the earlier wave uses only one mode, they are free from mode effects and therefore can be used as benchmark data. This procedure is explained and illustrated using data from the NKPS; see Hox et al. [26]. The first two waves of the NKPS were administered face-to-face. In the third wave, the data were collected using a sequential mixed-mode (CAWI, CATI, CAPI) design. For most variables of interest in the NKPS, measurement equivalence holds over modes, and mode measurement effects tend to become smaller if a correction is applied based on information from the previous single-mode data collection occasions.

18.4.2 Collect Additional Data for Estimating and Adjusting for Mode Effects

18.4.2.1 Reference Survey

Existing auxiliary information may be problematic, and assumptions as described above may not be met. For instance, the reference survey may have a different research goal, variables may be operationalized differently, the

definition of the population of interest may be different, or the data are simply too old. Sometimes researchers have the opportunity and funds to collect additional data. This should be planned carefully, and a form of experimental design is necessary. A simple design is a randomized between persons design, where a random subsample of the population is approached with a mixed-mode survey and a second random subsample gets a single-mode survey that serves as reference survey; this is the Vannieuwenhuyze et al. [64] approach. The sizes of the two random subsamples (i.e. the mixed-mode survey and the single-mode reference survey) need not to be the same, and the data from both subsamples can be used for the substantive analysis. A similar approach is used in the analysis of MultiTrait–MultiMethod experiments that were conducted in several countries of the ESS; in this case the regular ESS surveys were used as reference surveys [66].

18.4.2.2 Between Persons Experiments

De Leeuw [1] proposed to embed a randomized mode experiment for a subsample as part of the survey design. Even if for practical reasons a random subsample of the population is not feasible for all countries, a limited experiment can be designed where comparable subsamples in each of the countries (e.g. urban population) are studied to assess mode effects. This approach was used in the context of the ESS, where between subject mode experiments were carried out in the capitals of Hungary and Portugal. Three random groups (face-to-face interviews with show cards, face-to-face without show cards, and telephone interviews) were compared showing that use of show cards did not influence data quality neither positively nor negatively [67, 68].

18.4.2.3 Reinterview

Another design useful to investigate mode effects is the within subject design, which is a form of repeated measurement. An additional data collection phase is planned in the design of the mixed-mode survey, and in this sense the within subject design differs from using existing longitudinal data (e.g. NKPS). Klausch [69] proposes to use a single-mode follow-up of a random subsample of the respondents in a mixed-mode approach. As the same questionnaire and fieldwork procedures are used, this reinterview can act as a benchmark to separate selection and measurement effects. For a description and illustration of this method, see Klausch et al. [40]. A reinterview method was also used in the testing phase of a mixed-mode experiment in Hungary in the context of the ESS [68, pp. 6–7].

18.4.3 Choosing Benchmark Data

Whether one is using an existing survey as reference survey or whether one is collecting new data to serve as a benchmark, an important decision is which mode should be used as the gold standard and how it should be implemented.

The population (e.g. age range) should be the same as in the mixed-mode study, and the questions and response categories should be in a comparable format. In many cases a face-to-face interview is seen as the ideal mode and therefore used as reference survey. For instance, in the ESS the prescribed mode is face-to-face, and for mixed-mode studies in the ESS, face-to-face is the obvious choice as benchmark. In the reinterview method by Klausch [69] also a face-to-face interview was used, because traditionally this was the standard method at Statistics Netherlands. In those cases the mixed mode was also designed and implemented in such a way that it comes closest to the face-to-face interview it potentially may replace.

In some cases, face-to-face is not the obvious choice as benchmark. For instance, when sensitive data are collected a self-administered form of data collection is a far better choice as benchmark. Another example is the ISSP, which started out as a self-administered survey. Finally, in longitudinal studies it makes sense to use data from an earlier wave, as was done in the NKPS example [26] or preferably collect new benchmark data in the preferred mode.

18.5 Conclusions

In an optimal design for mixed-mode surveys, be it within a country or across countries, there are three phases that need attention. The first phase is the design phase: it is important to prevent mode measurement effects when designing the study. Some differences between modes are unavoidable; mode-inherent factors such as the presence or absence of interviewers cannot be changed. But other differences may be the result of differential implementation details; for example, explicitly offering a "do not know" category in one mode and not in the other may have severe consequences for data integrity as we discussed above. These design options are in principle under the control of the researchers and can be managed to counteract mode effects, for instance, through a unified mode design.

