Mobile Web Surveys: a First Look at Measurement, Nonresponse, and Coverage Errors

by

Christopher Antoun

**A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Survey Methodology)
in The University of Michigan
2015**

Doctoral Committee:

Professor Frederick G. Conrad, Chair
Professor Mick P. Couper
Assistant Professor Predrag Klasnja
Assistant Professor Brady T. West

DEDICATION

To Rachel and baby Rose

ACKNOWLEDGMENTS

I would never have been able to finish my dissertation without support and encouragement from many people. I would like to thank the members of my dissertation committee – Fred Conrad, Mick Couper, Brady West, and Predrag Klasnja – for their time and effort. I am grateful to the staff of CentERdata at Tilburg University in the Netherlands because this dissertation makes use of data obtained through the LISS panel administered by CentERdata and they carried out the experiment that is the foundation of this dissertation. In particular, I would like to thank Maarten Streefkerk who was my point of contact for the project.

I appreciate the generous financial support provided by Dr. James Morgan by way of a SEARCH Grant and Ms. Marilyn Lieberman by way of the Daniel Katz Dissertation Fellowship. These grants helped me focus on my dissertation research during this past year. I also appreciate the financial support provided by Ms. Jean Campbell through the Angus Campbell Graduate Fellowship.

I would like to thank my fellow students, including those how have moved on to other things, for their support, feedback, and friendship. I want to thank Kirsten Alcer for supervising me in my first research position at Survey Research Operations and for supporting my decision to enroll in the PhD program.

I would like to express my sincere thanks to Mick Couper who introduced me to the topic of mobile Web surveys and provided a two-semester guided reading course that helped shape this dissertation. Many (if not most) of the ideas in this dissertation came from our thought-provoking conversations, which always made me feel excited about this line of research.

I would like to express my deepest gratitude to my advisor and dissertation chair, Fred Conrad, for his guidance and encouragement. I was lucky to have his trust and support along the way. Despite his other commitments, he always made time for our weekly meetings (which I will greatly miss), interesting conversations about research ideas, and editing drafts of my work. Most importantly, he instilled in me a sense of confidence that I was up for task of producing original research, for which I cannot thank him enough.

Finally, I want to acknowledge the sacrifices made by my dear wife, Rachel, while I pursued this degree that are too numerous to describe. She was always there for me through the easy times and challenging times to ground me and motivate me to give my best effort.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ABSTRACT

This dissertation focuses on the use of smartphones for Web surveys. The current state of knowledge about whether respondents are willing and able to accurately record their answers when using such devices is evolving, but far from complete. The primary purpose of my research is therefore to investigate the implications of this new mode for various sources of error using a Total Survey Error (TSE) perspective.

Each chapter reports on a different aspect of a mode experiment that I designed to compare the effect of completion device (smartphone vs. computer) on survey errors. The experiment was carried out using the LISS panel (Longitudinal Internet Studies for the Social Sciences), a probability-based Web panel administered by CentERdata at Tilburg University in the Netherlands.

The first analysis (Chapter 2) compares response quality in the two modes. When using smartphones, respondents in this study really were more *mobile* and more engaged with the other people and other tasks compared to when using computers. Despite this, response quality – conscientious responding and disclosure of sensitive information – was equivalent between the two modes of data collection.

The second analysis (Chapter 3) investigates the causes of nonresponse in the mobile Web version of the experiment. I found that several social, psychological, attitudinal, and behavioral measures are associated with nonresponse. These include factors known to influence participation decisions in other survey modes such as

personality traits, civic engagement, and attitudes about surveys as well as factors that may be specific to this mode, including smartphone use, social media use, and smartphone e-mail use.

The third analysis (Chapter 4) estimates multiple sources of error simultaneously in the mobile Web version of the experiment. Errors are estimated as a mode effect against the conventional Web survey, which serves as the benchmark. I find few overall mode effects and no evidence whatsoever of measurement effects, but a significant impact of non-coverage bias for over one-third of the estimates.

Collectively, these findings suggest that non-observation errors (i.e., coverage and nonresponse), not measurement errors, are the largest obstacle to the adoption of mobile Web surveys for population-based inference.

Chapter 1: Introduction

Introduction

Researchers who use surveys to collect important information about society have long been interested in evaluating the quality of data they collect. The two sides of the Total Survey Error (TSE) paradigm emphasize different aspects of quality to be considered, including the quality of responses that are received (errors of measurement) and whether or not all sample members are observed (errors of representation). Along with other survey design features, the mode of data collection can impact these sources of error in a survey. There has been a rise in different modes of data collection over the past four decades, from mail and face-to-face surveys, to telephone surveys, and most recently to Web surveys. At each transition, researchers have tried to evaluate the quality of the data collected from the new mode relative to the traditional one (e.g., Groves and Kahn 1979; Couper 2000). The rise in smartphone ownership has led to the introduction of another mode: mobile Web surveys. My general interest in this dissertation is whether or not this mode decreases data quality relative to generic Web surveys.

The rise of Internet-enabled phones

In recent years a growing number of people – reportedly more than half of U.S. adults – have traded in their feature phones for larger multitouch smartphones that provide a better browser-based Web experience than earlier generations of mobile devices have done. According to the Pew Research Center, as of 2015 approximately

64% of U.S. adults owned a smartphone (up from 35% in 2011) and 89% of them used their phone to go online (Smith 2015). Moreover, 19% of U.S. adults rely on their phones for online access either because they have limited options for going online or because they have no landline Internet connection at home (Smith 2015). This transition to new technology has been rapid. Nancy Gibbs, writing in Time magazine, noted that “it is hard to think of any tool, any instrument, any object in history with which so many developed so close a relationship so quickly as we have with our phones” (Gibbs 2012 via Roe, Zhang, and Keating 2013).

Not surprisingly, the mobile revolution has affected Web surveys. According to AAPOR’s recent mobile technologies task force report (2014), “a non-ignorable and growing percentage of respondents are now accessing online surveys via their mobile browsers.” The percentage of surveys started on phones ranges from 6% to 43% in opt-in market research panels (Peterson et al. 2013; Kinesis 2013; Revilla et al. 2014), and from 2% to 25% in probability-based scientific panels (De Bruijne and Wijnant 2014; McGeeney 2015).

This change raises several questions related to the quality of data obtained from mobile Web surveys. For example, can respondents accurately record their answers when using a small screen? Who is willing to respond to such surveys? Do these surveys allow for inference to general populations? Careful evaluations of survey data quality in mobile Web surveys, which can inform approaches to reduce errors, are starting to be published but are still somewhat rare. Furthermore, most studies focus on only one source of error rather than multiple sources.

Specific aims

In this dissertation, I will investigate the impact of mobile Web surveys from a TSE perspective by focusing on three key error sources: *measurement*, *nonresponse*, and *coverage error*. Each source of survey error is the general focus of one chapter of this dissertation. The substantive chapters are organized as separate manuscripts.

This dissertation has three specific aims:

1) Compare the survey data generated by a mobile Web survey to that generated by a PC Web survey (*Chapter 2*).

2) Investigate the factors that influence participation when respondents are invited to complete a Web survey on a smartphone (*Chapter 3*).

3) Compare those with Internet access on their phones to those without access; then compare the resulting estimates of coverage errors to estimates of measurement and nonresponse errors (*Chapter 4*).

To address these aims, I designed a two-period crossover design experiment to compare the effect of completion device (smartphone vs. computer) on survey errors. I submitted this experiment as a research proposal to the CentERdata Institute for Data Collection at Tilburg University in the Netherlands. They accepted it and carried it out in 2013 using their LISS panel (Longitudinal Internet Studies for the Social Sciences), a probability-based Web panel in the Netherlands. The fact that the panel is comprised of a national probability sample instead of volunteers who self-select into the sample is important in order to support generalizations beyond my particular sample. To try to achieve full coverage of the panel, the LISS panel made extra efforts to include those without a smartphone by sending them an Android smartphone before the experiment.

Each dissertation chapter reports on a different aspect of this experiment. The experimental design will be described in more detail later on.

Defining "mobile Web"

As a starting point, it is important to define what *mobile* means and what it does not mean in this context. *Mobile* refers to the fact that respondents are using a wireless handheld device such as a smartphone or feature phone rather than a desktop or laptop computer to complete a survey, whether browser-based or app-based. Common phone models include feature phones (e.g., flip, bar, slider) and touchscreen smartphones with larger displays and more powerful processors (e.g., Android and iPhones) (Phonescoop 2012; Wikipedia 2013). *Mobile* does not refer to their connection, because mobile respondents can connect to either the mobile Internet via a cellular telephone network (e.g., 3G, 4G) or to the regular Internet via a wireless local area network (e.g., Wi-Fi). Additionally, the word *mobile* does not refer to the respondent herself who may actually be stationary, but rather the fact that handheld devices are easily carried and transported as part of ordinary use. PC Web, by contrast, refers to the fact that respondents are using desktop or laptop computers, which typically have larger screens, physical keyboards, a mouse or touchpad, more memory, and more computing power than phones[1]. In this context, then, the device used by the respondent (PC/mobile device) defines the mode (Web survey/mobile Web survey). I will use the term "mode" throughout this dissertation because it is part of the "lexicon of the survey research" (Couper 2011, p. 890), but it

[1] In everyday conversation, "PCs" almost always refer to Windows-based computers, whereas Apple computers are called "Macs" or "Macbooks". Since Apple computers are technically also *personal computers*, I refer to both types of computers as "PCs".

could be argued that only the device has changed between mobile Web and PC Web, not the mode.

Certainly, the line is blurring between smartphones and PCs, creating more of a continuum than a dichotomy. The newest generations of smartphones have relatively large screens (so-called phablets). Moreover, today's smartphones have more computing power than even the most powerful PCs of the past (or Apollo 11 when it landed on the moon for that matter [Gibbs 2012]). It goes both ways because PCs are also starting to look more like mobile devices. The newest laptop models are easily carried (and used away from home) and have touchscreen input in addition to physical keyboards. Also in the middle of the continuum are tablets, which are more similar to phones in some ways (touchscreen, easily transported) but more similar to PCs in other ways (screen size, computing power). There is still no label for Web surveys completed on tablets, and they are not a part of this research because early evidence suggests that data collected via PCs and tablets are similar (Guidry 2012; Lugtig and Toepoel 2015). Instead, my focus is on browser-based surveys accessed using smartphones and PCs. Next I will describe the recent empirical research on the topic and identify the gaps in knowledge that I aim to address with this dissertation.

Measurement on small devices

Early research focused on the usability issues related to screen size and touchscreen input that mobile respondents experience, particularly when completing unadjusted (or *passive*) Web questionnaires (Buskirk and Andrus 2012). These problems include the need to zoom in to select an answer and zoom back out to be able to see all response options, increased scrolling, reading text of small font-size, and the so-called “fat finger” problem of selecting an unintended response (Callagaro 2010; Peytchev and Hill 2010). Such issues, coupled with the fact that consumers now expect their favorite websites to be mobile-optimized (Online Publisher’s Association 2012), led many survey organizations to optimize their Web surveys (Macer 2012) in order to make them mobile-friendly, with features like large, touchscreen-friendly buttons for response options. This strategy is effective in some ways, as it seems to result in fewer breakoffs, shorter completion times, and higher survey satisfaction (Mavletova and Couper 2015; Baker-Prewitt 2013; Peterson et al. 2013). But mobile-optimized questionnaires still tend to produce more breakoffs and take longer to complete than PC Web versions of the same questionnaires (Mavletova and Couper 2015; De Bruijne and Wijnant 2013a; Mavletova 2013; Mavletova and Couper 2013; Wells, Bailey, and Link 2014).

Another wave of research has focused on the measurement properties of surveys that were optimized for mobile relative to PC Web surveys. A variety of different research designs have been used. These include observational studies that evaluate whether those who choose to use smartphones differ from PC Web respondents (e.g., Bosnjak et al. 2013a; Stapleton 2013; Toepoel and Lugtig 2014), classic split-ballot experiments to compare two data collection modes (e.g., Buskirk and Andrus 2014; De Bruijne and Wijnant 2013a; Mavletova 2013), and two-wave crossover experiments (Mavletova and Couper 2013). Researchers have also focused on a range of different indicators of response quality, such as response distributions (Buskirk and Andrus 2014; De Bruijne and Wijnant 2013a), the length of answers to open-ended questions (Wells et al. 2013; Peterson 2012; Mavletova 2013), the rate of selecting the same response for every item in a grid (“straightlining”) (McClain, Crawford, and Dugan 2012), and socially desirable responding (Mavletova and Couper 2013).

Results from these studies suggest that there are mode differences, but they are generally small. See Table 1.1 for a summary of these findings. The limitations of the work in this area are worth noting. Some studies are observational (marked with an “O”) wherein those who choose to use their phone may be different from those who choose to use their PC. Also, among the experimental studies where respondents are assigned to use either a smartphone or PC, some of them have uneven responses rates in their mode conditions (marked with “U” if the difference between the PC and mobile Web response rates is 10% or more), making it difficult to disentangle mode effects and selection effects.

