Survey Research Methods (2018)
Vol. 12, No. 2, pp. 9999-10013
doi:10.18148/srm/2018.v12i2.7402

ISSN 1864-3361
http://www.surveymethods.org

Mixed-Mode: Past, Present, and Future

Edith D. de Leeuw
Department of Methodology & Statistics
Utrecht University
The Netherlands

Mixed-mode surveys have been around since the late 1980s. In the past thirty years, major changes in technology and society influenced and changed data collection and survey methodology. However, in those years, mixed-mode strategies remained part of the daily survey practice, although the type of mix implemented followed the changes in technology and data collection methods. In this paper, I summarize the state of the art in traditional mixed-mode surveys and discuss implications for mixed device surveys.

Keywords: offline surveys; online surveys; mobile surveys; equivalence; multiple modes; multiple devices; mode selection effect; mode measurement effect; prevention; adjustment

1 Introduction

One of the earliest mentions of mixed-mode in a methodological publication was by Don Dillman and Jon Tarnai in 1988, who discussed the advantages and challenges of mixing face-to-face, mail, and telephone surveys. Thirty years later, mixing survey modes appears almost inevitable, especially for academic research and official statistics. Today, online surveys have been added to the mix and auxiliary observational data are becoming more and more important. Furthermore, the growing need for comparative international data has led to an increase in international survey programmes, and as individual countries differ in survey technology and survey customs, mixed-mode is almost inevitable in international comparative programmes (de Leeuw, Suzer-Gurtekin, & Hox, 2018). Mixed-mode surveys not only became a buzzword in innovation, as Dillman and Tarnai (1988) stated, but were adopted as a necessity in field practice (Biemer & Lyberg, 2003; Blyth, 2008). Mixed-mode strategies have been with us for a long time now and according to Tourangeau (2017) its usage is only expected to increase.

Contact information: Edith D. de Leeuw, Plantage Doklaan 40, NL-1018 CN Amsterdam, The Netherlands (email: edithl@xs4all.nl)

* Editors' note: This article is not the kind of paper usually published in SRM. However, SRM is the journal of the European Survey Research Association (ESRA), and this text is a reviewed and revised version of the keynote address of the 2017 ESRA conference in Lisbon, Portugal. By publishing this article, SRM documents the keynote speech of this important event of its funding institution.

Data collection techniques used in present and future mixed-mode approaches will undoubtedly be different from the ones mentioned by Dillman and Tarnai in 1988; since the eighties of the last century a technological revolution has taken place (Couper, 2013), which deeply influenced data collection methods, and mobile data collection, administrative and social media data will undoubtedly be part of the future mix (de Leeuw et al., 2018). Today, just as in 1988, there are four important reasons for using a mixed-mode survey design: improving coverage, increasing response rates and reducing nonresponse error, reducing cost, and a potential for better measurement (e.g., through reduced social desirability in self-administered modes). However, there are also potential drawbacks, such as an increased logistic burden and a potential for mode specific measurement error when data from different subgroups are collected through different modes.

In the next sections, I will summarize the state of the art in mixed-mode surveys and discuss implications for mixed-device surveys and the future. I will focus on the Total Survey Error (TSE) perspective, the distinction between wanted and unwanted effects of mixing data collection modes, and the importance of designing and adapting to the needs of the respondent.

2 Present

2.1 Common types used

At present, we see many forms of mixed-mode surveys being used in practice. A distinction should be made between using multiple modes of contact to communicate with sample units and using multiple modes of response, that is, using multiple modes for data collection. A clear typology of multiple mode survey systems, the rationale for using it, and the potential effect on survey quality is given by de Leeuw,

Dillman, and Hox (2008, Fig. 16.1) and de Leeuw (2005, Fig. 1).

Multiple modes of contact. The practice of using a different mode or even multiple modes to contact respondents has been around for a long time; a prime example is the use of pre-notifications or advance letters in interview surveys (de Leeuw, Callegaro, Hox, Korendijk, & Lensvelt, 2007; Tourangeau, 2017). Today, employing multiple modes of contact is a necessity and current best practice for the establishment of probability-based online panels (Blom et al., 2015). A major challenge for sampling and contacting potential panel members is the absence of general sampling frames of Internet users, as representative email lists from which random samples could be drawn are not available, except for special populations (e.g., employees of a certain company, teachers of a specific school). To solve this problem, a probability sample is drawn using off-line sampling frames (e.g., a registry-based or address-based sample), and the resulting sample is then approached with off-line contact methods to solicit their participation. For example, in establishing the Dutch LISS panel, all sample units received an advance letter with a brochure, and were then contacted by telephone or face-to-face, depending on the availability of telephone numbers. This was followed by a 10-minute recruitment interview. Furthermore, respondents, who did not have Internet access, but were willing to participate in the online panel, were provided with the necessary equipment. Finally, all willing respondents, both those who initially already had internet access and those who were provided with Internet, were given detailed instructions (Scherpenzeel & Das, 2011). Similar procedures were later followed in establishing online panels in Europe (e.g. GIP, ELIPSS, GESIS Panel) and in the USA (e.g., Knowledge network/GFK), see for instance Blom et al. (2015), Bosnjak et al. (2018).

A second example of using multiple modes of contact is the combination of traditional paper advance letters and email messages to "push respondents to the web" (Dillman, 2017). Through paper advance letters, researchers can convey legitimacy and trust, which is of the utmost importance in online surveys. An advance letter may take away fear or distrust, making it less likely that following emails are seen as "spam." It makes people aware of the coming surveys, and may raise interest. Furthermore, in a paper letter, researchers can include a prepaid incentive, which is far more effective in achieving a good response than promised incentives (for an overview, see Singer, 2002; Singer & Ye, 2013). However, contacting respondents by paper mail, and then requesting them to go to a computer or other device, connect to the web, and type in a URL and password, is not an ideal situation and this may take its toll on response rates. Therefore, in push-to-the web approaches, for those of which an email address is known, the paper letter is followed-up with an email notification containing an easily clickable URL, allowing the respondent to immediately access the survey questions.

In sum, multiple modes of contact are employed to improve coverage (e.g., online panels) and reduce nonresponse and nonresponse error by improving the opportunities to persuade people to respond (e.g., advance letters, push-to-the-web methods). Advance letters in general improve response in interviews (e.g., de Leeuw et al., 2007) and are current best practice in face-to-face and telephone interviews (Tourangeau, 2017). When different modes of communication are used, but the data collection itself is in one single mode, there are no threats to mode effects in measurement and using multiple modes of contact is beneficiary as it improves coverage and response (de Leeuw, 2005).

Multiple modes of response. Improving coverage and response at affordable costs are also the main reasons for using multiple modes of response. These modes may be offered at the same time, using a so-called concurrent mixed-mode design, or by following up one data collection mode with one or more different modes in a sequential mixed-mode design. A prime example of a sequential mixed-mode approach is the American Community Survey, in which cost-effective self-administered modes are followed-up with more expensive interview modes to reduce nonresponse (American Community Survey, 2014). A good example of concurrent mixed-mode design is the use of an across countries mixed-mode approach in international comparative surveys, such as the International Social Survey Programme (ISSP); for an overview see de Leeuw et al. (2018) Other examples are employing specific data collection methods for specific subpopulations, such as ethnic minorities (see for instance Kappelhof, 2017), or providing those without Internet or those who prefer to respond off-line with paper mail surveys (Bosnjak et al., 2018, e.g., GESIS panel,). A special case of a concurrent mixed-mode is a targeted design in which persons with a high response propensity are allocated to the more cost effective web or mail modes, while reserving the costly face-to-face interviews for persons with a low response propensity, though empirical examples are still scarce (Lynn, 2017).

A well-known form of mixed-mode in the response phase is the use of a self-administered mode within an interview mode for special, sensitive questions. Here, all respondents are questioned in a different mode for part of the questionnaire with the specific goal of reducing social desirability bias for specific questions and reducing overall measurement error. This form goes back to the old days of paper and pencil interviews, where respondents were handed over a paper, self-administered, questionnaire for a specific subsection of the questionnaire. Self-administration of sensitive questions is now standard best method in CAPI-surveys, where CASI or Audio-CASI modules are available and the interviewer hands over the computer to the respondent for sensitive parts of the interview (e.g., de Leeuw, Hox, & Kef, 2003; Turner et al., 1998).