In cross-national surveys, achieving measurement invariance is of paramount importance, and much attention has been paid to designing equivalent questionnaires and striving for measurement invariance across countries [24, 52] (see also the cross-cultural survey guidelines at ccsg.isr.umich.edu). When different modes are added to a design that already includes different countries and languages, measurement problems are likely to increase, and rules for equivalent questionnaire design are of the utmost importance. In cross-national research much effort is already going into the translation of questionnaires; analogously one can argue that each mode has its own language and that one should pay attention to the peculiarities of each mode language.

The use of well-established and validated constructs and multiquestion scales across countries and modes helps achieving measurement equivalence. What is minimally needed for valid comparisons across countries and modes is partial measurement invariance: at least two items (questions) per scale (construct) are invariant (see also Chapters 40 and 41). In the design phase, one option is to add some extra questions per scale, at the expense of having fewer scales (variables/constructs) available for substantive analysis. In survey design, the tendency is to use short scales in order to maximize the number of concepts that can be measured. When measurement equivalence is an issue, this strategy may backfire, because there is no room to drop problematic questions in the analysis phase. A promising new approach to measurement invariance testing is to allow some "wriggle room" for a small amount of measurement variance; this is called approximate invariance. The various ways this can be accomplished are discussed in a special issue of the journal *Frontiers in Psychology* [70].

After the data are collected, the next two phases are (i) estimating the potential mode measurement effect due to different modes and (ii) adjusting for any such bias. During analysis one then has to estimate both mode selection effects and mode measurement effects. To differentiate between these two effects, statisticians need additional information. In survey practice, often only demographic data are available, and often the assumption must be made that demographic information collected in a mixed-mode study is not sensitive to mode effects and that they are useful covariates. However, demographic data are in general weak predictors. Therefore, a richer source of additional data will help to achieve a better estimation of mode effects. These auxiliary data can come from a reference survey, a reinterview, or an embedded randomized mode experiment. The same data are also necessary for adjustment techniques. This feeds back into the design phase; when survey researchers design a mixed-mode international survey, they must design the data collection in such a way that the necessary additional information will be available for later analysis and adjustment.

Acknowledgment

The authors thank Ana Villar and Don Dillman for their helpful and insightful comments.

References

1 de Leeuw, E.D. (2005). To mix or not to mix. Data collection modes in surveys. *Journal of Official Statistics* 21 (2): 233–255.

2 de Leeuw, E.D., Dillman, D.A., and Hox, J.J. (2008). Mixed mode surveys: when and why. In: *International Handbook of Survey Methodology*, EAM Series (ed. E.D. de Leeuw, J.J. Hox and D.A. Dillman), 299–316. New York: Taylor & Francis/Psychology Press.
3 Tourangeau, R. (2017). Mixing modes: tradeoffs among coverage, nonresponse, and measurement error. In: *Total Survey Error in Practice* (ed. P. Biemer, E. de Leeuw, S. Eckman, et al.), 115–132. Hoboken, NJ: Wiley.
4 Brick, J.M. and Williams, D. (2013). Explaining rising nonresponse rates in cross-sectional surveys. *Annals of the American Academy of Political and Social Science* 645: 36–59.
5 de Leeuw, E.D. and de Heer, W. (2002). Trends in household survey nonresponse: a longitudinal and international comparison. In: Survey Nonresponse (ed. R.M. Groves, D.A. Dillman, J.L. Eltinge and R.J.A. Little), 41–54. New York: Wiley.
6 Blanke, K., and Luiten, A. (2014). Query on data collection for social surveys. ESSnet project: data collection for social surveys using multiple modes. http://ec.europa.eu/eurostat/cros/system/files/Query_report_DCSS.pdf_en (accessed August 2016).
7 Biemer, P.P. and Lyberg, L.E. (2003). *Introduction to Survey Quality*. Hoboken, NJ: Wiley.
8 Dillman, D.A., Smyth, J., and Christian, L.M. (2014). *Internet, Phone, Mail, and Mixed-Mode Surveys*. Hoboken, NJ: Wiley.
9 Baker, R., Blumberg, S.J., Brick, J.M. et al. (2010). AAPOR report on online panels. *Public Opinion Quarterly* 74 (4): 711–781.
10 Internet World Stats (2017a). Internet Stats and Facebook Usage in Europe. http://www.internetworldstats.com/stats4.htm#europe (accessed June 2017).
11 Couper, M.P. (2008). *Designing Effective Web Surveys*. New York: Cambridge University Press.
12 Mohorko, A., de Leeuw, E., and Hox, J. (2013). Internet coverage and coverage bias in Europe: developments across countries and over time. *Journal of Official Statistics* 29 (4): 1–15.
13 Sterrett, D., Malato, D., Benz, J. et al. (2017). Assessing changes in coverage bias of web surveys in the United States. *Public Opinion Quarterly* 81 (special issue): 338–356.
14 Dillman, D.A. and Tarnai, J. (1988). Administrative issues in mixed mode surveys. In: *Telephone Survey Methodology* (ed. R.M. Groves, P.P. Biemer, L.E. Lyberg, et al.), 509–528. New York: Wiley.
15 Groves, R.M. and Lepkowski, J.M. (1985). Dual frame, mixed mode survey designs. *Journal of Official Statistics* 1 (3): 263–286.
16 Blyth, B. (2008). Mixed mode: the only "fitness" regimen? *International Journal of Market Research* 50 (2): 241–266.
17 Pennell, B.-E., Cibelli-Hibben, K., Lyberg, L. et al. (2016). A total survey error perspective on surveys in multinational, multiregional, and multicultural