Table 1.1: Summary of observed mode differences between mobile and PC Web

Study	Significant measurement differences
Buskirk and Andrus 2014[U]	Mode effect for question about smartphone apps
De Bruijne and Wijnant 2013a[U]	Mode effect for question about preferred device for going online
Lattery, Park Bartolone, and Saunders (2013)[O]; Stapleton 2013[O]	Mobile respondents less likely to select right-most (invisible) scale points
Mavletova 2013[U] ; Wells, Bailey and Link 2014; Peterson 2012[O];	Mobile respondents provide shorter answers to open-ended questions
Mavletova and Couper 2013[U]	Mobile respondents disclose less alcohol consumption than PC respondents
Peterson et al. 2013	Several mode effects for questions administered using either slider bars or spin wheels
Wells, Bailey, Link 2014	Mobile respondents provide shorter answers to open-ended questions

But technology tends to moves faster than science, and so other topics which are critical to understanding the implications of mobile Web surveys on response quality have received relatively little attention. One such topic is respondents’ use context. Smartphones are used in a variety of different settings (e.g., away from home, around other people, while multitasking). While it has been noted that researchers have less control over respondents' circumstances and surroundings (i.e., use context) in mobile Web than in traditional Web surveys, there has been little focus on how this on-the-go style of responding impacts response quality. For example, when filling out surveys on their phones, respondents may be more easily distracted and have less privacy. Another topic that has received little attention is how familiarity with smartphones may affect response quality. This is because most research in this area has been conducted on samples of smartphone owners without including people who are unfamiliar or uncomfortable using smartphones (who may have the most trouble recording their answers on such devices).

Chapter 2 overview

In Chapter 2, I will describe the potential reasons for differential measurement error between mobile Web and PC Web surveys. I will then use these consideratations to derive three hypotheses: when using smartphones, respondents will be 1) less conscientious because of increased multitasking and distractions; 2) less honest because they are more likely to be away from home and around other people; and 3) less accurate when recording their answers to certain questions on small smartphone screens. I will then test for differences in response quality using several indicators of least-effort responding, socially desirable responding, and input error. I will also investigate the effect of using an unfamilar phone on response quality.

Nonresponse in mobile Web surveys

By enabling respondents to complete surveys where it is convenient, mobile Web surveys could have higher participation rates than other modes of data collection. Yet, early research suggests just the opposite – those using mobile devices to access Web surveys have both significantly lower response rates and significantly lower completion rates (i.e., more breakoffs) than those using PCs. Figure 1.1 shows the breakoff rates across several studies. It is striking that all of these studies observed higher breakoff rates in mobile Web than PC Web, even for surveys that are optimized for small screens. Some researchers have suggested that Web questionnaires are simply more burdensome to fill out on a phone compared to a PC (Mavletova and Couper 2013).

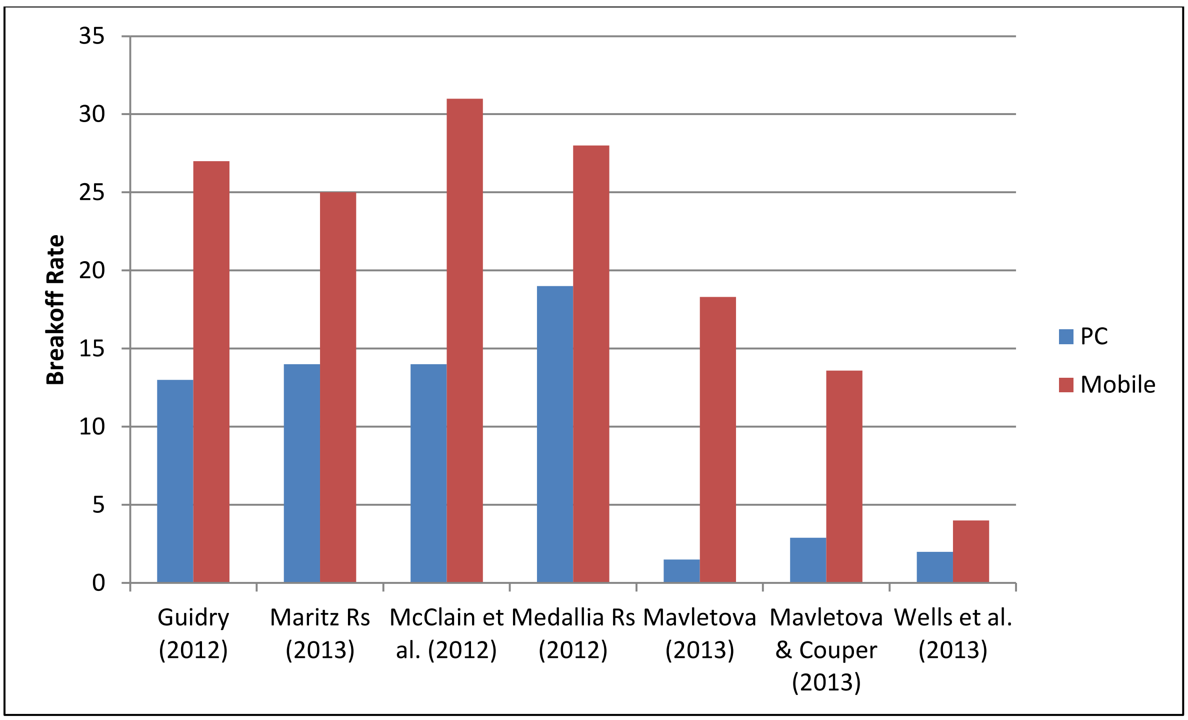


Figure 1.1: Breakoff rates across several mobile Web studies
Note: Source for "Maritz Rs" is Buskirk (2013).

The pattern for nonresponse looks similar to the pattern for breakoffs. As shown in Figure 1.2, several studies have observed lower unit response rates in mobile Web vs.

PC Web. For example, in an experimental study, Buskirk and Andrus (2014) reported that PC response rates were approximately three times higher than smartphone (iPhone) response rates (22.8% vs. 67.9%).

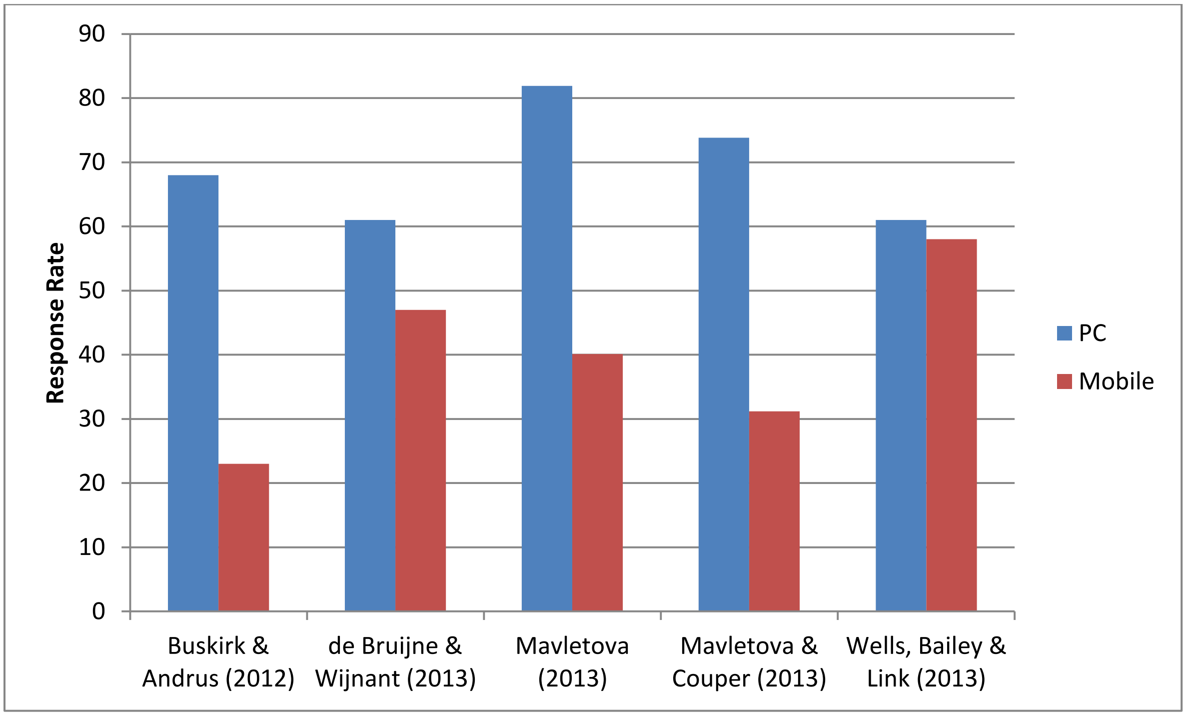


Figure 1.2: Response rates across several mobile Web studies

Unfortunately, little is known about the cause of this differential nonresponse. Early research suggests that those who are older, have less education (de Bruijne and Wijnant 2014), have lower incomes (Mavletova and Couper 2013) and have less trust in mobile Web surveys (Bosnjak, Metzger, and Gräf 2010) are less likely to participate in them.

Chapter 3 overview

In Chapter 3, I will review the potential causes of nonresponse in mobile Web surveys. I will separate the participation process in the mobile Web version of this study into two components: willingness to participate and survey response (conditional on willingness). I do this because different factors may affect each step. I will then model each of these steps using a range of social, psychological, attitudinal, and behavioral measures that were collected in prior panel waves.

Coverage error and total error in mobile Web surveys

To enhance survey measurement through the use of apps or other smartphone features, online panels are beginning to embark on mobile research by identifying a subsample of their members who are willing to use smartphones rather than computers, and then inviting them to participate using either their own phones or one that is provided by the panel. While these mobile-only surveys are not yet widespread, they soon could become more typical, especially in panels, because they can take advantage of the advanced features of smartphones. But it is still an open question whether such surveys allow for inference to general populations, given the fact that not all people have access to smartphones, and those who don't would be excluded from mobile-only research. As the recent AAPOR report (2014) put it:

> "Focusing on the widespread utility of mobile devices for data collection, there is still a question as to whether or not mobile is a niche methodology. It does appear to be a requirement in order to cover the increase in people taking online surveys via mobile devices and for specialty panels, but does it offer modes of collection robust enough for a general population survey? This remains to be seen."

Smartphone use is still far from universal, with coverage rates ranging from 18% to 60% in the U.S. and Europe, depending on the year and country (Fuchs and Busse 2009; Statistics Netherlands 2013). Moreover, the coverage rates vary across demographic groups. For example, Smith (2012) reports that those with mobile-access are younger, better educated, and more likely to be Black or Hispanic compared to those without access. Both Fuchs and Busse (2009) and Metzler and Fuchs (2014) report similar differences in several European countries between those with mobile Web access and those without it. But little is known about coverage errors for non-demographic variables or the size of such errors relative to errors from nonresponse or measurement.

Chapter 4 overview

In Chapter 4, I will evaluate multiple sources of error in a mobile Web survey. I will take advantage of the fact that respondents in the experiment being reported on were invited to respond using a smartphone and a PC at two subsequent points in time, and I will estimate total error in the mobile Web survey as a mode effect against the PC Web survey, which will serve as the benchmark. Because a balanced crossover design is used (rather than conducting the PC Web survey after the mobile Web survey for all respondents), the effects due to the time periods for conducting the surveys are not expected to contaminate the mode comparison. I will then decompose total error into its underlying components to explore the relative contributions of mode-specific noncoverage, nonresponse, and measurement errors for a range of non-demographic variables.

Conclusions

Survey completion via mobile devices is the new normal, according to several studies that have montiored device use among Web respondents (e.g., Kinesis 2013; Revilla et al. 2014). Researchers are grappling with this change in different ways. But strategies to limit mobile use are no longer viable. For example, efforts to block mobile users will exclude the small but growing numbers of mobile-only users (e.g., see Duggan and Smith 2013), and efforts to ask them to switch devices will likely lead to break-offs (McClain et al. 2012; Peterson 2012).

So while this change seems inevitable, survey researchers cannot assume that it will be innocuous. Instead, it is important for researchers to understand how the computing device used by respondents has an impact on survey errors. This way, survey researchers can realize the promise of mobile technology while producing high quality data.

Chapter 2: Effects of Mobile versus PC Web on Survey Response Quality: a Crossover Experiment in a Probability Web Panel

Summary

Survey participants are increasingly responding to Web surveys on their smartphones instead of their personal computers (PCs), and this change brings with it some potential challenges for them. Can they accurately record their answers when using a small screen and when distracted by walking or other tasks that could preoccupy them? The study reported here compares data quality in a conventional Web survey filled out on computers (PC Web) to a version of the same survey, simultaneously created for small screens, that respondents completed on smartphones (mobile Web). To reduce self-selection effects, a two-wave crossover design was applied to an existing probability-based panel sample, where participants (n = 1390) were invited to complete both a mobile and PC Web survey. I found that respondents in the mobile Web survey really were more *mobile* and more engaged with the other people and things around them, but this had little impact on the data quality indicators that I expected to be sensitive to respondents' context. This was evident in two ways. First, I found that despite the fact that respondents in mobile Web were more likely to multitask during the survey, they were at least as likely to provide conscientious, thoughtful answers as they did when responding on PCs. Additionally, despite the fact that respondents in mobile Web were

more likely to be around other people who could potentially look over their shoulder at their screens, they were as likely to disclose sensitive information. But while respondents' context of use did not have a large effect, screen size did. Respondents in mobile Web had trouble using their fingers to accurately move a small-sized slider bar and date picker wheel to the intended values. Overall, I find that people using smartphones can be careful and honest survey respondents, even when the context may be more distracting, as long as they are presented with question formats that are easy to use on small touchscreens.

Introduction

As respondents increasingly respond to surveys on their smartphones instead of their desktop and laptop computers, survey researchers have started to examine the measurement error implications of mobile Web use. The results from these assessments are likely to be a key factor in whether mobile Web surveys are widely adopted for social scientific research. If response quality is lower in mobile surveys relative to more widely-accepted methods like PC surveys, then researchers must find approaches to reduce measurement errors in mobile Web surveys or avoid using them altogether. But if response quality is as good in mobile Web surveys, then approaches for leveraging mobile use – such as encouraging mobile use in large-scale Web surveys or conducting mobile surveys designed to enhance measurement through the use of apps or other mobile-only features – become especially appealing. Hence, in considering how researchers should deal with mobile use, it is important to examine its potential impact on a respondent's ability to accurately report their answers.