The onset of online surveys (Couper, 2000) gave a new impulse to the use of multiple response modes for data collection (de Leeuw, 2005, 2011). This was partly driven by concerns about Internet coverage. Although Internet penetration has rapidly increased over the years, it still differs widely between countries. In Europe, the average Internet penetration rate was over 85% in December 2017 (Internet World Stats, 2018), but this ranges from an almost full 99% penetration for Iceland to 53% for San Marino. However, even for countries with a high Internet penetration, such as the Netherlands and Germany (both 96%), there is an important difference between those who have and those who have not: the digital divide. This is illustrated by two detailed reviews for Europe and the US (Mohorko, de Leeuw, & Hox, 2013; Sterrett, Malato, Tompson, & English, 2017). A combination of two or more modes of response in one study can reduce the undercoverage of specific subgroups, such as the elderly or the lower educated. A good example is the German GESIS panel, which combines online surveys with paper mail surveys for those without Internet. Another example is the first wave of the Dutch Labour Force Survey, for which a web and interview mix is used (Blanke & Luiten, 2014).

2.2 Effects on coverage

Do we reach our goals; does offering multiple modes reduce coverage error? There are only a few empirical studies that focus on mixed-mode and coverage. Most studies point to the necessity of including a second or even a third mode. In a study about immigrants in the Netherlands, Kappelhof (2015) states that different groups of respondents do participate in different modes (e.g., older female respondents more in CATI and younger and second generation respondents more online), but that a single mode CAPI results in the best reflection regarding socio-demographics of immigrants. When investigating mail and web combinations, both Messer and Dillman (2011) in the US and Bandilla, Couper, and Kaczmirek (2014) in Germany found that web only excludes segments of the population, and web plus mail gave a better representation on demographic variables. However, Klausch, Hox, and Schouten (2015) in a multiple mode (mail, telephone, web, face-to-face) experiment of Statistics Netherlands only found improvement of representativity on demographics when traditional mail and telephone surveys were followed up with a face-to-face interview; the demographic representativeness of the web survey in the Netherlands was already optimal and the face-to-face follow-up did not change this. Finally, Cornesse and Bosjnak (2018) looked into survey characteristics and representativeness; in their meta-analysis they found evidence that mixed-mode surveys when compared to single mode surveys were more representative.

2.3 Effects on nonresponse

Nonresponse has been worrying survey researchers for a long time, as testified by numerous publications and almost three decades of yearly meetings of the International Workshop on Household Survey Nonresponse[1]. A study by de Leeuw and de Heer (2002) showed that nonresponse has been increasing over countries and years. In a follow-up, Luiten, de Leeuw, Hox, and Schouten (2017) collected additional data for recent years. Preliminary analyses (Luiten et al., 2017) showed that nonresponse is still increasing over the years. Nonresponse not only leads to smaller realized samples and fewer cases for analysis, but also brings with it the potential of nonresponse bias, that is, a difference between respondents and nonrespondents on key variables (Groves & Petcheva, 2008). In the past, single mode nonresponse follow-up studies have been performed to increase response and investigate potential for bias (e.g., Stoop, 2005). In the interest of cost-effectiveness, multiple modes are now used. A well-known format implements sequential mixed-mode studies, starting with the most affordable mode, such as web or mail, while using the more expensive interview modes to get at the hard to reach and unwilling respondents. The American Community Survey (American Community Survey, 2014) is a successful example of a cross-sectional survey in which multiple modes are offered in sequence, starting with the affordable mail/online modes and following-up nonrespondents with interviews. The Understanding Society Innovation Panels provide a good example in a longitudinal setting of both a sequential CATI-CAPI mixed-mode experiment (Cernat, 2015) and a sequential Web-CAPI experiment (Bianchi, Biffigandi, & Lynn, 2017).

In the early days of mixed-mode studies, offering respondents a choice (a form of concurrent mixed-mode design) was seen as an attractive option. The reasoning behind this is that people may differ in their preferences for certain data collection methods; offering them a choice can therefore be seen as respondent-oriented, may create goodwill, and boost response rates as a result (e.g., Olson, Smyth, & Wood, 2012). However, empirical evidence for this is lacking. Most studies suggest that a sequential mixed-mode design, in which one mode is offered after another, is better than a concurrent design, which offers respondents a choice of mode. In a meta-analysis of 19 experimental comparisons of web and mail, Medway and Fulton (2012) showed that offering a mode choice actually reduces response rates with on average 3.8 %. A potential explanation is that offering a choice complicates the decision: instead of one decision (will I respond or not), someone has two make two decisions (will I respond and if so in what mode), causing some potential respondents to postpone their decision. In addition, respondents who decide to answer online, but do not have the

[1] See http://www.nonresponse.org/

opportunity to go online immediately, may eventually forget to do so at all (e.g., Dillman, Smyth, & Christian, 2014, chap. 11).

There is some evidence that when a web choice-option is offered more prominently than a mail option, respondents tend to choose the web more often (Tancreto, Zelenak, Davis, Ruiter, and Matthews, 2012, op cit. Tourangeau, 2017). However, going one step further and "push respondents to the web", by sequentially offering web first while also making it easy and convenient to access the online survey, works best. Based on experimental evidence (see Dillman & Edwards, 2016, for a detailed overview), it is advised to start with a paper mail request to respond on the web; this contact will legitimize the survey and make it possible to send prepaid (unconditional) incentives. This paper mail letter is then followed by multiple communications through different contact modes (e.g., an email with an easily clickable URL) to enhance the earlier communications. Furthermore, different response modes are being offered in sequence: web first and later a follow-up by mail questionnaire. Here, the order (web first and then mail and not the other way around), is of importance; this is not only important for cost-efficiency, but also for an increase in response. Of course, if the preferred mode of certain people is already known, for instance, from earlier communications or from the recruitment interview for a longitudinal survey or a panel, offering them their preferred mode is better (Olson et al., 2012; Smyth, Olson, & Kasabian, 2014). A good example of such a tailored-to-preference approach is the German GESIS-panel, for which those who prefer to respond off-line and those without Internet are provided with paper mail questionnaires (Bosnjak et al., 2018).

Most research on the effect of mixed-mode designs on response has been based on evidence from cross-sectional surveys in which fresh samples are used, less is known about longitudinal surveys. For longitudinal studies, high and steady response rates are extremely important (Jaeckle, Lynn, & Burton, 2015). Respondents also have prior experience and are familiar with the standard mode, which could hamper the introduction of a new mode. Hence, researchers have been more hesitant to introduce mixed-mode designs in single mode longitudinal surveys, fearing potential adverse effects. A carefully implemented sequential mixed-mode experiment in the Understanding Society Innovation Panel sheds light on this. In wave 1 to 4, members were interviewed face-to-face, in wave 5 a sequential mixed-mode Web-CAPI approach was experimentally compared with the standard CAPI-only approach. Although there was clear evidence of cost-savings in the mixed-mode, the effect on response was less positive, indicating a slightly lower response for the mixed-mode approach in wave 5 (for details, see Jaeckle et al., 2015). The experiment was continued in the subsequent waves 6 and 7, where respondents were approached with the same (mix of) modes as in wave 5. Analysis of the full dataset (Bianchi et al., 2017, Tab. 2 & 3) showed that the response rate for mixed-mode recovers in wave 6 and 7 and is in fact even slightly higher for the mixed-mode than the single mode approach. No statistical differences between mixed-mode and single mode were found in overall response and in response amongst important subgroups. Furthermore, there were clear indications that a mixed-mode design leads to potential cost savings.

2.4 Effects on measurement

Win-Win: Another response mode for all respondents, for a subset of specific questions. Combining multiple modes can have a beneficial effect on measurement, but it depends on the type of mixed-mode design that is implemented. When a different mode is used for a specific part of the questionnaire, and provided that all respondents receive this mode, measurement will, in general, improve. An example is mixing self-administered and interview modes to exploit the strong points of both methods. Within an interview, a self-administered mode (paper questionnaire, CASI or A-CASI module) is used for sensitive questions, to enhance privacy for the respondent, and other (e.g., household roster or complex) questions are administered by an interviewer. Such designs are often used to reduce social desirability bias. Numerous studies (e.g., Kreuter, Presser, & Tourangeau, 2008; Tourangeau & Yan, 2007; Turner et al., 1998) have shown that self-administered forms (both paper mail and computerized or internet) resulted in less social desirability and more openness in sensitive answers than interviews. For an early meta-analysis on paper-and pencil methods see De Leeuw (1992, chap. 3). For a narrative overview that compares online and interview surveys both in Europe in the USA, see de Leeuw and Hox (2011); for a meta-analysis of US studies, see Tourangeau, Conrad, and Couper (2013, pp. 140–142).