contexts. In: *Total Survey Error in Practice* (ed. P. Biemer, E. de Leeuw, S. Eckman, et al.), 179–201. Hoboken, NJ: Wiley.

18 de Leeuw, E.D. and Berzelak, N. (2016). Survey mode or survey modes? In: *The SAGE Handbook of Survey Methodology* (ed. C. Wolf, D. Joye, T.W. Smith and Y.-c. Fu), 142–156. London: Sage.

19 Dillman, D.A., Smyth, J., and Christian, L.M. (2008). *Internet, Mail, and Mixed-Mode Surveys*. Hoboken, NJ: Wiley.

20 Heberlein, T.A. and Baumgartner, R. (1978). Factors affecting response rates to mailed questionnaires: a quantitative analysis of the published literature. *American Sociological Review* 43 (4): 447–462.

21 de Leeuw, E.D., Callegaro, M., Hox, J. et al. (2007). The influence of advance letters on response in telephone surveys: a meta-analysis. *Public Opinion Quarterly* 71 (3): 413–443.

22 Martin, P. (2011). A good mix? Mixed mode data collection and cross-national surveys. *ASK, Research and Methods* 20 (1): 5–26.

23 Skjak, K.K. and Harkness, J. (2003). Data collection methods. In: *Cross-Cultural Surveys Methods* (ed. J.A. Harkness, F.J.R. Van de Vijver and P.P. Mohler), 179–194. Hoboken, NJ: Wiley.

24 Harkness, J.A., Braun, M., Edwards, B. et al. ed. (2010). *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*. Hoboken, NJ: Wiley.

25 Dillman, D.A. and Edwards, M.L. (2016). Designing a mixed-mode survey. In: *The SAGE Handbook of Survey Methodology* (ed. C. Wolf, D. Joye, T.W. Smith and Y-c. Fu), 255–268. London: Sage.

26 Hox, J.J., de Leeuw, E.D., and Zijlmans, E.A. (2015). Measurement equivalence in mixed mode surveys. *Frontiers in Psychology* 6 (87): doi: 10.3389/fpsyg.2015.00087.

27 Blom, A.G., Bosjnak, M., Cornilleau, A. et al. (2016). A comparison of four probability based online and mixed mode panels in Europe. *Social Science Computer Review* 34 (1): 8–25.

28 Messer, B.J. and Dillman, D.A. (2011). Surveying the general public over the internet using address-based sampling and mail contact procedures. *Public Opinion Quarterly* 75 (3): 429–457.

29 Millar, M.M. and Dillman, D.A. (2011). Improving response to web and mixed-mode surveys. *Public Opinion Quarterly* 75 (2): 249–269.

30 Medway, R.L. and Fulton, J. (2012). When more gets you less: a meta-analysis of the effect of concurrent web options on mail survey response rates. *Public Opinion Quarterly* 76 (4): 733–746.

31 Olson, K., Smyth, J.D., and Wood, H.M. (2012). Does giving people their preferred survey mode actually increase survey participation rates? An experimental examination. *Public Opinion Quarterly* 76 (4): 611–635.

32 Luiten, A. and Schouten, B. (2013). Tailored fieldwork design to increase representative household survey response: an experiment in the survey of consumer satisfaction. *Journal of the Royal Statistical Society: Series A* 176 (1): 169–189.