There is a small but growing literature concerning the effects of respondents' mobile use on response quality. Several recent articles suggest that mobile Web surveys do not necessarily produce lower quality responses than PC Web surveys (Baker-Prewitt 2013; Bosnjak et al. 2013a; Gupta and Lee 2013; Lugtig and Toepoel 2015; Toepoel and Lugtig 2014; Zahariev, Ferneyhough, and Ryan 2009). However, others suggest that respondents using smartphones devote less effort to the survey task, either in the form of shorter answers to open-ended questions (Mavletova 2013, Wells et al. 2014) or selecting the first acceptable response option rather than considering the full set (primacy effects) (Lattery et al. 2013; Stapleton 2013). In addition, respondents using smartphones may be less willing to disclose sensitive information for some topics (Mavletova and Couper 2013).

Unfortunately, initial research in this area has encountered methodological limitations that may weaken causal claims about mobile use. In observational studies where respondents selected the device they used, it is unclear if these observed differences are due to device or to the characteristics of those who choose to use smartphones versus PCs. In experiments where respondents were randomly assigned to a device, relatively low response rates in the mobile condition increase the risk of non-equivalent experimental groups. Furthermore, most of the studies in this area were conducted using convenience samples of volunteers who may have an especially high level of comfort and familiarity with smartphones rather than probability samples, which are more likely to support generalizations beyond the sample.

In this chapter, I report on an experiment conducted to compare the effect of completion device (smartphone vs. computer) on survey reponse quality while limiting these threats to validity. To reduce self-selection effects, a two-wave crossover design was carried out in which participants were invited to complete both a mobile and PC Web survey (rather than just one survey). To study a diverse group of respondents, the sample consisted of participants already sampled to be part of a probability-based panel (rather than volunteers).

I examine input errors as well as several indicators of satisficing and socially desirable responding and their association with respondents' context. Specifically, I focus on the possibility that mobile respondents may be less conscientious because of increased multitasking and distractions; less honest because they are more likely to be away from home and around other people; and less accurate when recording their answers to certain questions on small smartphone screens. My focus is on browser-based surveys accessed using smartphones and PCs, and not tablets, because early evidence suggests that data collected via PCs and tablets are similar (Guidry 2012; Toepoel and Lugtig 2014). Before describing the experiment in more detail (in the Methods section), I will first outline the reasons why one might expect data quality to be lower when respondents use smartphones instead of PCs and further describe my research hypotheses.

Background

There are several different ways that smartphones might affect response quality in Web surveys. I highlight eight of them in Table 2.1. These considerations fall under three groups. First, smartphones have different *technical features* than computers. This includes such features as screens that are relatively small in size, touchscreen user input, and connection via a cellular Internet connection (when not connected to WiFi) that may vary in consistency and speed. Second, smartphones are used in more diverse *use* *contexts* than PCs (Cui and Roto 2008; Lee, Kim, and Kim 2005). Mobile respondents may actually be *mobile* (on-the-go), in different physical locations (away from home), and in different social situations (around other people). They may be at risk of having their attention taken away by multitasking (dividing attention between tasks) or distractions (being forced to switch between tasks). Third, data collected with smartphones may be affected by the *characteristics of their users* who are operating them. Respondents in mobile Web must have the fine-motor skills to precisely tap, drag, or spin a small target on a touchscreen. They must also know how to use mobile browsers and know how to record their answers using question formats that are unique to such browsers (e.g., date pickers). In the right-hand column of Table 1, I list my expectations for how these factors could compromise data quality in mobile Web surveys.

Table 2.1: Factors that could differ between PC Web and mobile Web survey modes and the implications for data quality in mobile Web

Factor	**Traditional Web**	**Mobile Web**	**Implications for data quality in mobile Web surveys**
Technical features			
Screen Size	Larger	Smaller	More difficult to read text and record intended answers because of small font size
User input	Physical keyboard and mouse	Touchscreen keyboard	More burdensome to type answers
Connection quality	Stronger, more consistent	Could be weaker, more intermittent	Increased breakoffs
Use Context			
Distractions/multitasking	Less common	Could be more common (e.g., ambient noise)	Reduced attention and effort
Presence of others	Less likely	Could be more likely	Less willingness to disclose embarrassing information
Environmental cues	Provided mostly by home environment	Provided by home and mobile (away from home) environment	Environment may have priming effect (i.e., won't give sufficient attention to unprimed info)
Necessary User Characteristics			
Fine-motor skills	Precise mouse movements	Precise touchscreen gestures	Increased input errors
Familiarity with commonly used input tools	Radio buttons, check boxes, etc.	Pickers, sliders, drag and drop approaches, etc.	Increased item missing data or breakoffs

These distinctions allow me to derive several hypotheses. One hypothesis is that respondents' *use context* will affect the attention and effort they devote to the survey task. Specifically increased distractions and multitasking will negatively affect effort. There is evidence that secondary tasks reduce the quality of task performance because they compete for mental resources. In a study of people who were asked to divide their attention while walking, Hyman and colleagues (2009) found that those who were walking while talking on a cell phone were less likely to notice unusual activities during their route that should have been quite easy to see (specifically, a unicycling clown).

Respondents who perform secondary tasks while completing a survey may also divide their attention. As a result, they may rely on a least-effort response strategy or *satisfice* instead of optimizing and performing all of the cognitive steps required to carefully answer a survey question (Tourangeau, Rips, and Rasinski 2000). The theory of satisficing, which originated from Simon's (1956, 1957) work on decision making and has since been generalized to surveys (Krosnick and Alwin 1987; Krosnick 1991), posits that respondents may provide simply satisfactory answers instead of optimal answers in order to minimize cognitive effort.

Given the evidence that multitasking has the potential to lead to increased reliance on cognitive shortcuts in PC Web surveys (Zwarun and Hall 2014) and telephone surveys (Lavrakas et al. 2010, Lynn and Kaminska 2012, Kennedy 2010), I would also expect a similar effect in mobile Web surveys. To the extent that increased multitasking occurs in mobile Web as compared to PC Web, it follows that when using smartphones, respondents may be less conscientious. In addition, respondents who are unfamiliar with smartphones will satisfice more if the mental resources required for using an unfamiliar phone diminish their processing ability.

The literature in this area is inconclusive. Initial mobile Web studies about satisficing have reached different conclusions. To illustrate this problem, consider respondent nondifferentiation or "straightlining" (the tendency to select the same response option for every item on a page). In studies in which respondents are randomized to device, researchers have found no significant device differences (Baker-Prewitt 2013; Lugtig and Toepoel 2015; Peterson et al. 2013), but in observational studies where respondents selected the device they used, researchers have found evidence of more straightlining behavior in mobile Web compared to PC Web (Guidry 2012; McClain et al. 2012). It is unclear if these differences are due to device or to differences between those who choose to use smartphones versus PCs.

Researchers have also found mixed results when evaluating the quality of answers to open-ended questions, another indicator of satisficing. For example, in an early mobile Web study, Peytchev and Hill (2010) examined a half-open question containing both closed-ended options and an open-ended "other" option. They found that mobile respondents tended to avoid choosing the ''other'' category that required typing. But in a more recent experiment using this question format, Wells et al. (2014) reported that mobile respondents chose the "other" category at the same rate as PC respondents. Similarly, when investigating the length of answers to open-ended questions, some have found that mobile respondents type shorter answers than PC respondents (Mavletova 2013, Peterson et al. 2013, and Wells et al. 2014); others have found no such device differences (Bosnjak et al. 2013a; Buskirk and Andrus 2014; Lugtig and Toepoel 2015; Toepoel and Lugtig 2014; Zahariev et al. 2009).

My second hypothesis focuses on the impact of respondents' setting as they complete surveys. I expect that being away from home and around other people will lead them to disclose less sensitive information. It is known that respondents may edit their answers in order to avoid revealing information that is embarrassing or threatening if others were to find out about it (e.g., Kreuter, Presser, and Tourangeau 2008; Tourangeau and Smith 1996). There is also evidence that the variation in interview privacy, i.e., whether third-parties are present, can affect honest reporting even in self-administered surveys. For example, in a computerized and paper-and-pencil self-administered survey of adolescents and young adults, Aquilino, Wright, and Supple (2000), found that respondents were less likely to disclose alcohol and marijuana use when parents or siblings were present. When using smartphones, respondents may be concerned not about a third-party overhearing their interview but about a third-party's prying eyes looking at their device screens (so-called "shoulder surfing"). Or, an awareness of the mere presence and close proximity of others may heighten the sensitivity of some questions. Thus to the extent that this occurs and that it diminishes respondents' sense of privacy, I expect them to disclose less sensitive information in mobile Web than PC Web (at least when others are present).

As with mobile Web studies that have investigated satisficing behaviors, there are some conflicting findings for socially desirable responding. Using a Web panel in Russia, Mavletova (2013) evaluated 16 questions with more or less socially desirable answers and found no overall difference in disclosure. Likewise, no significant device differences were reported for a customer satisfaction survey (Gupta and Lee 2013) or to an attitude question about one's own personal finances (Zahariev et al. 2009). But in a crossover experiment using a Web panel in Russia, Mavletova and Couper (2013) found lower levels of reporting of alcohol consumption and household income in mobile Web than PC Web (though they found no differences for their other three groups of measures). All of these studies relied on volunteer samples where the effect of device on socially desirable responding may be different from that for non-volunteers (e.g., if volunteers are more conformable disclosing sensitive information than non-volunteers).

A third hypothesis that I will test is that input accuracy will be lower in mobile Web than PC Web for certain question formats such as slider bars and pickers (sometimes called "spin wheels" or "spinners"). Here I focus on two types of input error. The first type -- motor-precision errors -- assumes that respondents know how to use the input tools but lack the motor control necessary to record an accurate answer. These errors are related to the well-known "fat finger" problem of accidently selecting the wrong target on touchscreen keypads (e.g., Bi, Li, and Zhai 2013). One challenge to recoding accurate answers on smartphone is touchscreen input. For some computer tasks there is evidence that motor performance is slower and more difficult for direct touch compared to mouse input (Forlines et al. 2007), especially for older adults who have reduced dexterity (Wood et al. 2005); this could be one reason why older adults are less likely to use mobile devices. Another challenge to recoding accurate answers on smartphones is screen size. Some question formats (e.g., slider bars, grids) are quite small when displayed on a smartphone screen, and trying to interact with a target on a screen that covers less screen space is harder according to Fitts' law, a longstanding human-computer interaction principle (Card, English, and Burr 1978). This rule states that a target's size and distance from a user determines how difficult it is to touch, other things being equal. While difficulty is commonly framed in terms of the time required to complete the task assuming error-free execution, difficulty could instead be framed in terms of accuracy under the assumption that response time is fixed. For example, in PC Web surveys there is evidence that shorter computerized Visual analog scales (VASs) (width is typically about 50 pixels) result in about 27% larger input errors, defined as the distance between the recorded answer and the intended value, than longer ones (200 pixels and 800 pixels) (Reips and Funke 2008).

The other type of input error occurs when respondents do not know how to use an input tool to begin with. These so-called knowledge-based mistakes (Norman 1981; Reason 1990) would occur among respondents who are unfamiliar with smartphones and fail to recognize the affordances of a particular input tool – e.g., that a picker affords spinning. If their troubleshooting efforts fail, and clear directions are not provided, they may be forced to skip the question entirely. By contrast, respondents who are familiar with smartphone devices probably know the touchscreen gestures required to use such tools.

Few experiments have evaluated input widgets in mobile Web. In a comparison of answers between mobile Web and PC respondents for a series of slider bar questions, Buskirk and Andrus (2014) found no significant differences. They also compared the reported number of cell phones owned via a single-wheel picker in mobile Web and a drop box in PC web and found no significant differences. Peterson et al. (2013), by contrast, found that the responses recorded using slider bars in mobile Web differed significantly from responses recorded on grids in PC Web for 10 of 15 questions, though not necessarily in a discernable way. They also report significant differences between responses recorded using spin wheels in mobile Web and responses recorded on grids in PC Web for 6 of 15 questions.

In sum, initial empirical work supports the possibility that mobile administration could either harm data quality or could have no adverse effects. It is important to investigate this topic to determine if the findings from earlier research are replicable in a general population sample using an experimental design that separates measurement effects from selection effects.

Methods

Experimental design

A special experimental design was needed given the consistent problem of uneven response rates between the mobile and PC Web conditions. Following Mavletova and Couper (2013), I opted for a crossover design (see Johnson 2010) where the same participants were invited to complete two surveys in sequence, once using their PC and once using a smartphone (see Figure 2.1). This way each respondent who completed both surveys could serve as his or her own control, making group differences less impactful. Participants were randomized to one of two sequences of modes (mobile Web first or PC Web first). As other survey researchers have done when using a crossover design (e.g., Gupta and Lee 2013; Mavletova and Couper 2013), I included a month-long “wash-out” period between the waves. This time period was presumed to be long enough to minimize the potential effect of answering questions in the first period persisting into the second period, but short enough to minimize the likelihood that there were substantial real changes (e.g., lost employment, got married) between surveys.