Data collected with survey methods may also be combined with data from other sources, such as administrative data (e.g., Calderwood & Lessof, 2009). Examples are SHARE in Germany and NHIS in the USA; for a discussion of linking surveys and administrative data, see Sakshaug and Antoni (2017). Recently, there have been experiments with mixing online data collection methods with new, innovative methods. Examples, in a medical context, include the use of smart weighing scales to automatically record weight and fat percentages of respondents in an online panel, and the use of accelerometers in an activity study, another example is the use of GPS data in mobility studies; for details, see Scherpenzeel (2017). A major challenge when it comes to enhancing survey data with data from other sources is respondents' consent to linkage and potential non-consent bias (e.g., Sakshaug & Kreuter, 2012). For instance, Scherpenzeel (2017, Tab. 1) reports on the willingness to participate in a study with smart devices by LISS panel members, a group that is,

in principle, willing, and responds to online questionnaires about once a month. In four studies, around 75% of the panel members completed the invitation questionnaire; this reflects the customary monthly response in the LISS panel. From these respondents, 37% replied they were willing to participate in a smart phone GPS study (which brings the total to 28% of wiling LISS-members), while 56% and 57% replied that they were willing to use smart weighting scales and accelerometers (which brings the total to 40% and 44 % respectively). An overview regarding mobile devices and consent with special request is given by Couper, Antoun, and Mavletova, 2017, pp 141-142). However, more research is needed to get insight in socio-demographic and attitudinal correlates of willingness and on how to best approach respondents with requests to use new technologies for research in addition to their survey responses.

Danger of mode effects: Different modes for different respondents. When different groups of respondents complete a survey in different modes there is a risk of differential measurement errors. Previous research suggests a dichotomy between self-administered and interview modes. Self-administered modes, be it with paper and pen or online, perform better with sensitive questions and result in less social desirability and more openness in answers than face-to-face and telephone interviews (e.g., De Leeuw, 1992; Tourangeau et al., 2013). So, if nonrespondents in a mail survey are followed-up with face-to-face interviews in a sequential mixed-mode design, this social desirability effect may influence their answers on sensitive questions, for example, respondents in the interview mode may underreport the amount of alcohol they consume. Similarly, when in a longitudinal study, a switch is made from interview to web, this may threaten to comparability over waves of responses to the more sensitive questions. However, with more neutral questions a mode switch does not necessarily have to threaten comparability over waves Allum, Conrad, and Wenz (2018) showed that a mode switch from face-to-face to web in the Understanding Society Innovation panel only produced small differences in recall accuracy across modes.

Some modes are more disparate from each other than others, and the more modes differ in important factors, the greater the risk of measurement effects. Not surprisingly, most mode differences have been found between interview surveys (be it telephone or face-to-face) and self-administered surveys (be it paper and pen or online). In an early meta-analysis, De Leeuw (1992, chap. 3) found that mail surveys resulted in less social desirability than both telephone and face-to face interviews, and that both interview modes did not differ from each other. Furthermore, face-to-face interviews produced the least amount of item nonresponse (missing data). A recent meta-analysis (Tourangeau et al., 2013, p. 140..142) confirms that online surveys also result in less social desirability than interviews. Regarding other indicators, like satisficing and item-nonresponse, the evidence is mixed. Some studies report less in interviews, others do not find any difference; for a narrative review, see de Leeuw and Hox (2011, pp. 56–61).

Telephone and face-to-face interviews have much in common. Both are interviewer- administered with potential positive effects, like interviewer assistance and motivation, and potential negative effects regarding privacy and social desirability. Both are also mainly using aural channels for information transmission, as questions are read out loud and respondents convey their answers orally (De Leeuw, 1992; C. Roberts, 2007). Consequently, when we focus on measurement the differences in data quality between comparable telephone and face-to-face interviews are small (Groves & Kahn, 1979; Jaeckle, Roberts, & Lynn, 2010); for a detailed review and meta-analysis, see De Leeuw (1992) and de Leeuw and Van der Zouwen (1988).

Similarly, online and paper mail questionnaires have much in common; both are self-administered with a high degree of privacy; both use visual communication and may use graphical language; in both, the respondent is the locus of control, and both are self-paced (e.g., Couper, 2011; De Leeuw, 1992; Dillman et al., 2014). Empirical research confirms that, in many cases, self-administered paper and online questionnaires are very similar regarding mode measurement effects. For instance, computerized tests have been found to be measurement equivalent to their earlier paper versions, both for cognitive abilities and more subjective self-assessments. For a detailed overview, see de Leeuw and Hox (2011, pp. 56–59).

In sum, there is a lower risk of mode measurement effects when online surveys and self-administered questionnaires are used in a multiple mode design than when interviews and self-administered surveys are mixed. However, much depends on how questionnaires are designed and implemented in both modes. Furthermore, the danger of mode differences is greatest when sensitive topics are the focus of the study.

3 Mode selection vs mode measurement effects

3.1 Wanted vs unwanted mode effects

Although in early publications the term mode effect is often used as a general term, we should realize that mode effects as such do not exist. Tourangeau et al. (2013) point out that mode effects have two components: differential non-observation error (who responds and who does not in which mode) and differential observation error (how do they respond in a mode). What is generally labelled as mode effect is the net effect of both non-observation and measurement error differences between the modes (Hox, de Leeuw, & Klausch, 2017; Kappelhof, 2017; Tourangeau, 2017). In other words, there is a confounding of mode selection effects

and mode measurement effects. This is clearly illustrated in Figure 1.

When we use multiple modes in one survey, we aim at mode selection effects where each mode brings in different respondents. Selection in a mixed-mode study is a desired effect that reduces the coverage and nonresponse errors of single mode designs, and it is one of the main reasons why mixed-mode approaches are implemented. If there was no mode selection effect and the same type of respondents react in the same proportions to all modes, adding a second or a third mode would only bring in more of the same, and we might as well stick to the most affordable mode of data collection (e.g., online survey). Mode measurement effects, on the other hand, are a source of unwanted measurement error, for which researchers ideally want to correct. It is therefore important to discern between wanted mode selection effects and unwanted mode measurement effects.

3.2 Evidence for mode measurement effects after controlling for selection effects

Few mode comparison studies have explicitly addressed and corrected for mode selection effects when studying mode measurement effects. In an early study, De Leeuw (1992, chap. 5) used covariate analysis to investigate differences between face-to-face, telephone, and mail surveys. Her results confirm earlier mode comparisons. After correcting for selection effects, she found that the mail survey resulted in more self-disclosure on sensitive questions, but also in more item-nonresponse than both interview modes. No differences between the modes were found for acquiescence, extremeness, and primacy and recency effects, after adjustment for selection. Allum et al. (2018) found only small differences in recall accuracy after a mode switch from face-to-face to web, after correcting for selection effects. Heerwegh and Loosveldt (2011) analyse a mixed-mode mail-telephone study into crime victimization. After controlling for differences between the modes in socio-demographic variables, they still find a more favourable attitude to the police in the telephone mode, and this is attributed to social desirability. A similar conclusion was reached by Gordoni, Schmidt, and Gordoni (2012) when comparing face-to-face and telephone interviews. After correcting for socio-demographic differences between the modes, measurement errors tended to be higher in the telephone mode (for a detailed overview, see Hox et al., 2017).

Kappelhof (2017) describe how in special surveys of non-western immigrants in the Netherlands, after correcting for mode selection effects, there is less social desirability online than in interviews, especially when interviewers and respondents have a shared background. Finally, Tourangeau (2017, Tab. 6.3) reviews an election study in which mail and telephone survey modes are used and where information was available about whether sample members had actually voted. Careful analysis showed that there were mode selection effects, but also mode measurement effects, with the result that the telephone mode showed the largest effects of underreporting of voting.