33 United States Census (2016). Design and methodology report. http://www.census.gov/programs-surveys/acs/methodology/design-and-methodology.html (accessed September 2016).
34 Hochstim, J.R. (1967). A critical comparison of three strategies of collecting data from households. *Journal of the American Statistical Association* 62: 976–989.
35 Siemiatycki, J. (1979). A comparison of mail, telephone, and home interview strategies for household health surveys. *American Journal of Public Health* 69: 238–245.
36 Fowler, F.J., Gallagher, P.M., Stringfellow, V.L. et al. (2002). Using telephone interviews to reduce nonresponse bias to mail surveys of health plan members. *Medical Care* 40: 190–200.
37 Beebe, T.J., Davern, M.E., McAlpine, D.D. et al. (2005). Increasing response rates in a survey of Medicaid enrollees: the effect of a prepaid monetary incentive and mixed modes (mail and telephone). *Medical Care* 43: 411–414.
38 Kappelhof, J. (2015). The impact of face-to-face vs. sequential mixed-mode designs on the possibility of nonresponse bias in surveys among non-western minorities in the Netherlands. *Journal of Official Statistics* 31 (1): 1–31.
39 Bandilla, W., Couper, M. P., and Kaczmirek, L. (2014). The effectiveness of mailed invitations for web surveys and the representativeness of mixed-mode versus Internet only samples. *Survey Practice*, 7(4). http://www.surveypractice.org/index.php/SurveyPractice/article/view/274/html_11 (accessed August 2016).
40 Klausch, T., Hox, J., and Schouten, B. (2015). Selection error in single and mixed mode surveys of the Dutch general population. *Journal of the Royal Statistical Society: Series A* 178 (4): 945–961.
41 Internet World Stats (2017b). Internet World Stats. http://www.internetworldstats.com/ (accessed June 2017).
42 Cornesse, C. and Bosnjak, M. (2016). The association between survey characteristics and representativeness: a meta-analysis of two common approaches. Paper presented at the International Workshop on Household Survey Nonresponse, Oslo (August).
43 Brick, J.M. and Kalton, G. (2010). Survey mode choices: data quality and costs. *Proceedings of the Ninth Conference on Health Survey Research Methods*, Hyattsville, MD: US National Center for Health Statistics (May 2010).
44 Tourangeau, R., Conrad, F.R., and Couper, M.P. (2013). *The Science of Web Surveys*. New York: Oxford University Press.
45 Couper, M.P. (2011). The future of modes of data collection. *Public Opinion Quarterly* 75 (5): 889–908.
46 de Leeuw, E.D. (1992). *Data Quality in Mail, Telephone and Face to Face Surveys* (Chapter 2, pp. 13–20; Chapter 3, pp. 21–33; Appendix A, pp. 127–134). Amsterdam: TT Press.
47 de Leeuw, E.D. and Hox, J.J. (2011). Internet surveys as part of a mixed mode design. In: *Social and Behavioral Research and the Internet: Advances in*

Applied Methods and Research Strategies, EAM Series (ed. M. Das, P. Ester and L. Kaczmirek), 45–76. New York: Routledge/Taylor & Francis.

48 Braun, M. and Harkness, J.A. (2005). Text and context: challenges to comparability in survey questions. In: *Methodological Aspects in Cross-National Research* (ed. J.H.P. Hoffmeyer-Zlotnik and J.A. Harkness), 95–107. Mannheim, Germany: ZUMA.

49 Harkness, J.A. (2008). Comparative survey research: goals and challenges. In: *International Handbook of Survey Methodology*, EAM Series (ed. E.D. de Leeuw, J.J. Hox and D.A. Dillman), 56–77. New York: Taylor & Francis/Psychology Press.

50 Dillman, D.A. and Christian, L.M. (2005). Survey mode as a source of instability across surveys. *Field Methods* 17: 30–52.

51 Dillman, D.A. (2000). Mail and Internet Surveys. New York: Wiley.

52 Harkness, J.A., van de Vijver, F.J.R., and Mohler, P. Ph., eds. (2003). *Cross-Cultural Survey Methods*. Hoboken, NJ: Wiley.

53 Wine, J.S., Cominole, M.B., Heuer, R.E., and Riccobono, J.A. (2006). Challenges of designing and implementing multimode instruments. Paper presented at the Second International Conference on Telephone Survey Methodology, Miami, FL (January 2006). https://www.rti.org/pubs/TSM2006_Wine_paper.pdf (accessed August 2016).