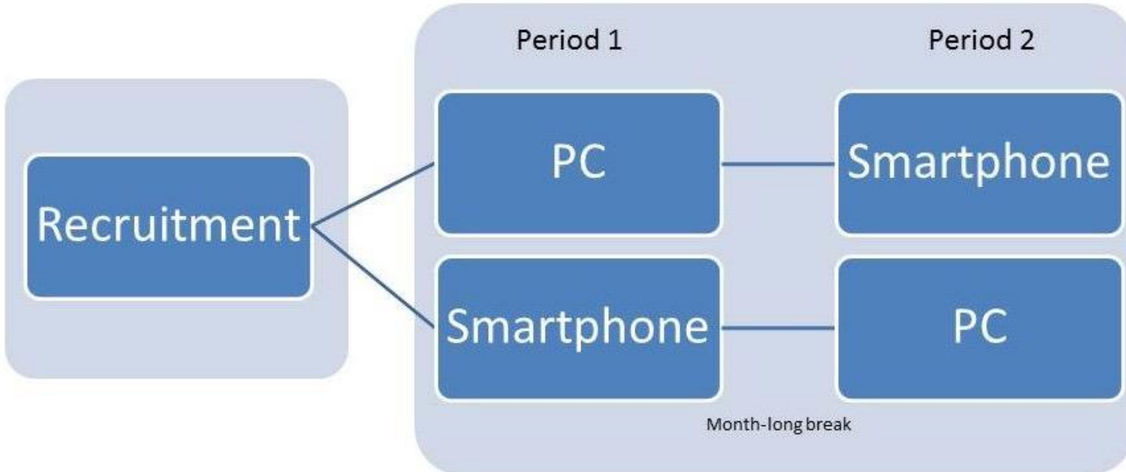


Figure 2.1: Design of mode experiment comparing mobile Web and PC Web

Questionnaire optimization

The same 46 questions were asked in both modes and in both periods. They were programmed using Blaise survey software (Statistics Netherlands). The mobile version was programmed to be optimized for small devices using a special C-Moto stylesheet (De Bruijne and Wijnant 2013b). Both questionnaires used a paging design with wide navigation buttons (for *back* and *next*) presented at the bottom of each page. Examples of the questionnaire layouts are presented in Figure 2.2[2].

[2] There was a LISS panel logo and Tilburg University logo in PC version that was not included in the mobile version. All panelists already know the organizations involved in this Web panel, so there is no reason to expect that this difference in branding might matter.

2.2a: Mobile Web questionnaire (question 1)

2.2b: PC Web questionnaire (question 1)

Figure 2.2: Examples of questionnaire layout

A restriction for the experiment was that participants were asked to use either an iPhone or Android device. The mobile questionnaire was designed for and tested on these devices in order to standardize its layout to the extent possible. To try to achieve full coverage of the panel, the LISS panel made extra efforts to include those without the necessary phone. These panel members were sent a loaned Android smartphone (Samsung Gio) approximately one week before the mobile Web survey.

Sample

All participants were LISS panel members. The panel consists of nearly 8,000 individuals in the Netherlands age 16 and older who are invited to take Web surveys every month. Its original members were selected in 2007 based on a probability sample of Dutch-speaking households from the Netherlands population register. Selected members were recruited by mail, phone, and in-person visits, and those without Internet access were loaned equipment to provide access (for more information about LISS, see Scherpenzeel 2011).

Recruitment

Recruitment for this experiment was carried out from July 1-August 26, 2013. Among the 5486 panel members who completed the recruitment and screening questionnaire, 2250 indicated they were willing to participate (n = 1389 iPhone or Android users; n = 861 users of other phones and non-users); a small number of these willing panelists (n = 38) were considered to be ineligible because of a programming error. Due to a limited number of borrowed phones, iPhone and Android phone owners were oversampled (n = 990 iPhone or Android users; n = 400 users of other phones and non-users). The selected cases were then randomly assigned to a sequence, either mobile Web first (n = 695: 495 with own phones and 200 with borrowed phones) or PC Web first (n = 695: 495 with own phones and 200 with borrowed phones). Cases were invited to participate in the second survey in the sequence, regardless of whether they completed the first survey.

Data collection

Data collection was carried out from October 7-October 29, 2013 for period 1 and from December 2-December 31, 2013 for period 2. Invitations to the surveys were sent by e-mail. Participants using a loaned phone had the option to automatically log in to the survey by selecting a bookmark on their phone's home screen (containing an encrypted version of their login credentials)[3].

Two reminder emails were sent to nonrespondents near the end of each month of data collection. The normal cash incentive for the LISS panel (15 Euros per hour of time spent completing surveys) was provided as payment for participation in the experiment.

The completion rates by mode and by period are shown in Table 2.2. Those participants who self-selected to complete the Web questionnaire in a different mode than they were assigned (i.e., *treatment crossovers*) were noncompliant and counted as nonrespondents[4]. Device information was extracted from browser log files (user agent strings). 25.4% of respondents used more than one device. When this occurred I assume that they switched devices early on in the questionnaire, and I use information about the browser that was used last by them. This last-session strategy has been used by other researchers when analyzing user agent strings (De Bruijne and Wijnant 2013a; Toepoel and Lugtig 2014). As one might expect for a PC Web panel the level of unassigned device use was higher in mobile Web (11.9%) than in PC Web (3.9%). As a result the

[3] This was not the first mobile Web experiment conducted in the panel, so some of these panelists may have been asked to participate in other mobile surveys.
[4] I did not do an intent-to-treat analysis because I am interested in the effect of mode as received, not as assigned.

overall response rates (AAPOR response rate 2, see AAPOR 2011) were lower in mobile Web (73.4%) than PC Web (81.4%), and this was the case for both periods (bolded in Table 2). The number of breakoffs was quite small in both modes (mobile Web: n = 12; PC Web: n = 4).

Table 2.2: Completion rates in mobile Web and PC Web by period

	Invitations by mode			
	Mobile Web		**PC Web**	
	n	**%**	**n**	**%**
Period 1				
Number of invitations	695		695	
Starts: all devices	598	86.0	623	89.6
Starts: assigned device	**522**	**75.1**	**594**	**85.5**
Completes: assigned device	517	74.4	592	85.2
Period 2				
Number of invitations	689[a]		695	
Starts: all devices	582	84.5	562	80.9
Starts: assigned device	**494**	**71.7**	**537**	**77.3**
Completes: assigned device	487	70.7	535	77.0
Periods 1 and 2 combined				
Number of invitations	1384		1390	
Starts: all devices	1180	85.3	1185	85.3
Starts: assigned device	**1016**	**73.4**	**1131**	**81.4**
Completes: assigned device	1004	72.5	1127	81.1
	Unique participants			
	n		**%**	
Number of participants invited	1390			
Number who participated…	1262		90.8	
…in both surveys	895		64.3	
…in mobile Web only	129		9.3	
…in PC Web only	238		17.1	

[a] Six panel members dropped out of the experiment after period 1.

Respondent demographics

In total, 1262 panelists started one or more of the surveys using the assigned device. This sample of participants was nearly identical to the full LISS panel in terms of age (median age: 44.0 vs. 43.0 years old in the LISS panel) and gender (percent male: 49.1% vs. 49.1%), but those in the sample of participants were relatively well educated compared to the full panel (college degree and above: 38.7% vs. 26.3%).

It is unlikely that nonresponse affected the comparability of groups because most of the invited panelists participated in both surveys and could serve as their own controls. To be sure, I checked this for demographic characteristics. According to ordinal logistic mixed models that were used to account for the overlap between samples, those participants in the mobile condition and those participants in the PC Web condition did not differ reliably in terms of age, gender, or education (Table 2.3).

Table 2.3: Demographics of participants in mobile Web and PC Web

	Mobile Web sample (n=1016)	PC Web sample (n=1131)	F Test (Type III)
Characteristic	%	%	
Age			F (1, 885) = 0.27, *p* = n.s.
16-29	22.5	23.6	
30-44	27.6	25.7	
45-59	27.3	27.7	
60-87	22.6	23.0	
Gender(=male)	49.9	49.8	F (1, 887) = 0.05, *p* = n.s.
Education			F (1, 882)[a] = 0.07, *p* = n.s.
High School or less	34.9	36.2	
Vocational/Junior College	25.3	25.0	
College and above	39.8	38.7	

[a]9 cases had missing values for education

Questionnaire

The main section of the instrument contained 33 items. The supplementary section of the questionnaire contained 7 debriefing questions that were asked to allow measurement of the respondent's setting (e.g., location, presence of other people, how many other tasks the respondent did while completing the questionnaire) and 6 standard LISS debriefing questions that were asked about the questionnaire itself (e.g., satisfaction with survey). The questionnaire was in Dutch. The wording for the complete set of survey questions (translated to English) is provided in Appendix A.

The median completion time was longer in mobile Web (17.4 minutes) than PC Web (9.8 minutes). This difference might reflect usability problems, network latency, different levels of respondent engagement, or any combination of these factors. Comparable time differences have been reported in other studies (De Bruijne and Wijnant 2013a; Mavletova 2013; Mavletova and Couper 2013; Wells et al. 2014).

Response quality indicators

Data quality was evaluated in several ways. As described in Table 2.4, I used five different indicators of satisficing (i.e., providing adequate but not optimal responses): rounded numerical responses, non-differentiation to political attitude items, incorrect answers to a cognitive reflection test (CRT), short answers to an open-ended question, and non-selection of the “other” option for a half-open question. For each measure, I assume that larger means reflect increased tendency to satisfice.

Table 2.4: Indicators of satisficing

Measure	Summary	Description/Justification
Rounding	This is the number of rounded answers (divisible by 10) provided to eight questions. The questions consist of two behavioral frequency questions Q2 and Q3 (adapted from Toepoel and Couper 2011 and Schober et al. 2012) and six political attitude questions using a feeling thermometer scale ranging from 0 (“very unfavorable) to 100 (“very favorable”) (Q12a-Q12f).	For the behavioral frequency questions, it is assumed that respondents who provide rounded answers have a greater tendency to use impression-based estimation, rather than recall-and-count strategies, as a mental shortcut (Conrad, Brown, and Cashman 1998; Manski and Molinari 2010). For the political attitude questions, it is assumed that respondents who provided rounded answers, using only part of the scale (i.e., heaping) rather than the full 101-point scale (Holbrook et al. 2014). The two types of questions are combined for an overall rounding measure, but the results are unchanged when the two are analyzed separately.
Non-differentiation	This is the degree to which respondents failed to differentiate between political attitude items with their answers. It was measured using numerical responses to the six political attitude questions.	The computation of nondifferentiation scores followed a series of steps. I first recoded the ratings to range from 0 to 1 (instead of 0 to 100) and took the mean of the root of the absolute differences between all 15 pairs of items in the battery; the resulting scores ranged from 0 to 0.659. To produce an index where higher scores indicated more straightlining, I then rescaled the score using a method developed by Chang and Krosnick (2009). I subtracted the value of the largest score (0.659) from each score and then divided the resulting scores by the inverse of the largest score (-0.659). This produced scores that ranged from 0 (indicating the lowest observed level of nondifferentiation) to 1 (indicating that a respondent straightlined and gave the exact same answer to all 6 questions in the battery).
Incorrect answers to cognitive reflection test	This is the number of incorrect answers to a three-item cognitive reflection test (CRT) developed by Frederick (2005).	According to Frederick (2005), CRT measures the ability to resist reporting the impulsive answer that springs quickly to mind and instead report the correct answer that comes slowly to mind after careful deliberation. They have been shown to predict performance on a wide range of tasks used in social psychological research (see e.g., Toplak, West, and Stanovich 2011). Consider the first test item: “A bat and a ball cost $1.10. The bat costs $1.00 more than the ball. How much does the ball cost?” Here the impulsive answer is “10 cents,” but the difference between 10 cents and $1.00 is only 90 cents, not $1.00; so the correct response is “5 cents”. A tendency toward satisficing should result in more incorrect answers to CRT questions.
Short open-ended response	This is a binary measure of whether respondents typed one word or less in	The target of the question was plural, as in one’s “hobbies,” and not one’s singular hobby. It is assumed that respondents who provide a one-word answer or no answer at all have a

	response to an open-ended question about their hobbies.	greater tendency to satisfice.
Non-selection of the “other” option for a half-open question	This is a binary measure of whether the “other” category was selected for a question (Q8). The question (“What is your favorite vegetable?) contained 5 closed-ended response options (beans, broccoli, kale, carrots, and spinach) and an “other specify” box.	If respondents did not prefer one of the five vegetables listed, they were faced with the decision to either type their answer (requiring more effort) or select one of the other closed-ended options instead (requiring less effort).

I also used six indicators of socially desirable responding. They consist of five individual items [adapted from Lynn and Kaminska (2012) and Schober et al. (2012)] and one summary measure of the total count of socially undesirable answers. I make the common assumption that more socially undesirable answers reflect more honest reporting (infrequent exercise, binge drinking, driving while intoxicated, negative views about immigrants, and frequent TV viewing), which has been supported by a record-check study (Kreuter et al. 2008).

Finally, I used two interactive question formats that use finger or stylus touch: a slider bar that respondents touch and drag along a semantic dimension; and a date picker that respondents touch and spin vertically. So that reporting errors do not just add undetected noise, I compare responses to a verifiable, objective measure. I asked respondents to record their age using the slider scale and birth year using the date picker. The benchmark values are obtained from the LISS data archive[5], and deviations from them are regarded as error.