4 TSE Perspective for handling multiple mode surveys

From a Total Survey Error perspective one wants the best of all worlds and aims to reduce overall error. When researchers employ mixed-mode surveys, they should therefore carefully design the total survey process. The steps involved in a high quality mixed-mode study are: (1) designing data collection and equivalent questionnaires, (2) diagnosing, that is, estimating mode effects by separating intended mode selection effects from unintended mode measurement effects, and (3) if necessary, adjusting for unintended differential mode measurement error. It is important to note that auxiliary data are required for both diagnosis and adjustment, just as auxiliary data are required for nonresponse adjustment. Furthermore, just as in studies into nonresponse bias and adjustment, researchers need to think ahead and already incorporate the possibilities for collecting auxiliary information in the design phase.

4.1 To mix is to design

Questionnaire design. Early methodological publications mainly focused on preventing unwanted mode measurement effects by means of good questionnaire design (de Leeuw, 2005; Dillman, 2000). As Dillman and Christian (2005) point out, different data collection modes have different traditions in questionnaire format. Examples include explicitly offering, or not offering, do-not-know options, check all that applies vs yes-no responses, using a matrix (grid)-format vs one question at a time. However, question format does influence response distributions, even within a single mode (e.g., Christian, Dillman, & Smyth, 2008; Smyth, Dillman, Christiian, & Stern, 2006), and changing the question format in one mode and not in another may exacerbate mode measurement effects, which in turn may result in differential measurement error between modes. In daily practice, questionnaire designers may therefore unintentionally enhance mode measurement effects by routinely using different question formats across modes. This practice is especially dangerous when comparisons between groups or countries are the goal of the study, as differential mode measurement effects threaten the comparability of the data collected (de Leeuw et al., 2018; Tourangeau, 2017). To avoid this, the first step in a mixed-mode study is designing and using equivalent questionnaires.

Equivalent questionnaires are not the lowest common denominator. The aim of designing equivalent questionnaires is to improve the questionnaire in general and get better instruments. A good starting point is the unified (uni-) mode construction, in which particular efforts are made to present a

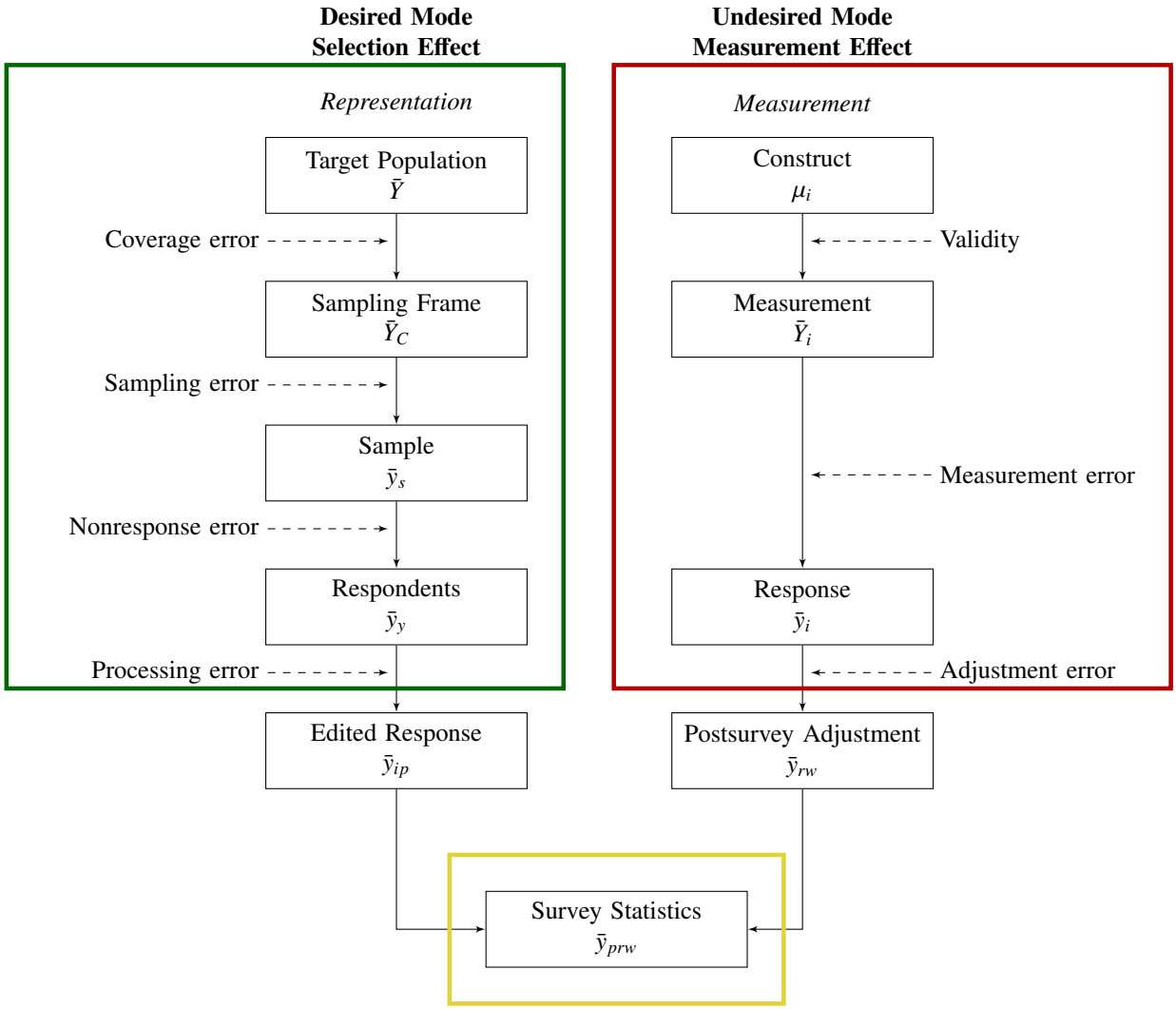

Figure 1. Influence of Wanted Mode Selection Effects and Unwanted Mode Measurement Effects on the resulting Estimated Survey Statistic. Adapted and extended Figure 2.2. (**groves09**)

similar stimulus across modes (Dillman & Edwards, 2016). To achieve this, the same question structures and wording should be used across modes and similar instructions and visual formats (on screen and paper) are essential (e.g., Dillman et al., 2014). An early example of the required design steps to arrive at equivalent questionnaires can be found in De Leeuw (1992, pp. 36–38); detailed guidelines are given in Dillman et al. (2014, chap. 11)[2] . In the above paragraphs, the words "equivalent" and "similar" are used on purpose. Just as a literal translation from one language into another sometimes does not make sense (e.g., Harkness, 2008); the exact same wording may be perfectly understandable in one mode, but somewhat strange in another. If questionnaire designers have good reasons to deviate from the exact wording or format across modes, they should do so. The purpose of equivalent questionnaire design is to maximize data quality in a specific mode and minimize differences in data across modes[3]. To do so it is necessary to follow the rules of communication and information transmission associated with specific modes; for a comprehensive introduction into mode differences, see Couper (2011), de Leeuw and Berzelak (2016). A simple example is the use of instructions or explanations for the respondent; in order to make sense these must be adapted to the mode (e.g. "click next to continue" online or "I will now read out a list of…" in interviews).

Using the build-in intelligence in electronic questionnaires (online or CAPI and CATI) moreover increases the

[2]For an overview see also the student resources accompanying Dillman et al. (2014, chap. 11); see http://bcs.wiley.com/he-bcs/Books?action=chapter&bcsId=9087&itemId=1118456149&chapterId=103125.

[3]E.g., see http://www.census.gov/srd/mode-guidelines.pdf

quality of data, reduces response burden and meets respondents' expectations (e.g., routing, automatic fills). Finally, the flow of the question-answer sequence and the rapport with the respondent have different requirements in interviewer-administered vs self-completion modes. When an interviewer has to read out the same instructions time and again when asking a set of questions, it may seriously hamper the flow of the interview. Hence, interviewer instructions in questionnaires often state, "repeat if necessary", and a well-trained interviewer will notice whether this is necessary, whereas the instruction in a self-administered questionnaire should always be accessible for the respondent[4].

The following paragraph offers two detailed examples of the creative use of the intelligence of computer-assisted systems and interviewer skill to construct equivalent questions that optimize the quality of data collection in each mode and minimize differences between the modes.