54 de Leeuw, E.D., Hox, J.J., and Boevé, A. (2016). Handling do-not-know answers: exploring new approaches in online and mixed-mode surveys. *Social Science Computer Review* 34 (1): 116–132.

55 Couper, M.P., Antoun, C., and Mavletova, A. (2017). Mobile web surveys: a total survey error perspective. In: *Total Survey Error in Practice* (ed. P. Biemer, E. de Leeuw, S. Eckman, et al.), 133–154. Hoboken, NJ: Wiley.

56 Roberts, A., Leeuw, E.D., Hox, J. et al. (2013). [In Dutch: We can put more fun in it: Online questionnaire design using an auto-advance carrousel matrix instead of a standard grid] Leuker kunnen we het wel maken. Online vragenlijst design: standard matrix of scrollmatrix? In: Ontwikkelingen in het Marktonderzoek 2013. 38^{e}Jaarboek van de MOA [In Dutch: Developments in Market Research] Jaarboek 2012 (ed. A.E. Bronner, P. Dekker, E. De Leeuw, et al.), 133–148. Haarlem: SpaarenHout.

57 Christian, L.M., Dillman, D.A., and Smyth, J.D. (2005). Instructing Web and Telephone Respondents to Report Date Answers in Format Desired by the Surveyor. *Technical Report 05–067*. Washington, DC: Social and Economic Sciences Research Center, Washington State University. https://www.sesrc.wsu.edu/dillman/papers/2005/instructingwebandtelephone.pdf (accessed January 2017).

58 Groves, R.M. and Couper, M.P. (1998). *Nonresponse in Household Interview Surveys*. New York: Wiley.

59 Groves, R.M., Dillman, D.A., Eltinge, J.L., and Little, R.J.A. (2002). *Survey Nonresponse*. New York: Wiley.

60 Hox, J., de Leeuw, E., and Klausch, T. (2017). Mixed mode research: issues in design and analysis. In: *Total Survey Error in Practice*, (ed. P. Biemer, E. de Leeuw, S. Eckman, et al.), 511–530. Hoboken, NJ: Wiley.

61 Groves, R.M., Fowler, F.J., Couper, M.P., Lepkowski, J.M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology*, 2 (p. 248), Hoboken, NJ: Wiley.

62 Vannieuwenhuyze, J. and Loosveldt, G. (2013). Evaluating relative mode effects in mixed-mode surveys: three methods to disentangle selection and measurement effects. *Sociological Methods and Research* 42 (1): 82–104.

63 Lee, S. (2006). Propensity score adjustment as a weighting scheme for volunteer panel web surveys. *Journal of Official Statistics* 22 (2): 329–349.

64 Vannieuwenhuyze, J., Loosveldt, G., and Molenberghs, G. (2010). A method for evaluating mode effects in mixed-mode surveys. *Public Opinion Quarterly* 74 (5): 1027–1045.

65 Cernat, A. (2015). Evaluating mode differences in longitudinal data: moving to a mixed mode paradigm of survey methodology. PhD thesis. University of Essex.

66 Revilla, M. (2015). Comparison of the quality estimates in a mixed-mode and a unimode design: an experiment from the European Social Survey. *Quality and Quantity* 49 (3): 1219–1238.

67 Jäckle, A., Roberts, C., and Lynn, P. (2010). Assessing the effect of data collection mode on measurement. *International Statistical Review* 78 (1): 3–20.

68 Jäckle, A., Roberts, C., and Lynn, P. (2006). Telephone Versus Face-to-Face Interviewing: Mode Effects on Data Quality and Likely Causes. Report on Phase II of the ESS-Gallup Mixed Mode Methodology Project. *Institute for Social and Economic Research Working Paper 2006 (41)*, Colchester. https://www.iser.essex.ac.uk/files/iser_working_papers/2006-41.pdf (accessed September 2016).

69 Klausch, T. (2014). Informed design of mixed-mode surveys: evaluating mode effects on measurement and selection error. PhD dissertation. Utrecht University. http://dspace.library.uu.nl/handle/1874/300673 (accessed October 2016).

70 van de Schoot, R., Schmidt, P., DeBeuckelaer, A. et al. (2015). Editorial: measurement invariance. *Frontiers in Psychology* doi: 10.3389/fpsyg.2015.01064.