Data analysis

The results are based on comparisons of the full set of mobile-Web cases (n = 1016) to the full set of PC-Web cases (n = 1131). Mixed-effects models and marginal models were used instead of repeated-measure ANOVA so that respondents with missing values were not dropped from the analysis altogether (see West et al. 2014). For continuous measures, linear mixed models (LMM) were used with a random effect of respondent. They were estimated in SAS using proc mixed. For the models of the response quality indicators, I assessed model diagnostics by using quantile–quantile plots

[5] This background information is collected through a separate survey and can be updated by the household contact person every month.

and distribution plots both for the random slopes and for the conditional residuals. All model assumptions were adequately met. As an example, see Tables B.1-B.4 in Appendix B for the model diagnostics for the disclosure model that will be presented later on (in Table 2.7).

The model for a continuous measure indexed by *t* (*time period* = 1, 2) on the *i*-th respondent is

$$Y_{ti} = (\hat{\beta}_0 + b_{0i}) + \hat{\beta}_1 Mode_{ti} + \hat{\beta}_2 Period_{ti} + \hat{\beta}_3 Seq_i + \varepsilon_{ti}$$

where

$\hat{\beta}_1$ = effect estimate of survey mode (mobile vs. PC Web);
$\hat{\beta}_2$ = effect estimate of time period (period 1 vs. period 2);
$\hat{\beta}_3$ = effect estimate of sequence (mobile-PC vs. PC-mobile);
b_{0i} = random respondent effect; and
ε_{ti} = error term

For categorical measures, marginal logistic models estimated using GEE (Generalized Estimating Equations) were used to describe the effect of each predictor, averaged across all participants. They were estimated in SAS using proc genmod (with a REPEATED statement). A compound-symmetric (exchangeable) structure was specified for the structure of the correlation among responses within subjects[6]. Robust standard errors were used rather than model-based ones so that the standard errors are less sensitive to misspecification of the correlation structure. For the models of the response quality indicators, I assessed model diagnostics by using Pearson standardized residual distribution plots as a function of the predictors. No violations of model assumptions were detected. As an example, see Tables B.5-B.10 in Appendix B for the residual plots

[6] When the unstructured and compound-symmetric correlation structures were applied, the quasi-likelihood information criteria for model fit QIC and QICu (Pan 2001) were nearly identical for the two correlations structures.

both from a model predicting short open-ended responses and from a model predicting the admission of ever driving while intoxicated that will be presented later on (in Tables 2.5 and 2.7, respectively). In addition, I correlated the predicted values (probabilities) with the actual (binary) values. These correlations ranged from .01 to .43 suggesting at least reasonable correspondence between predicted and observed values for some models, but a general lack of predictive power of the independent variables (survey mode, period, and sequence) for other models.

The model for a binary measure indexed by *t* (*time period* = 1, 2) on the *i*-th respondent is

$$logit(\Pr(Y_{ti} = 1)) = \hat{\beta}_0 + \hat{\beta}_1 Mode_{ti} + \hat{\beta}_2 Period_{ti} + \hat{\beta}_3 Seq_i$$

where

$\hat{\beta}_1$ = effect estimate of survey mode (mobile vs. PC Web) averaged across all participants;
$\hat{\beta}_2$ = effect estimate of time period (period 1 vs. period 2) averaged across all participants;
$\hat{\beta}_3$ = effect estimate of sequence (mobile-PC vs. PC-mobile) averaged across all participants; and
$\hat{\beta}_0$ = intercept term

Least squares (LS) means were estimated from these models using the ESTIMATE statement (for binary measures, LS means are reported on the probability scale). For all measures, a test of carryover was conducted based on the sequence effect; there was no evidence of a carryover effect from period 1 to period 2, which suggests that the inclusion of the month-long washout period worked as planned.

Results

Setting where Web questionnaires were completed

Before evaluating response quality, I consider the context in which surveys were completed in mobile Web relative to PC Web because it is perceived that these differences will impact response quality.

Respondents to the mobile Web survey really were more *mobile* than PC Web respondents, apparently taking advantage of their smartphone's small size, light weight, and mobile connectivity. This is evident in several ways. When using smartphones, respondents were more than four times as likely to report that they were away from home or work when completing the survey (7.0%) than when using computers (1.6%) (*OR* 4.91, *95% CI* 2.80-8.61, $p < .001$). When using smartphones, more respondents reported that they walked or moved around when completing the survey (11.9%) than when using PCs (8.9%) (*OR* 1.39, *95% CI* 1.04-1.85, p=.027). In addition, more respondents in mobile Web reported that they traveled between two different locations when completing the survey (3.1%) than in PC Web (0.2%) (*OR* 16.74, *95% CI* 3.96-70.76, $p < .001$). Despite these differences in mobility, the majority of mobile Web respondents (85.3%) reported being at home and nowhere else when completing the survey.

Respondents to the mobile Web survey were not only more *mobile*, but also more likely to divide their attention between other tasks. When using smartphones, respondent were more likely to report having multitasked while completing the survey (54.3%) than when using PCs (44.4%) (*OR* 1.61, *95% CI* 1.32-1.95, $p < .001$). As shown in Figure 2.3, this result was consistent across different levels of multitasking: mobile Web respondents were more likely to have completed one other task (30% vs. 25%), two other tasks (17%

vs. 14%), three other tasks (5% vs. 4%), and four or more other tasks (2% vs. 1%), respectively. When using smartphones, marginally more respondents in mobile Web reported being distracted by the things going on around them (38.8%) than when using computers (35.2%; *OR* 1.23, *95% CI* 1.00-1.53, $p = .054$). These distractions may have caused some respondents to take breaks from the survey; when using smartphones, a substantial percentage of respondents used more than one session (39.7%), which was not the case when using PCs (10.3%). In addition, respondents were more likely to report that they were around other people while filling out the questionnaire using smartphones (37.4%) than when doing so using PCs (29.9%) (*OR* 1.55, *95% CI* 1.25-1.93, $p < .001$).

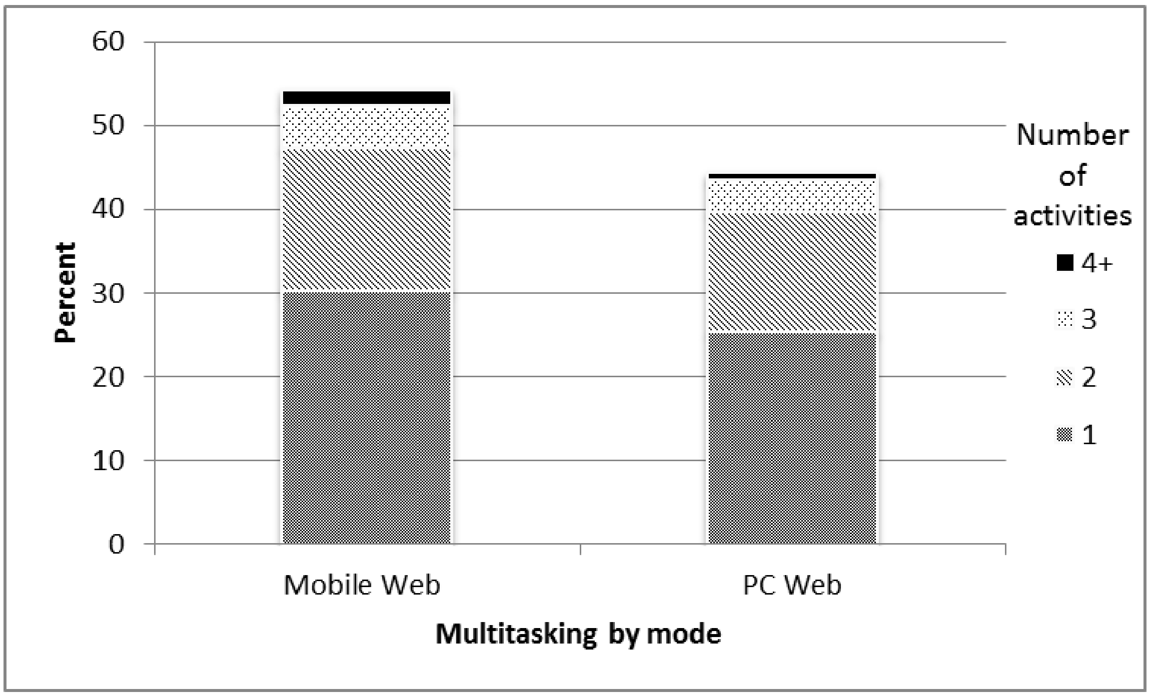


Figure 2.3: Self-reported number of other activities completed while filling out the Web survey

Response quality: satisficing

The fact that the respondents using a mobile device were more likely to divide their attention between other things while completing the survey raises the possibility that they might be less conscientious while answering than they would be in conventional

Web surveys. I consider five indicators of satisficing to understand what happened in this experiment. First I consider the effect of device and then I consider the association between divided attention and satisficing. Despite the increased prevalence of multitasking and distractions in mobile, respondents in mobile Web appeared to be at least as conscientious while completing the questionnaire as they were in PC Web (Table 2.5). Only one of the five indicators of satisficing differed between the two survey modes and, unexpectedly, it suggested less satisficing in mobile Web. Fewer respondents in mobile Web provided an unacceptably short answer (of one word or fewer) to the open-ended question than PC Web respondents (46.5% vs. 54.7%, $p < .001$). When those who skipped the open-ended question altogether were removed, the device difference remained statistically significant. Additional tests revealed the same pattern of longer answers recorded on phones; respondents in mobile Web typed significantly more words (adjusted means: 2.11 vs. 1.95, $p = 0.016$) and more characters (adjusted means: 15.3 vs. 13.6, $p < .001$) than PC Web respondents. I expand on what this result may mean in the Discussion section of this chapter.

Table 2.5: Adjusted differences in indicators of satisficing between mobile and PC Web

Indicator	*n*	Mobile *Mean (SE)*	PC *Mean (SE)*	*p*
Rounding Average number of rounded answers per respondent	2142	4.30 (0.07)	4.34 (0.06)	n.s.
Non-differentiation Degree to which respondents failed to differentiate between political attitude items with their answers	2142	0.35 (0.01)	0.34 (0.00)	n.s.
Superficial cognitive processing Number of incorrect answers to cognitive reflection test	2131	1.88 (0.03)	1.85 (0.03)	n.s.
Short open-ended response Percentage describing their hobbies with one word or fewer	2144	**46.49** **(1.52)**	**53.72** **(1.47)**	**<.001**
Avoiding "half-open other" response Percentage choosing closed-ended response instead of typing open-ended response	2144	72.05 (1.38)	71.04 (1.33)	n.s.

NOTE 1: Means/percentages adjusted for the effects of survey period and sequence.
NOTE 2: Sample sizes vary due to item missing data for one or more of the items used to form an indicator.
NOTE 3: Likelihood ratio tests for the random respondent effects were significant for all three of the LMMs (fit for continuous measures), suggesting substantial within-person correlation even after accounting for other variables (mode, survey period, and sequence). The estimated marginal correlations among observations for the same respondent were 0.61 and 0.54 for the GEE models (fit for binary measures).

Why were mobile Web respondents at least as conscientious as PC respondents? To help answer this question, I consider the associations between satisficing and divided attention and familiarity with smartphones. I fit a series of multivariate models using two predictor variables that were expected to affect attention to the survey itself: multitasking (yes vs. no) and being distracted (yes vs. no). I also included two predictors that serve as proxies for familiarity with smartphones: smartphone owner (yes vs. no) and type of device used to complete the survey (owned vs. loaned). (These two variables are different because some smartphone owners used a borrowed phone while others used their own phone.) Age, gender, and education were also included as control variables (Table 2.6). I found the same mode difference for the length of open-ended responses in these larger models. There was a significant main effect of one of the familiarity variables: respondents who used their own devices to complete the survey were less likely to record short answers, perhaps because they were more comfortable typing on the touchscreen keypad than those using a borrowed device. I also assessed if there were two-way interactions between either the attention variables or familiarity variables and mode, and found no significant interaction effects at the .05 level.

These further analyses revealed that, contrary to my expectations, divided attention and familiarity with smartphones did not have a substantial impact on these satisficing indicators.