Equivalence I: Do-not-know-option in interviews and online surveys. In both telephone and face-to-face interviews, a "do-not-know" option is usually not explicitly offered to the respondent, but it is accepted when a respondent spontaneously answers do-not-know. In quality surveys, interviewers are well-trained to use a friendly "probe" after a spontaneous "do-not-know" answer to reduce uninformative answers (e.g., Gwartney, 2007; Sudman & Bradburn, 1982) and this is also often the standard-option in software for computer-assisted interviewing (Bethlehem & Biffigandi, 2012). However, in online surveys, web designers are inclined to either explicitly offer a "do-not-know"-option or to omit the do-not-know answers and make the question mandatory, followed by a - usually not so friendly - error message when a respondent tries to skip the question.

Respondents will clearly not perceive these web formats in the same way as they perceive a friendly interviewer probe after a spontaneous do-not-know-answer. In fact, these different formats result in different missingness patterns (de Leeuw, 2011). For instance, Heerwegh and Loosveldt (2011) report more missingness and do-not-know answers in an internet survey than in a face-to-face survey in which interviewers did probe after a do-not-know. However, when a do-not-know is directly accepted by the interviewer, the opposite may occur, and web surveys result in less item-nonresponse (Fricker, Galesic, & Yan, 2005, e.g.,).

A pioneering study by Wine, Cominole, Heuer, and Riccobono (2006) suggested a clever way of designing equivalent solutions for the "do-not-know" option in interviews and online questionnaires. Forced to change from a telephone interview to a web survey, these researchers wanted to keep and use the advantages of an interview as much as possible and implement these online. In the Wine et al. (2006) approach, interviewer texts were analysed and used to emulate desired interviewer behaviour online, such as a friendly probe.

This approach was experimentally investigated by de Leeuw, Hox, and Boeve (2015) for online and telephone interviews, and by Baghal and P. (2015) for online and face-to-face interviews. In the online version of both studies, an emulated –friendly worded- interviewer probe after a question is left unanswered reduces item-nonresponse to the same degree as an interview with probing. Furthermore, de Leeuw et al. (2015) showed that both within and across modes, explicitly offering do-not-know leads to higher missingness; that probes reduce the amount of missing data in both (online and CATI) modes, and that probing does not negatively influence the respondents' evaluation of the survey. For more details, see de Leeuw et al. (2015).

In sum, when it is standard practice in an organization to use interviewer probes after non-informative answers, it is wise to also use similar, emulated, friendly probes after important questions in online surveys.

Equivalence II: Avoid matrix (grid) questions. In online surveys, series of similar questions are often presented in a matrix (grid) format, where the rows are the questions and the columns the common set of response options. All questions are presented simultaneously and visually together. In contrast, interview questions are asked sequentially, one at a time. These different formats are a clear example of how asking the same question across modes offers respondents different stimuli in different modes, which in this case may give rise to differential context effects caused by adjacent questions.

Furthermore, even within one mode (i.e., online surveys) researchers have been concerned about the effect of matrix questions on the resulting data quality. In general, completion times in web surveys are faster when grids are used than when item-by-item questions are used (for an overview, see Tourangeau et al., 2013, Tab. 4.1). This led to the fear that this faster pace in grid questions would result in satisficing behaviour, such as straight-lining or non-differentiation (Zhang & Conrad, 2014). A better visual design of grids and dynamic feedback may help to counteract the problems associated with grids (e.g., Couper, 2013; Kaczmirek, 2011).

An alternative to presenting a series of sub-questions in a grid or matrix is the auto-advance question, also known as HSM-format or carrousel question Callegaro, Manfreda, and Vehovar (2015, p. 82). Questions are presented on the screen one-by-one, mimicking the sequential interview process. After an answer has been chosen and submitted, the next question flies in using auto-advance procedures; there is no need to click on a "next" button, keeping respondent burden light. A visual navigation bar provides respondents with an overview and enables them to go back to a previous question. For an illustration of the auto advance format, see Figure 2; for a more detailed description, see de Leeuw and

[4]Examples Deirdre Giessen and Sjoertje Vos, Utrecht Summer school course, Survey Design, implementation, and data processing, 2016

Figure 2. Example of auto advance-carrousel format in Dutch. On top is the general introduction. Below the text of the question. This is the first question in a series of 5 (see navigation bar below response categories). A seven point response scale with radio buttons is used (Totally agree – Totally disagree). The Previous and Next button are disabled (grey).

Berzelak (2016).

Compared with traditional grid questions in an online survey, the auto-advance format performed well and resulted in less straight-lining and higher evaluations (A. Roberts, Leeuw, Hox, Klausch, & De Jongh, 2013). The auto-advance format also greatly reduced mode differences between online surveys and interviews, making the responses comparable across modes (J. Berzelak, 2014). This approach may also be a suitable form of optimization for mobile phones and other devices, making stimuli on mobile phone, tablet, and pc/laptop equivalent.

In sum, a new approach to question presentation with a careful analysis of the interview process, instead of simply implementing old paper and pen (grid) formats to online surveys, results in equivalent stimuli both in mixed-mode (online-interview) surveys and in mixed-device studies.

4.2 Estimating and adjusting for mode effects

Reducing mode measurement effects as far as possible is only the first step in a well-planned mixed-mode survey. The second and third step are estimating mode effects, thereby separating intended mode selection effects from unintended mode measurement effects and adjusting for these unintended mode measurement effects.

Estimation of mode effects. During the diagnosis/estimation phase it is important to discern between desired differential mode selection effects, which help reduce coverage and nonresponse error, and undesired differential mode measurement effects. Only then is it possible to estimate the undesired mode measurement effect while controlling for the desired selection effects, and if necessary adjust for undesired differential measurement error. To be able to do so, it is necessary to have access to extra information for analysis, and, in the design phase, researchers should already investigate and plan which sources of auxiliary data are available, and which auxiliary data can be collected additionally. In other words, researchers should design a mixed-mode study with the statistical analysis of selection and measurement effects in mind.

Often, researchers have access to some background variables, such as frame data, or collect background information for general nonresponse adjustment; sometimes, researchers can access administrative records or do additional record checks. In a mixed-mode study, the availability of good background variables is essential. Biographical (e.g., age, gender) and other background data (e.g., education, urbanicity) form a good starting point, but the more data the better. For instance, de Leeuw (2005) suggests an experimental approach to collecting additional data. This approach was used by Jaeckle et al. (2010), who performed several mode experiments in the ESS context. Another example is the work of Klausch (2014), who combined a split ballot approach with repeated measures through reinterviews. Successful non-experimental approaches use data from longitudinal surveys (Cernat, 2015) or data from an existing reference survey (Vanniewuenhuyze, 2013). For an introduction with examples, see de Leeuw et al. (2018, sec. 4).

Adjusting for unintended mode measurement effects. There are several statistical approaches for adjustment of unintended mode measurement effects while controlling for the intended mode selection effects. For instance, regression with mode as binary predictor and biographical covariates to control for selection; multigroup SEM analysis, which compares multi-mode groups, again while controlling for selection; potential outcomes (imputation) for missing modes using a matching procedure (e.g., predictive mean matching); or using a reference survey of a similar population or population data as golden standard.

These approaches all require a certain amount of statistical expertise; although covariate approaches and multigroup SEM comparisons have by now been described in statistical handbooks and good software is widely available. For more details: two good introductory publications on this topic are written by Kolenikov and Kennedy (2014) and Hox et al. (2017).

5 Never a dull moment

In 1956, the British Astronomer Royal predicted that space travel would be technologically impossible for a long time to come. One year later, the first Sputnik-satellite was successfully launched; in 1968, the first man walked on the moon, and since February 2018, a Tesla roadster with a dummy Starman is circling in space while its music system plays David Bowie's Space Oddity.[5] Predicting is difficult,

[5] https://www.cnbc.com/2018/02/06/this-is-what-a-tesla-roadster-looks-like-floating-through-space.html. For a live view of starman see (Note, you will not hear the music as soundwaves do not travel in the vacuum of outer space) https://www.youtube.com/watch?v=aBr2kKAHN6M&feature=youtu.be.

especially when the future is concerned. But of one thing I am sure, the life of a survey methodologist/statistician will never be a boring one.

The history of surveys is full of changes and adaptations. Perhaps the first mail survey (of Scottish ministers) was carried out by Sinclair in 1788, using written letters and 23 reminders, and resulting in a 100% response rate. The first social survey interviews were designed and initiated in the UK by Booth with publications in 1902; and in 1912, Bowley, again in the UK, introduced structured interviews and sampling techniques and even performed nonresponse analysis. Finally, in 1934, Neyman introduced probability sampling and sampling errors (for a historical overview, see e.g., Bethlehem, 2015; de Heer & de Leeuw, 1999; de Heer, de Leeuw, & van der Zouwen, 1999).