Table 2.6: Estimated parameters in multivariate models regressing satisficing indicators on experimental factors, demographic control variables, attention to the survey, and device familiarity

	Rounding	Non-differentiation	Number of incorrect answers to cognitive reflection test	Short open-ended response	Avoiding "half-open other" response
Predictor	*Est. (SE)*	*Est. (SE)*	*Est. (SE)*	*Est. (SE)*	*Est. (SE)*
Intercept	4.912***	0.358***	1.908***	-0.133	1.246***
	(0.159)	(0.012)	(0.082)	(0.153)	(0.175)
Experimental factors					
Mode: Mobile vs. PC Web	-0.041	0.006	0.029	-0.282***	0.053
	(0.075)	(0.004)	(0.024)	(0.064)	(0.071)
Period: 1 vs. 2	-0.016	0.008	0.110***	-0.098	-0.031
	(0.074)	(0.004)	(0.024)	(0.063)	(0.071)
Sequence: mobile first vs. PC first	-0.157	0.007	-0.062	-0.036	0.061
	(0.106)	(0.008)	(0.058)	(0.105)	(0.116)
Demographic control variables					
Age: 16-34 vs. 35+	0.049	-0.037***	-0.035	0.508***	-0.048
	(0.123)	(0.009)	(0.067)	(0.121)	(0.132)
Gender: male vs. female	-0.136	0.023**	-0.411***	0.203	0.129
	(0.106)	(0.008)	(0.058)	(0.105)	(0.116)
Education: <college vs. >college	-0.504***	-0.036***	0.449***	0.152	0.095
	(0.109)	(0.008)	(0.059)	(0.108)	(0.118)
Divided Attention					
Multitasked: yes vs. no	-0.010	0.011	0.015	0.037	-0.084
	(0.096)	(0.006)	(0.036)	(0.086)	(0.095)
Distracted: yes vs. no	0.019	-0.002	-0.057	-0.040	0.056
	(0.101)	(0.006)	(0.038)	(0.090)	(0.099)
Device familiarity					
Smartphone owner: yes vs. no	-0.440	-0.029	0.025	0.526	-0.367
	(0.286)	(0.021)	(0.156)	(0.290)	(0.300)

Device used: own smartphone vs. borrowed smartphone	0.277	0.018	-0.216	-0.564*	-0.257
	(0.277)	(0.021)	(0.151)	(0.280)	(0.286)
Estimated variance of random respondent effect[c]	1.781***	0.014***	0.858***	0.529[d]	0.531[d]

*p<0.05, **p < 0.01, ***p < 0.001
N = 2122 for all models
[c]Likelihood ratio tests used for significance testing
[d]Estimated marginal correlation among observations for the same respondent

Response quality: disclosure

The fact that respondents in mobile Web were more likely to be away from home and around other people when filling out the questionnaires than PC Web respondents raises the possibility that mobile respondents might be less willing to disclose sensitive information. I consider five items with more or less socially undesirable answers to understand what happened in this experiment. Contrary to my expectations, mobile respondents appeared to be as willing to disclose sensitive information as PC respondents. As shown in Table 2.7, respondents in mobile Web provided the same total number of socially undesirable answers (1.88) as PC Web respondents (1.90) (p = .55) on average. Moreover, none of the means for the five indicators of socially desirable responding differed significantly between the two survey modes. I also examined the pattern of results, but found no consistent effects; the estimated means suggest that mobile Web elicited slightly more truthful answers for three items (exercise, TV viewing, attitudes towards immigrants) and slightly less truthful answers for the other two items (binge drinking, driving while intoxicated). Thus, the completion device had no consistent effect on disclosure of sensitive information, despite the fact that respondents in mobile Web were in less private settings than PC Web respondents when completing the questionnaires.

Table 2.7: Adjusted differences in socially undesirable responding between mobile and PC Web

Indicator	Mobile *Mean (SE)*	PC *Mean (SE)*	*p*
Overall disclosure Total number of socially undesirable answers	1.88 (0.03)	1.90 (0.03)	n.s.
Individual items Percent reporting the selected socially undesirable answer:			
Exercise less than once per week	16.12 (1.12)	17.83 (1.13)	n.s.
Watch TV for more than 3 hours per day	45.09 (1.50)	45.84 (1.45)	n.s.
Binge drank in past 30 days	43.76 (1.50)	41.88 (1.44)	n.s.
Ever driven while intoxicated	42.60 (1.48)	41.52 (1.43)	n.s.
Immigrants make country worse	40.30 (1.49)	42.68 (1.45)	n.s.

N = 2142
NOTE 1: Means/percentages are adjusted for the effects of survey period and sequence.
NOTE 2: A likelihood ratio test for the random respondent effect was significant for the LMM (fit for the overall disclosure indicator). The estimated marginal correlations among observations for the same respondent were 0.58, 0.69, 0.64, 0.74, and 0.58 for the GEE models (fit for binary measures).
NOTE 3: For the overall model, the distribution plots and quantile–quantile plots and for the conditional residuals are shown in Appendix B.

Why were mobile Web respondents as willing to disclose sensitive information as PC respondents? To test whether socially desirable responding was directly affected by the setting in which respondents completed the Web survey, I fit a series of multivariate models using two measures of privacy: presence of other people (yes vs. no) and physical location (at home vs. away from home). As before, variables for age, gender, and education were included as controls.

As shown in Table 2.8, I found the same null effect of mode. I found significant period effects both for overall disclosure and for two out of the five individual items, indicating respondents were more likely to disclose sensitive information when answering some questions for the second time. The third party presence variable was negatively associated with disclosure in the overall model and in four of the five models for individual items; but it only reached statistical significance in the model predicting reports of binge drinking. The effects of respondent location were mixed and, unexpectedly, being away from home was positively associated with reports of binge drinking. I also assessed if there were two-way interactions between the privacy variables and mode, and found no significant interaction effects (at the $p < .05$ level). The interactions between physical location and mode were in the expected direction for all 6 models, suggesting that being away from home reduced disclosure more in mobile Web than PC Web. This raises the possibility that this effect would be detected in a larger sample. The interactions between presence of others and mode were in the expected direction for the overall model but only for two of the five models for individual items.

Apart from the model predicting reports of binge drinking, these analyses indicate that being away from home and in the presence of others did not have a significant effect on disclosure. Mobile Web respondents could apparently record honest answers, even when in low privacy settings.

Table 2.8: Estimated parameters in multivariate models regressing socially undesirable responding on experimental factors, demographic control variables, context variables, and user characteristics.

	Overall disclosure	Exercise less than once per week	Watch TV for more than 3 hours per day	Binge drank in past 30 days	Ever driven while intoxicated	Immigrants make country worse
Predictor	*Est. (SE)*	*Est. (SE)*	*Est. (SE)*	*Est. (SE)*	*Est. (SE)*	*Est. (SE)*
Intercept	1.525***	-1.438***	-0.252	-1.182***	-0.488***	-0.965***
	(0.070)	(0.163)	(0.131)	(0.141)	(0.136)	(0.136)
Experimental factors						
Mode: Mobile vs. PC Web	-0.017	-0.110	-0.027	0.078	0.044	-0.099
	(0.028)	(0.081)	(0.055)	(0.062)	(0.055)	(0.063)
Period: 1 vs. 2	-0.101***	-0.226**	-0.228***	-0.074	0.004	0.007
	(0.028)	(0.079)	(0.055)	(0.062)	(0.056)	(0.063)
Sequence: mobile first vs. PC first	0.012	0.150	-0.028	0.000	-0.037	0.015
	(0.059)	(0.141)	(0.109)	(0.112)	(0.116)	(0.108)
Demographic control variables						
Age: 16-34 vs. 35+	0.181**	0.112	-0.584***	1.121***	-0.379**	0.544***
	(0.063)	(0.148)	(0.121)	(0.122)	(0.126)	(0.115)
Gender: male vs. female	0.554***	-0.225	0.047	1.021***	1.238***	0.270*
	(0.059)	(0.140)	(0.110)	(0.116)	(0.115)	(0.109)
Education: <college vs. >college	0.147*	0.006	0.651***	-0.003	-0.620***	0.555***
	(0.060)	(0.144)	(0.114)	(0.114)	(0.119)	(0.112)
Privacy[a]						
Around other people: yes vs. no	-0.045	-0.027	-0.053	-0.195*	0.083	-0.033
	(0.042)	(0.113)	(0.074)	(0.088)	(0.082)	(0.084)
Away from home: yes	0.102	0.178	-0.045	0.391**	-0.053	0.094

vs. no						
	(0.059)	(0.161)	(0.113)	(0.120)	(0.113)	(0.121)
Estimated variance of random respondent effect[b]	0.833***	0.586[d]	0.669[d]	0.600[d]	0.700[d]	0.580[d]

*p<0.05, **p < 0.01, ***p < 0.001
N = 2122 for all models
[b]Likelihood ratio test used for significance testing
[d]Estimated marginal correlation among observations for the same respondent

Response quality: slider and date picker

I next test whether my expectation that the reporting errors for a slider question and date picker question will be larger when respondents record their ages using a smartphone compared to a PC. For both input widgets, errors are interpreted as formatting mistakes rather than mistakes in the cognitive response process; I assume that respondents know their age and don't need to formulate it when asked. For the slider question, respondents were asked to record their age using a horizontal slider scale with end labels ("0" and "100"); the slider was positioned at the far left-hand-side of the scale at the outset. No numeric feedback about the location of the slider was provided; while this would not be done in practice, as it might communicate to respondents that they should estimate their ages, I wanted to make the input task challenging. The size of the scale and the way users interacted with it differed by mode. On smartphones, the scale was relatively narrow and the bar was moved by touching and dragging it, whereas on PCs the scale was relatively wide and the slider bar was moved by down-clicking and dragging it (Figure 2.4).

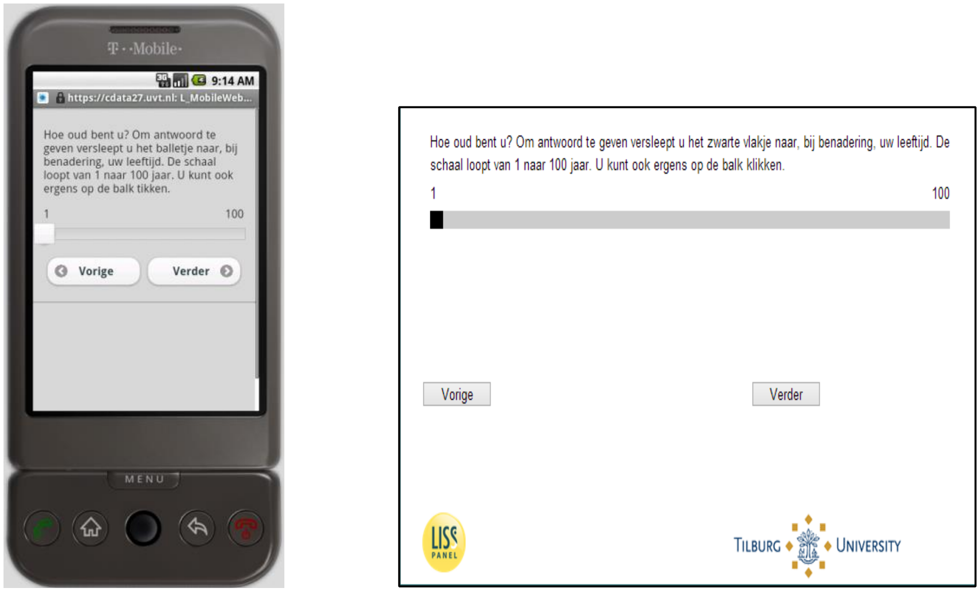


Figure 2.4: Slider question in mobile and PC Web

For the date picker, respondents recorded their date of birth in mobile Web by using spin wheels (one for month, day, and year, respectively) and in PC Web by using drop-down boxes (Figure 2.5). I avoided using spin wheels in PC Web because they are typically used on touchscreen devices (especially iPhones) and not PCs. While this introduces a potential confound between widget and device, I opted for naturalness (i.e., ecological validity) over perfect experimental symmetry. For this question, the benchmark value is respondent birth year according to the panel records. I focus on the recorded year of birth because it is available in the panel archives while the day and month of birth are not.

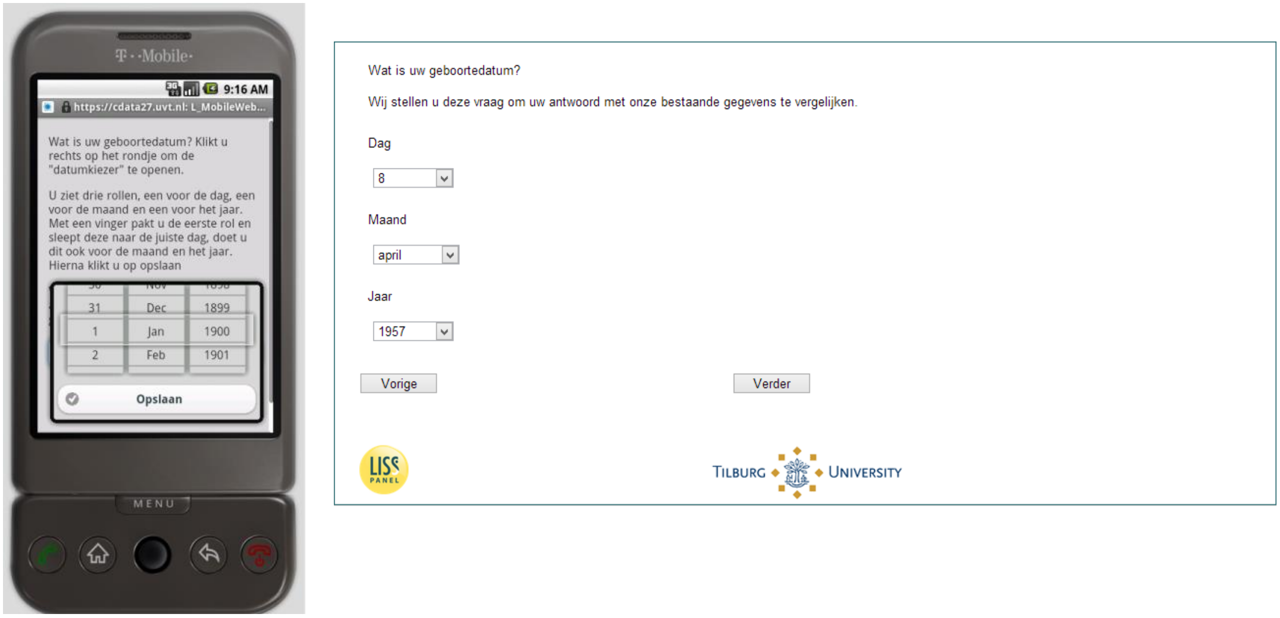


Figure 2.5: Date picker/drop box question in mobile/PC Web

First I focus on the slider results. Two respondents in mobile Web failed to move the slider bar (i.e., they recorded an answer of zero). For those who moved the slider, I classified outliers as differences of more than three interquartile ranges from the mean discrepancy (more than 15.7 years from their actual age in this case). This identified 16 extreme ratings from the mobile Web survey and 11 extreme ratings from the PC Web survey that were excluded from other analyses.