We have come a long way from the first mail coaches in Scotland and the social surveys in 19th century England to the present-day online and mixed-mode designs. Along the way, we learned how to perform quality telephone surveys, how to use all the assets of computer-assisted interviewing, and how to use the World Wide Web and online panels.

Society and technology are constantly changing, and our tools need to change too. We have just barely learned to meet the challenges of mixed-mode, and new challenges and opportunities are already presenting themselves. A major technological challenge facing survey researchers today are mobile devices, such as smartphones and tablets, which are increasingly used by our respondents to access the Internet. Web surveys are now morphing from a computer-oriented into a multi-device-oriented concept, and although a mixed-device survey is not a mixed-mode survey in the traditional sense of the word, as all devices are self-administered, the devices used vary widely in screen size, data entry interface, and social customs regarding their usage. For instance, tablets have larger screens and are more often used at home or in meetings, while smartphones are smaller and are continuously at hand in trains, buses, and on the road. The question therefore arises whether answers obtained from a smartphone are comparable to those obtained from tablets, and if answers obtained from these mobile devices are comparable to answers obtained from a pc or laptop.

Nevertheless, these technical challenges are the least of our problems, and we are now learning how to adapt questionnaires to mobile phones (for overviews, see Couper, Antoun, & Mavletova, 2017; Peterson, Griffin, LaFrance, & Li, 2017). In fact, mobile-first design is becoming a buzzword in online and mixed-mode surveys. More difficult to handle than technological challenges are the differences in social customs between pc/laptop and mobile device usage. Mobile devices are used at all places at all times. Tablets are a relaxed form of laptops and are with us on the couch, the comfy chair, or even in bed to read some papers, go online, browse new catalogues, do some online shopping, log into Facebook, and watch streamed TV. Furthermore, whereas in the old days a pc, and to a lesser extent a laptop, was a shared device in most households, tablets and smart phones are truly personal devices. Communication through mobile devices is private, users only want to connect within their own personal or business network and strangers are not allowed in. This poses challenges when it comes to persuading new, potential respondents to complete a questionnaire (e.g., Dillman, 2017) and response rates through mobile devices are generally lower than through PC (Couper et al., 2017, Tab. 7.2).

Finally, the way people, especially the younger generations, communicate through smartphones is completely different from the way we traditionally used the telephone or the web. Communication usually comes in the form of very brief and often sequential online interactions (e.g., texting, use of "WhatsApp", "SnapChat"), and not in that of a long telephone conversation. Perhaps, the traditional form of the online survey no longer connects with respondents and should be programmed more like a sequence of short questions and answers (Conrad, Schober, Antoun, & Yan, 2017).

When mixed device surveys are used, the attention span of respondents is shorter, and surveys should be brief. This poses a challenge to the traditional surveys. A potential solution to reduce questionnaire length for an individual respondent, while still collecting data on a larger set of questions for the total population of interest, is to split the questionnaire into smaller parts, and give groups of respondents only a part of the questionnaire (Graham, Taylor, Olchowski, & Cumsille, 2006). One popular questionnaire split is the three-form design. The questionnaire is split into four parts: part X, core questions posed to all respondents and equal parts A, B and C which are only posed to subsets of respondents. Modern missing data methods are then used to analyse the complete data set. I refer to Graham (2012) for an introduction to these methods, and a discussion on how to realize an optimal split for the questionnaire.

Mobile devices also present us with unique opportunities for improved measurement and additional observations. Examples include time-use diaries on a smartphone combined with experience sampling through a message or pop-up at random moments, or mobility studies using GPS-coordinates. Other examples include using mobile phone functions, or special devices or apps to collect additional information in health surveys or to measure physical activity (e.g., accelerometers). Once again, the problem here is the willingness to comply with such special requests. Even among potentially cooperative persons, such as members of online panels, compliance with these types of requests is not high. For an introduction to mixing online data collection with innovative methods, see Scherpenzeel (2017).

Lastly, we have entered the era of big data, of observational and available (found) data. At the 2017 ESRA conference, several sessions focussed on how surveys and big

data can work together[6] and in autumn 2018 a large conference is devoted to combining big data and survey science[7]. Intelligently combining big data and survey outcomes offers great opportunities to reduce survey burden and add precise measurement to additional variables to gain better insights. However, we should always remember our survey methodology and critically examine the purpose and origin of the available data itself by using the TE framework and asking ourselves who is covered, who is excluded, whether or not sampling is involved and what type of sampling, and how the main variables are operationalized (e.g., Couper, 2013; Hox, 2017). For instance, on social media, image management may distort our scraped data. After all, on Facebook or Instagram all food looks delicious and all babies are beautiful, have clean diapers, and never cry.

6 Conclusions

Mixed-mode and mixed-device surveys will be with us in the near future, offering opportunities and posing challenges. We, survey methodologists, are clever enough to meet the challenges of changing technology. At present, smart phones are being used in smart surveys, and questionnaires are being optimized for mixed-mode and mixed device surveys. More and more statistically sophisticated techniques become available to estimate and adjust for mode and device effects, and we learn to scrape the web, use qualitative text analysis, and add these data to survey outcomes. To paraphrase Dickens (1859): "it is the best of times, it is the age of wisdom". Or is it?

If we forget about the individuals behind the data, if we do not meet the changes in social customs and society and the differential adaptation of technology in different subgroups and cultures, it may well be: "the worst of times, the age of foolishness". Not all respondents adapt at the same rate, and when we are busy optimizing our questionnaires, we should take care not to be too advanced and risk losing less tech-savvy respondents, while at the same time being too basic may bore the smartphone generation. We need further and innovative research on how to best persuade the public to respond to our surveys, on the most successful way to ask them for permission to use, tracking devices, or other facilities on their smartphone, or how to get their informed consent for accessing their medical files, or other administrative data and link this information to survey data. If we combine technology and socio-psychological insights, we will meet respondents' needs, gather high quality data, and live in the best of all possible times and worlds.

Acknowledgements

This paper was presented as a keynote address at the ESRA conference in Lisbon in July 2017[8]. I thank Annelies Blom, Caroline Roberts, and the organizers for giving me the opportunity to present these ideas and Ulrich Kohler for encouraging the submission of this paper. I am also grateful to Bella Struminskaya, Joop Hox, Don Dillman, and Elizabeth Verwey for their helpful comments on an earlier draft. Finally, I want to thank Lars Lyberg, who urged me to start writing about mixed-mode surveys way back in 2005.

References

Allum, N., Conrad, F., & Wenz, A. (2018). Consequences of midstream mode switching in a panel survey. *Survey Research Methods*, *12*(1), 43–58.

American Community Survey. (2014). Design and methodology, chapter 7: data collection and capture for housing units. Retrieved from https://www2.census.gov/programs-surveys/acs/methodology/design_and_methodology/acs_design_methodology_ch07_2014.pdf

Baghal, T. & P., L. (2015). Using motivational statements in a web instrument design to reduce item missing rates in a mixed mode context. *Public Opinion Quarterly*, *79*, 568–579.

Bandilla, W., Couper, M., & 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).

Berzelak, J. (2014). Mode effects in web surveys (doctoral dissertation). University of Ljubljana, Slovenia. Retrieved from http://dk.fdv.uni-lj.si/doktorska_dela/pdfs/dr_berzelak-jernej.pdf

Bethlehem, J. (2015). Sunday shopping-the case of three surveys. *Survey Research Methods*, *9*(3), 221–230.

Bethlehem, J. & Biffigandi, S. (2012). *Handbook of web surveys*. Wiley Series in Survey Methodology. Hoboken, NJ: Wiley & Sons.

Bianchi, A., Biffigandi, S., & Lynn, P. (2017). Web-face-to-face mixed-mode design in a longitudinal survey: effects on participation rates, sample composition, and costs. *Journal of Official Statistics*, *33*(2), 385–408. doi:10.1515/JOS-2017-0019

Biemer, P. & Lyberg, L. (2003). *Introduction to survey quality*. Wiley Series in Survey Methodology. Hoboken, NJ: Wiley & Sons.