As expected, motor performance on the slider question was slightly worse on a smartphone than a PC. Fewer respondents recorded accurate answers (i.e., within one year of their age in the LISS panel frame to account for those who had a recent birthday) in mobile Web (26.5%) than PC Web (34.1%) ($p < .001$), though the difference is modest. The low accuracy rates in both modes are probably related to the fact that no numeric feedback about the position of the slider was given (and not errors in the records that the LISS panel keeps about their panelists). The average deviation was 3.68 years in mobile Web compared to only 3.04 years in PC Web (adjusted mean difference=0.64, $t(861)=5.69$, $p < .001$). The errors had a directional effect on the overall means recorded via sliders; respondents tended to pass the place on the scale where they should have stopped which inflated the overall estimates of age by 1.4 years in mobile Web (recorded = 45.8 vs. actual = 44.4) and 0.6 years in PC Web (recorded=45.2 vs. actual=44.6).

Were respondents less precise in mobile because they took less time? My analysis shows that respondents actually took more time in mobile (median=27 seconds) than in PC Web (median=22 seconds) and were still less precise, so it is unlikely a result of time differences between the two modes. To further explore the input errors, I divided the 100 point scale into seven intervals (15-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75+) and calculated the mean absolute input error for each interval. As shown in Figure 2.6, the discrepancy between LISS records and survey responses is smallest for the middle intervals, and no different for the two modes; for the left-most and right-most intervals the absolute error is greater for mobile than PC responses (15-24 interval: mode difference=1.75 years, $p < .001$; 25-34 interval: 1.42 years, $p < .001$; 65-74 interval: 0.67 years, $p = .061$; 75+ interval: 2.37 years, $p = .003$). When looking at signed errors rather than absolute errors, the pattern of errors in mobile at these intervals suggest a *regression to the mean*. Young respondents tended to run pass the place on the scale where they should have stopped (which was the general trend) and older respondents tended to fall short of the place where they should have stopped.

The fact that I find a mode effect for both older and younger respondents indicates that this effect is not just a result of poor precision among older respondents. Rather, it indicates that it is easier to find the midpoint of the mobile scale than it is to find other points. Future studies on smartphone slider bars should test for this same pattern using other verifiable measures (e.g., respondent height and weight) or by instructing respondents to move a slider to a particular set of values using different length scales.

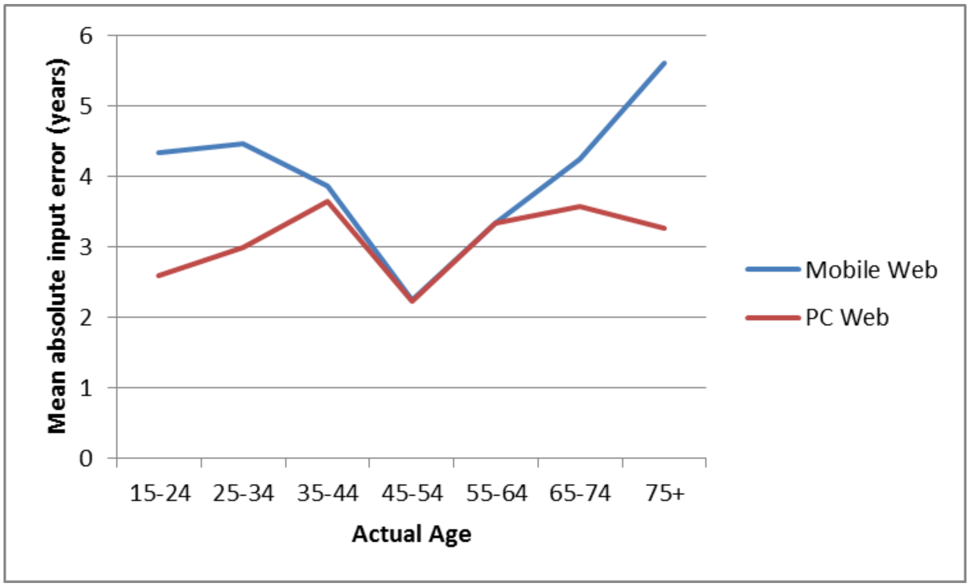


Figure 2.6: Slider input errors in mobile and PC Web

Next I consider the date picker widget. I first focus on item missing data, which for this tool refers to failing to move at least one of the wheels to change the recorded date from its default value (January 1, 1900). As I expected, more respondents in mobile Web failed to answer the question (*n* = 52, 5.14%) than PC Web respondents (*n* = 0, 0.0%). While it is possible that some respondents skipped the question in mobile Web in order to save time and effort, further analysis suggests the opposite. Item nonresponders actually spent more time on the question (median = 141 seconds) than the item responders (median = 89 seconds), so the item missings are unlikely a result of satisficing. Instead, respondents skipped this question after spending some time trying to record their answer. I cannot rule out the possibility that the date picker did not work properly on some devices; of the 52 missing values, 33 came from those using loaned

Android devices, 14 from those using their own Android devices, and only 5 from those using their own iPhones. When the input tool did work properly, respondents may have failed to record an answer because they lacked technical knowledge about how to use touchscreen gestures to spin the picker wheel (i.e., knowledge-based mistakes).

For those who moved at least one of the wheels (and were not classified as item nonresponders), I classified outliers as differences in the 99th percentile or above (more than 35 years in this case) and excluded these cases from other analyses. As with the slider, motor performance was worse on a smartphone than a PC. Conditional on being able to move the wheel, fewer respondents recorded accurate answers in mobile Web (96.0%) than PC Web (98.9%) ($p < .001$), though the differences were modest and responses were mostly accurate in both modes. The average size of the input error was 0.19 years in mobile Web compared to only 0.08 years in PC Web (adjusted mean difference=-0.11, $t(797)=2.34$, $p = .02$). Device type (loaned phone vs. Android vs. iPhone) was not associated with the size of the errors. Overall, the lower accuracy rate in mobile along with the relatively large number of item missings in mobile indicate that the date picker was moderately difficult to use.

Was there consistency between the picker responses and slider responses? After I converted the birth year into an age in years, I found the correlations between the two answers to be quite high in both modes (mobile: .96; PC: .97). But fewer respondents in mobile Web were able to record the same answers (27.4%) than PC Web respondents (35.0%) ($p < .001$). This is further evidence that the slider and picker were difficult to use on a smartphone.

Discussion

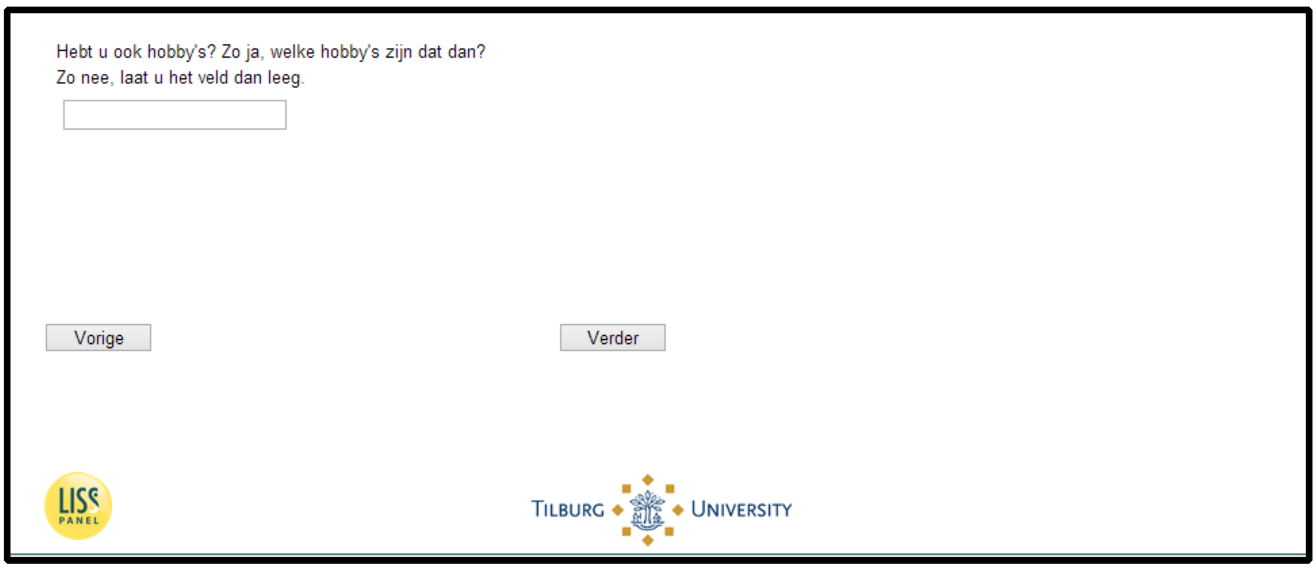
This chapter reported on a crossover experiment conducted using a probability-based scientific panel to compare survey measurement error when respondents answer questions on mobile devices to when they answer on PCs. There are three main conclusions. First, respondents in the mobile Web survey really were more *mobile* and more engaged with the other people and things around them compared to PC Web respondents. Specifically, they were significantly more likely to report multitasking and being in the presence of other people while completing the questionnaires. These findings add to a growing body of literature suggesting that researchers have less control over respondents' circumstances and surroundings (i.e., use context) in mobile Web than in traditional Web surveys (De Bruijne and Wijnant 2013a; Mavletova 2013; Mavletova and Couper 2013; Revilla et al. 2014; Toepoel and Lugtig 2014).

Second, despite respondents being more *mobile*, I found no serious perils in mobile administration. It had little impact for the measures that I used which should be sensitive to respondents' context. In fact, response quality – conscientious responding and disclosure of sensitive information – was equivalent between mobile and PC Web. This is because differences in context are modest and because respondents' use context was generally not predictive of their response quality. For example, being around other people did not seem to affect socially desirable responding. One possibility is that respondents did not believe that others could look over their shoulder to read such a small screen. Another is that they were surrounded by strangers with no prior knowledge of the information requested (rather than family and friends) and third-party effects depend on just who is nearby (Aquilino 1993; Aquilino et al. 2000). Or, respondents did not find the items to be highly sensitive, since almost half endorsed each item (except for the item about exercise).

In any case, my findings add to the growing number of studies that find device equivalence for such data quality indicators as item missing data (Lugtig and Toepoel 2015; Toepoel and Lugtig 2014), the rate of non-substantive responses (Mavletova 2013), the number of answers to a check-all-that-apply question (Lugtig and Toepoel 2015 and Toepoel and Lugtig 2014), and primacy effects (Mavletova 2013; Toepoel and Lugtig 2014). I did find significantly longer open-ended responses when mobile devices were used. This was contrary to my expectation that respondents would be reluctant to type open-ended answers using their smartphone's touchscreen keypad, as others have found (Mavletova 2013; Peterson et al. 2013; Wells et al. 2014). It is in line, however, with recent findings that people have become more comfortable typing on mobile devices over time (Bosnjak et al. 2013a; Buskirk and Andrus 2014; Lugtig and Toepoel 2015; Toepoel and Lugtig 2014; Zahariev et al. 2009), which some have suggested is a result of social learning that took place during the so-called text revolution (Link 2014). A future analysis could investigate whether panelists' texting behavior affected the length of their answers. The longer answers in mobile Web could also be related to the fact that voice-to-text options and auto-complete, more widely used on smartphones than PCs, are facilitating open-ended answer entry for some respondents (AAPOR 2014).

An alternative explanation for this result could be that text box size – which is known to affect the length of open-ended answers in regular Web surveys (see Couper, Traugott, Lamias 2001; Smyth et al. 2009) and mobile Web surveys (Wells et al. 2014) – carries different meanings on different sized screens. As shown in Figure 2.7, even though the absolute text box size was comparable in the two modes, the text box on the mobile device (Fig. 2.7a) occupied a larger proportion of overall screen space than the PC version of the text box (Fig. 2.7b). When respondents viewed the two text boxes, they may have inferred the requested amount of information not from the absolute size of the text box but from its size relative to their device's screen size. In any case, this suggests the need for future research, perhaps with a narrative-type open-ended question, about the relationship between text box size and screen size.

2.7a: Open-ended question in mobile Web



2.7b: Open-ended question in PC Web

Figure 2.7: Text box format for open-ended question

Third, respondents made more input error in mobile Web than PC Web for certain question formats. This was evident for a slider question that required using fine-motor skills and for a date picker that required knowledge about the how to spin its wheels. This conclusion supports the recent AAPOR task force (2014) recommendation that "… some of the more sophisticated ways of recording a response in traditional online surveys (such as slidebars, drag and drop approaches, or drop-down boxes) may be much more difficult to utilize by respondents on a mobile device and should be avoided." Other question designs that may produce more measurement error in mobile Web relative to PC Web surveys include grids, check-all-that-apply approaches, and questions that use long lists of response options (Peterson et al. 2013; Stapleton 2013). Radio buttons, which are sometimes larger in (optimized) mobile surveys than on PCs if they are displayed as wide touchscreen friendly buttons, appear to be a viable alternative to some of these formats. In the case of slider bars, for example, radio buttons should not only improve accuracy in mobile Web, but may also reduce completion time and item missing data in PC Web (Cook et al. 2001; Couper et al. 2006). But radio buttons do not work for all items (e.g., grids), and so future research in needed about other question formats for small devices.

Given these results, it is my view that a distinction should be made between question content (sensitive or not, burdensome or not) and question design (layout, type of input tool, etc.). Mobile Web and PC Web are subtle variants of each other when it comes to the content that they can be used effectively to gather; questions with sensitive content or requiring burdensome response processes could be just as effectively administered on smartphones as they could be on PCs. On the other hand, the effectiveness of different question designs seems to vary by mode. Input tools that are commonly used in mobile surveys (e.g., sliders and pickers) actually produced substantial measurement error in mobile Web, while other question designs (e.g., open-ended questions, numeric entry) performed as well in mobile Web as PC Web.