Blanke, K. & Luiten, A. (2014). Query on data collection for social surveys. ESSnet project: data collection for social surveys using multiple modes. Retrieved from http://ec.europa.eu/eurostat/cros/system/files/Query_report_DCSS.pdf_en

[6] https://www.europeansurveyresearch.org/conference/programme17

[7] https://www.bigsurv18.org

[8] A copy of the presentation is available in the online supplement to this article.

Blom, A., Bosnjak, M., Cornilieau, A., Cousteaux, A.-S., Das, M., Douhou, S., & Krieger, U. (2015). A comparison of four probability-based online and mixed-mode panels in Europe. *Social Science Computer Review*, *34*(1), 8–25.

Blyth, B. (2008). Mixed mode: the only "fitness" regime. *International Journal of Market Research*, *50*(2), 241–266.

Bosnjak, M., Dannwolf, T., Enderle, T., Schaurer, I., Struminskaya, B., Tanner, A., & Weyandt, K. (2018). Establishing an open probability-based mixed-mode panel of the general populaion in Germany: the GESIS panel. *Social Science Computer Review*, *36*(1), 103–115.

Calderwood, L. & Lessof, C. (2009). Enhancing longitudinal surveys by linking to administrative data. In P. Lynn (Ed.), *Methodology of longitudinal surveys* (pp. 55–72). New York: Wiley & Sons.

Callegaro, M., Manfreda, K., & Vehovar, V. (2015). *Web survey methodology*. London: Sage.

Cernat, A. (2015). The impact of mode design on measurement errors and estimates of individual change. *Survey Research Methods*, *9*(2), 83–99.

Christian, L., Dillman, D., & Smyth, J. (2008). The effect of mode and format on answers to scalar questions in telephone and web surveys. In J. Lepkowski, C. Tucker, M. Bryck, E. de Leeuw, L. Japec, P. Lavrakas, ... R. Sangster (Eds.), *Advances in telephone survey methodology* (p. 5). Hoboken, NJ: Wiley & Sons.

Conrad, F., Schober, M., Antoun, A., C. Hupp, & Yan, H. (2017). Text interviews on mobile devices. In P. Biemer, E. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. Lyberg, ... B. West (Eds.), *Total survey error in practice* (pp. 299–318). Hoboken, NJ: Wiley & Sons.

Cornesse, C. & Bosjnak, M. (2018). Is there an association between survey characteristics and representativeness? A meta-analyis. *Survey Research Methods*, *12*(1), 1–13.

Couper, M. (2000). Web surveys: a review of issues and approaches. *Public Opinion Quarterly*, *64*(4), 464–494.

Couper, M. (2011). The future of modes of data collection. *Public Opinion Quarterly*, *75*(5), 889–908.

Couper, M. (2013). Is the sky falling? New technology, changing media, and future of surveys. *Survey Research Methods*, *7*(3), 145–156.

Couper, M., Antoun, C., & Mavletova, A. (2017). Mobile web surveys. In P. Biemer, E. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. Lyberg, ... B. West (Eds.), *Total survey error in practice* (pp. 133–154). Hoboken, NJ: Wiley & Sons.

De Leeuw, E. (1992). *Data quality in mail, telephone and face to face surveys*. Amsterdam: TT-Publ.

de Heer, W. & de Leeuw, E. (1999). Surveying individuals and households; a passionate history. In J. Van Maarseveen & M. Gircour (Eds.), *A century of statistics: counting, accounting and recounting in the Netherlands* (pp. 45–65). Amsterdam: Stichting beheer IISG.

de Heer, W., de Leeuw, E., & van der Zouwen, J. (1999). Methodological issues in survey research a historical review. *BMS*, *64*, 25–48.

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

de Leeuw, E. (2011). Counting and measuring online: the quality of Internet surveys. *BMS, 114*, 68–78.

de Leeuw, E. & Berzelak, N. (2016). Survey mode or survey modes? In C. Wolf, D. Joye, T. Smith, & Y.-C. Fu (Eds.), *The sage handbook of survey methodology* (pp. 142–156). Los Angeles: Sage.

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

de Leeuw, E. & de Heer, W. (2002). Trends in household survey nonresponse: a longitudinal and international comparison. In R. Groves, D. Dillman, J. Eltinge, & R. Little (Eds.), *Survey nonresponse* (pp. 41–54). New York: Wiley.

de Leeuw, E., Dillman, D., & Hox, J. (2008). Mixed mode surveys: when and why. In E. de Leeuw, J. Hox, & D. Dillman (Eds.), *International handbook of survey methodology* (pp. 299–316). Psychology Press, EAM series. New York: Taylor and Francis.

de Leeuw, E. & Hox, J. (2011). Internet surveys as part of a mixed-mode design. In M. Das, P. Ester, & L. Kaczmirek (Eds.), *Social behavioral research and the Internet* (chapter 3). New York: Taylor and Francis group.

de Leeuw, E., Hox, J., & Boeve, A. (2015). Handling do-not-know answers. *Exploring new approaches in online and mixed-mode surveys. Social Science Computer Review, 34*, 116–132.

de Leeuw, E., Hox, J., & Kef, S. (2003). Computer-assisted self-interviewing tailored for special populations and topics. *Field Methods*, *15*, 223–251.

de Leeuw, E., Suzer-Gurtekin, Z., & Hox, J. (2018). The design and implementation of mixed mode surveys. In *Advances in comparative survey methodology* (chap. 18). Hoboken, NJ: Wiley & Sons.

de Leeuw, E. & Van der Zouwen, J. (1988). Data quality in telephone and face to face surveys: a comparative meta-analysis. In R. Groves, P. Biemer, L. Lyberg, J. Massey, W. N. II, & J. Waksberg (Eds.), *Telephone survey methodology*. Hoboken, NJ: Wiley & Sons.

Dickens, C. (1859). *A tale of two cities*. London: Chapman and Hall.

Dillman, D. (2000). *Mail and internet surveys*. New York: Wiley & Sons.

Dillman, D. (2017). The promise and challenge of pushing respondents to the Web in mixed-mode surveys. Survey Methodology, Statistics Canada, Catalogue No. 12-001-X, Vol. 43, No. 1. Retrieved from http://www.statcan.gc.ca/pub/12-001-x/2017001/article/14836-eng.htm

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

Dillman, D. & Edwards, M. (2016). Designing a mixed-mode survey. In C. Wolf, D. Joye, T. Smith, & Y.-C. Fu (Eds.), *The sage handbook of survey methodology* (pp. 255–268). Los Angeles: Sage.

Dillman, D., Smyth, J., & Christian, L. (2014). *Internet, phone, mail, and mixed-mode surveys; the Tailored Design Method*. New York: Wiley & Sons.

Dillman, D. & Tarnai, J. (1988). Administrative issues in mixed mode surveys. In R. M. Groves, P. Biemer, L. Lyberg, J. Massey, L. Nicholls II, & J. Waksberg (Eds.), *Telephone survey methodology* (pp. 509–528). New York: Wiley & Sons.

Fricker, S., Galesic, R., M. Tourangeau, & Yan, T. (2005). An experimental comparison of web and telephone surveys. *Public Opinion Quarterly*, *69*(1), 370–392.

Gordoni, G., Schmidt, P., & Gordoni, Y. (2012). Measurement invariance across face-to-face and telephone modes: the case of minority-status collectivistic-oriented groups. *International Journal of Public Opinion Research, 24*, 185–207.

Graham, J. (2012). *Missing data. Analysis and design*. New York: Springer.

Graham, J., Taylor, B., Olchowski, A., & Cumsille, P. (2006). Planned missing data designs in psychological research. *Psychological Methods, 11*, 323–343.

Groves, R. & Kahn, R. (1979). *Surveys by telephone: a national comparison with personal interviews.* New York: Academic Press.

Groves, R. & Petcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: a meta-analysis. *Public Opinion Quarterly*, *72*(2), 167–189.

Gwartney, P. (2007). *The telephone interviewers' handbook.* San Francisco: Jossey-Bas.

Harkness, J. (2008). Comparative survey research: goals and challenges. In E. de Leeuw, J. Hox, & D. Dillman (Eds.), *International handbook of survey methodology* (pp. 56–77). New York/London: Lawrence Erlbaum Associate.

Heerwegh, D. & Loosveldt, G. (2011). Assessing mode effects in a national crime victimization survey using structural equation models: social desirability bias and acquiescence. *Journal of Official Statistics, 27*, 49–63.

Hox, J. (2017). Computational social science methodology, anyone? *Methodology, 13*, 3–12. doi:10.1027/1614-2241/a000127

Hox, J., de Leeuw, E., & Klausch, T. (2017). Mixed mode research: issues in design and analysis. In P. Biemer, E. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. Lyberg, ... B. West (Eds.), *Total survey error in practice,* (pp. 511–530). Wiley Series in Survey Methodology. Hoboken, NJ: Wiley & Sons.

Internet World Stats. (2018). Internet world stats. Retrieved from http://www.internetworldstats.com/

Jaeckle, A., Lynn, P., & Burton, J. (2015). Going online with a face-to-face household panel: effects of a mixed mode design on item and unit non-response. *Survey Research Methods*, *9*(1), 57–70.

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

Kaczmirek, L. (2011). Attention and usability in internet surveys: effects of visual feedback in grid questions. In M. Das, P. Ester, & L. Kaczmirek (Eds.), *Social behavioral research and the internet* (pp. 191–214). New York: Taylor and Francis group.

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.

Kappelhof, J. (2017). Survey research and the quality of survey data among ethnic minorities. In P. Biemer, E. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. Lyberg, ... B. West (Eds.), *Total survey error in practice,* (pp. 235–252). Wiley Series in Survey Methodology. Hoboken, NJ: Wiley & Sons.

Klausch, T. (2014). Informed design of mixed-mode surveys: evaluating mode effects on measurement and selection error. Ph.D. Dissertation Utrecht University. Retrieved from https://dspace.library.uu.nl/handle/1874/300673

Klausch, T., Hox, J., & 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.

Kolenikov, S. & Kennedy, C. (2014). Evaluating three approaches to statistically adjust for mode effects. *Journal of Survey Statistics and Methodology, 2*, 126–158.

Kreuter, F., Presser, S., & Tourangeau, R. (2008). Social desirability bias in CATI, IVR, and Web surveys. *The effect of Mode and Question Sensitivity. Public Opinion Quarterly*, *72*(5), 847–865.

Luiten, A., de Leeuw, E., Hox, J., & Schouten, B. (2017). Response trends: the (new) international questionnaire on

nonresponse. Paper presented at the 7th International conference of the European Survey Research Association (ESRA 2017), Lisbon.

Lynn, P. (2017). From standardized to targeted survey procedure for tackling nonresponse and attrition. *Survey Research Methods*, *11*(1), 93–103.

Medway, R. & 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.

Messer, B. & Dillman, D. (2011). Surveying the general public over the internet using addressed based sampling and mail contact procedures. *Public Opinion Quarterly*, *75*(3), 429–457.

Mohorko, A., de Leeuw, E., & Hox, J. (2013). Internet coverage and coverage bias in europe: developments across countries and over time. *Journal of Official Statistics*, *29*(4), 1–15.

Olson, K., Smyth, J., & Wood, H. (2012). Does giving people their preferred survey mode actually increases participation rates? *Public Opinion Quarterly*, *76*(4), 611–635.

Peterson, G., Griffin, J., LaFrance, J., & Li, J. (2017). Smartphone participation in websurveys : choosing between the potential for coverage, nonresponse, and measurement error. In P. Biemer, E. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. Lyberg, . . . B. West (Eds.), *Total survey error in practice* (pp. 203–234). Hoboken, NJ: Wiley & Sons.

Roberts, A., Leeuw, E., Hox, J., Klausch, T., & De Jongh, A. (2013). Leuker kunnen we het wel maken. Online vragenlijst design: standard matrix of scrollmatrix? [We can put more fun in it: Online questionnaire design using an auto-advance carrousel matrix instead of a standard grid]. In A. Bronner, P. Dekker, E. de Leeuw, J. Paas, K. de Ruyter, A. Smidts, & J. Wieringa (Eds.), *Ontwikkelingen in het Marktonderzoek 2013. 38eJaarboek van de MOA Jaarboek 2012.* (pp. 133–148). Haarlem: SpaarenHout.

Roberts, C. (2007). Mixing modes of data collection in surveys: a methodological review. ESRC/NCRM Methods Review Paper # 008. Retrieved from http://eprints.ncrm.ac.uk/418/1/MethodsReviewPaperNCRM-008.pdf

Sakshaug, J. & Antoni, M. (2017). Errors in linking survey and administrative data. In P. Biemer, E. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. Lyberg, . . . B. West (Eds.), *Total survey error in practice* (pp. 557–573). Hoboken, NJ: Wiley & Sons.

Sakshaug, J. & Kreuter, F. (2012). Assessing the magnitude of nonconsent biases in linked survey and administrative data. *Survey Research Methods*, *6*(2), 113–122.

Scherpenzeel, A. (2017). Mixing online panel data collection with innovative methods. In S. Eifler & F. Faulbaum (Eds.), *Methodische Probleme von Mixed-Mode-Ansätzen in der Umfrageforschung.* Schriftenreihe der ASI - Arbeitsgemeinschaft Sozialwissenschaftlicher Institute. Wiesbaden: Springer VS.

Scherpenzeel, A. & Das, M. (2011). True longitudinal and probability-based internet panels: evidence from the Netherlands. In M. Das, P. Ester, & L. Kaczmirek (Eds.), *Social and behavioral research and the internet: advances in applied methods and research strategies* (pp. 77–104). New York: Routledge.

Singer, E. (2002). The use of incentives to reduce nonresponse in household surveys. In R. Groves, D. Dillman, J. Eltinge, & R. Little (Eds.), *Survey nonresponse* (pp. 163–178). New York: Wiley & Sons.

Singer, E. & Ye, C. (2013). The use and effects of incentives in surveys. *Annals of the American Academy of Political and Social Science, AAPS*, *645*(1), 112–141.

Smyth, J., Dillman, D., Christiian, L., & Stern, M. (2006). Comparing check-all and forced choice question formats in web surveys. *Public Opinion Quarterly*, *70*, 66–77.

Smyth, J., Olson, K., & Kasabian, A. (2014). The effect of answering in a preferred versus a non-preferred survey mode on measurement. *Survey Research Methods*, *8*(3), 137–152.

Sterrett, D., Malato, J., D. Benz, Tompson, T., & English, N. (2017). Assessing changes in coverage bias of web surveys in the United States. *Public Opinion Quarterly*, *81, special issue*, 338–356.

Stoop, I. (2005). *The hunt for the last respondent*. The Hague: SCP.

Sudman, S. & Bradburn, N. (1982). *Asking questions*. San Francisco: Jossey-Bass.

Tancreto, J., Zelenak, M., Davis, M., Ruiter, M., & Matthews, B. (2012). 2011 American Community Survey Internet tests: results from the first test in April 2011. 2012 American Community Survey Research and Evaluation Report Memorandum Series #ACS12-RER-13-R2, DSSD 2012 American Community Survey Memorandum Series #ACS12-MP-01-R2.

Tourangeau, R. (2017). Mixing modes: tradeoffs among coverage, nonresponse, and measurement error. In P. Biemer, E. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. Lyberg, . . . B. West (Eds.), *Total survey error in practice* (pp. 115–132). Hoboken, NJ: Wiley & Sons.

Tourangeau, R., Conrad, F., & Couper, M. (2013). *The science of web surveys*. New York: Oxford University Press.

Tourangeau, R. & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, *133*(5), 859–883.

Turner, C., Forsyth, B., O'Reilly, J., Cooley, P., Smith, T., Rogers, S., & Miller, H. (1998). Automated self-interview and the measurement of sensitive behavior. In M. Couper, R. Baker, J. Bethlehem, C. Clark., W. I. Nicholls, & J. O'Reilly (Eds.), *Computer-assisted survey information collection* (pp. 455–473). Hoboken, NJ: Wiley & Sons.

Vanniewuenhuyze, J. (2013). Mixed-mode data collection: basic concepts and analysis of mode effects. Retrieved from https://lirias.kuleuven.be/handle/123456789/399508

Wine, J., Cominole, M., Heuer, R., & Riccobono, J. (2006). Challenges of designing and implementing multimode instruments. Presented at the Second International Conference on Telephone Survey Methodology, Miami, FL, January 2006. Retrieved from https://www.rti.org/pubs/TSM2006_Wine_paper.pdf

Zhang, C. & Conrad, F. (2014). Speeding in the web:the tendcy to answer very fast and its association with straightlining. *Survey Research Methods*, *8*(2), 127–135.