It should be noted that this study focused on a single Web panel in the Netherlands comprised of experienced survey takers, and the pattern of results might be specific to the sample being studied. Participation by LISS panelists was also conditional on agreeing to join the mobile Web experiment and using the assigned device. The panelists who agreed to participate were better educated than the full panel. My analysis also assumes that there were no undetected crossover effects. It would certainly be beneficial to conduct more research on this topic in a cross-sectional survey or in another country. Nevertheless, there are several reasons why these findings may generalize to other settings. Since LISS is a PC Web panel, its members should not be any more familiar with mobile Web surveys than people outside of the panel. Indeed, at the time of this experiment, de Bruijne and Wijnant (2014) found in an independent analysis that only 1.6% of LISS panel surveys were completed using smartphones. Another reason that these results may generalize to other settings is that the sample was quite diverse (in terms of age, for example) because it is a probability panel and not a panel of volunteers. Furthermore, the Netherlands is quite similar to the other high-income countries in terms of its internet penetration rates. For instance, approximately 57% of U.S. adults use their cell phones to go online (Duggan and Smith 2013) compared to 60% of internet users in the Netherlands (Statistics Netherlands 2013).

Smartphones are powerful tools for capturing new kinds of data. As this technology becomes more pervasive in people's lives, researchers must find innovative ways to leverage it to collect high-quality data for social and medical science research. This may require new ways of thinking about data collection, ranging from "modular" survey design to reach those who use phones in short bursts (e.g., Kelly and Stevens

2013) to app-based passive collection of travel data using GPS and health data using bluetooth sensors (e.g., De Nazell et al. 2013). In the meantime, while the mobile revolution rages on, survey practitioners can take heart in knowing that respondents using smartphones can be careful and honest survey respondents, even when distracted by the other people and things around them, as long as they are presented with question formats that are easy to use on small touchscreens.

Chapter 3: Influence of Sample Person Characteristics on Nonresponse in a Mobile Web Survey

Summary

Web surveys are now completed on a mix of different computing devices, and so these surveys are subject to nonresponse not only from the sample members who (intend to) use PCs but now also from the sample members who (intend to) use mobile devices. Yet little is known about the causes of nonresponse in mobile Web surveys. In this chapter, I examine how sample person characteristics influence nonresponse in an online panel for an individual who is prompted to use a mobile device (in particular, a smartphone). Panelists (n = 5486) were asked whether they were willing complete a survey on a smartphone before a subset of willing panelists (n = 1388) were invited to do so. The results reveal that several social, psychological, attitudinal, and behavioral measures are associated with willingness to participate, and to a lesser extent survey response (conditional on expressing willingness). I found that several factors that are known to influence the nonresponse process in other survey modes such as personality traits, civic engagement, and attitudes about surveys also play a role in the nonresponse process in mobile Web surveys. In addition, I identified other factors that may be specific to this mode, including smartphone use, social media use, and smartphone e-mail use.

Introduction

Researchers who conduct surveys have long been interested in the factors that influence response and nonresponse. Among other data collection features, mode can have an effect on response rates and on just which respondent characteristics affect survey participation. There is a literature on the causes of nonresponse in Web surveys that are completed on computers and laptops (e.g., Fan and Yan 2010; Keusch 2015). Web surveys are now completed on a mix of different computing devices though, and so these surveys are subject to nonresponse not only from the sample members who (intend to) use PCs but now also from the sample members who (intend to) use mobile devices. This nonresponse can occur in generic Web surveys for an individual who intends to use a mobile device rather than a PC but fails to participate. The focus here is on nonresponse that occurs in a Web survey in an online panel for an individual who is prompted to use a mobile device (in particular, a smartphone) but uses another device or does not respond at all.

There is little information available about the causes of nonresponse in mobile Web surveys, which can inform approaches to reduce or adjust for it. Most of the research in this area is limited to using demographic correlates of response. Furthermore, most studies only focus on samples of volunteers who have already expressed willingness to participate. Few studies consider stated willingness to do a mobile Web survey and survey response separately.

This chapter tries to address these research gaps by examining how social, psychological, attitudinal, and behavioral measures influence the nonresponse process in mobile Web surveys. I separate the participation process into two components, willingness to participate and the survey response (conditional on willingness), because different factors may affect each step. I review the survey design attributes and sample person characteristics that may influence the nonresponse process. I then use data from a recent mobile Web survey conducted in a probability-based panel to model stated willingness and survey response.

Survey design factors

Several design features that respondents are aware of when they are invited to participate in a mobile Web survey have implications for nonresponse. First, the mode used for prenotifications, contact, and reminders (e.g., e-mail, SMS, mail) influences survey response. De Bruijne and Wijnant (2014) and Mavletova and Couper (2014) both found that SMS invitations increase mobile response rates compared to e-mail invitations. The design of e-mail invitations and their degree of optimization for small screens may prove to be an important factor (for examples, see Buskirk 2013). The timing and frequency of invitations and reminders may also have implications for nonresponse since mobile Web use peaks in the evening (Lipsman and Aquino 2013) invitation may be more effective if sent in the evening, although effects due to the season or day of the week are not well understood. The type of login requirements (automatic vs. manual) are another factor since manually typing a login and password is particularly burdensome when using a phone and may discourage people from accessing a survey.

Design features that are decided on well before inviting sample persons to participate may also have implications. Using an app-based approach rather than a browser-based approach may increase the likelihood of nonresponse since not all smartphone users are willing and able to install research apps (Revilla et al. 2014). The use of enhanced data collection that requires such tasks as activating GPS, taking photos, or connecting to bluetooth enabled sensors may lead to high levels of nonresponse (Revilla et al. 2014). In addition, the stated length of the survey is a factor that may be of special concern in mobile Web surveys. Long questionnaires may increase the perceived burden of participation and reduce response propensities since people tend to use their phones to go online for relatively brief periods of time (Cui and Roto 2008). The stated topic and description of the survey may also matter; specifically, any indication that personal information will be collected may reduce response rates (Walton et al. 2013). In longitudinal or panel surveys, the length and frequency of previous surveys may be a factor. Survey sponsor and data collection organization have implications for nonresponse (e.g., Groves, Presser, and Dipko 2004), although their role in mobile Web surveys is not well understood. Finally, the size and type (e.g., conditional vs unconditional, cash vs. lottery prizes) of the incentive may affect participation in mobile Web surveys. Mavletova and Couper (2015) found that offering a larger cash incentive for completion on a mobile device compared to a PC was successful in increasing the proportion of mobile Web respondents.

While studying nonresponse in a panel makes it difficult to generalize to cross-sectional surveys, one advantage is that information is available about more than just survey design features. Detailed information about panel members has already been collected in earlier waves that can be used to study the respondent factors that influence participation. According to leverage-salience theory (Groves, Singer, and Corning 2000), person-specific factors (e.g., attitudes about surveys, topic interest, personality traits) are key factors in participation decisions because they determine the importance of different survey attributes (e.g., survey topic, survey sponsor, type and amount of incentive) to potential respondents, and whether such attributes act as motivators or deterrents in their decision.

As mentioned earlier, I separate the participation process into two components: willingness to participate and the actual participation. Next, I consider the respondent factors that may influence each process.

Willingness to participate in mobile Web surveys

Initiating a survey that requires smartphones to participate often involves identifying eligible sample members who are willing to use smartphones rather than PCs to complete surveys. Identifying such people may involve a screener questionnaire to ask about their interest in participating. The percentage who are willing to participate in such studies ranges from 8% to 61%, depending on the country and survey setting (Mavletova and Couper 2014; Revilla et al. 2014; Wells et al. 2014).

The likelihood of identifying people who are willing to use smartphones to complete a Web survey may depend on several factors. Access and use of the mobile Internet (i.e., coverage) is one consideration. If a smartphone is required to participate in future surveys, then non-owners are ineligible. On the other hand, if phones are provided to non-owners to achieve full coverage of the sample, then both owner and non-owners may be eligible, though non-owners may be less likely to comply with the request to participate if they don't know how to use smartphones. The wording, timing, and location of the request to participate may affect the likelihood of compliance. For example, the chances of compliance might be greater when answering this question at the beginning of a questionnaire rather than at the end of a long and burdensome questionnaire, and also when asking about “willingness” (e.g., are you willing to participate in mobile Web surveys) rather than asking about “intention to participate” (e.g., do you intend to participate in a mobile Web surveys) because the former is related to merely complying with a request while the latter implies that the respondents plans to participate.

Perceived burden of participation is another potential correlate of willingness. This may be influenced by experience, comfort, and familiarity with mobile Web along with respondents’ use habits (e.g., the amount of time spent each day using smartphones) and their phone type (full touchscreens vs. navigation buttons).

General personality traits have been shown to predict individual behavior, including internet use (Tuten and Bosnjak 2001), mobile phone use (Butt and Phillips 2008), and willingness to join Web panels (Bosnjak et al. 2013b), and may also predict willingness to participate in a mobile Web survey. The most widely accepted representation of personality includes the following dimensions: Extraversion, Emotional Stability, Agreeableness, Conscientiousness, and Openness to Experience[7] (Costa and McCrae 1992). Specifically, individuals’ willingness to try something new or their feeling of obligation to acquiesce to a request, which are likely to be related to traits such as openness to experience and agreeableness, may increase their chances of complying with the request to participate.

Another personality trait that appears to be relevant to participation in mobile Web surveys is need for cognition, which refers to an individual’s tendency to engage in and enjoy effortful cognitive activities (Cacioppo and Petty 1982). This is because

[7] Extraversion reflects the tendency to have active engagement with other people. Emotional Stability is closely related to neuroticism and reflects individual differences in the tendency to feel anger, anxiety, and depression. Agreeableness is characterized by high levels of cooperation and trust. Conscientiousness is marked by being organized and dependable. Openness to Experience reflects a willingness to try new things.

completing a survey on a smartphone device may seem like a cognitively demanding task.

Actual participation

A distinction can be made between stated willingness and response; just because people are willing to participate does not mean that they will do so. Response rates among those who express willingness to participate in mobile Web surveys range from 23% to 58%, depending on the survey setting (Buskirk and Andrus 2014; Wells et al. 2013). An individual's decision to respond will be influenced by several factors. Situational factors like the time and place that they receive the invitation may be a correlate of nonresponse if certain settings do not lend themselves to completing a Web survey. These include the factors that are related to the mobility afforded by mobile devices: physical location (at home vs. away from home), physical mobility (whether they are on-the-go or not), the nature of other tasks they are engaged in (multitasking, distractions), and the presence of other people. Opening the invitation on a computer rather than a smartphone may decrease the likelihood of response since it would require respondents to manually type the URL into their smartphone rather than simply click the link in the invitation. Perhaps for this same reason, De Bruijne and Wijnant (2014) found that response rates were lower in a mobile Web survey among those who do not read emails on their smartphones. Individuals who are busy with other activities may be less likely to respond, especially since mobile Web surveys might be expected to take longer to complete than PC Web surveys.

Several socio-demographic factors have been shown to affect participation decisions in mobile Web, although the reasons why are not yet well understood. In a mobile Web survey carried out in Russia, Mavletova and Couper (2014) found that respondents reported higher monthly household income than nonrespondents. In a mobile Web survey carried out in the same probability-based online panel in the Netherlands (LISS) as this one, De Bruijne and Wijnant (2014) found that respondents were younger and more educated than nonrespondents.

Attitudes about surveys that reflect differences in people's views about the value, enjoyment, and burden of completing surveys affect participation decisions in other modes (e.g., Couper et al. 1998) and may also play a role in mobile Web surveys.

Several concepts from sociological research appear to be relevant to the decision to participate, although little is known about their role in mobile Web surveys. Civic engagement, which refers to civic or political actions that are designed to affect others in the community (e.g., volunteer work, active membership in groups, voting), may increase the likelihood of response if mobile Web surveys are viewed as a way to do this (e.g., Couper, Singer, and Kulka 1998). Social integration, which refers to the strength of one's identification with mainstream culture (e.g., strong social support, larger social networks, and increased participation in social activities), may also increase the likelihood of response, either because participation in mobile Web surveys is viewed as a social norm or because these surveys are viewed as a kind of "social" activity (e.g., Groves and Couper 1998, Chapter 5).

Several concepts from psychological research also appear to be relevant to the decision to participate, although they have not yet been tested in a mobile Web survey. The first is Davis's (1986) technology acceptance model (TAM), which posits that the belief that a technology is easy to use has a causal effect on the belief that it is useful and enjoyable, which in turn motivates a potential user to actually use the technology. In line with this hypothesis, Bosnjak and colleagues (2010) found that perceptions about whether mobile Web surveys are enjoyable and trustworthy are predictive of the intention to participate in such surveys.

Another area of psychological research that may be relevant to mobile participation focuses on compliance heuristics (Cialdini 1988), which in the survey context refer to rules that a sample member might apply when deciding whether to respond, especially when they do not have the time or motivation to make careful decisions (Groves, Cialdini, and Couper 1992). Specifically, the consistency principle, which posits that people have a basic desire for consistency that leads them to engage in similar types of behavior over time, may be relevant to mobile participation. In this context, it suggests that sample members who already use smartphones to complete surveys – or to complete survey-like tasks that require providing information online like filling out forms or posting content on social media – will be more inclined to respond to a mobile Web survey compared to those who don't use their smartphones in this way.

Respondents' smartphone use habits may also affect participation for practical reasons. For example, those who rarely use a smartphone to go online or do so for brief browsing sessions may not have sufficient time to complete a mobile Web survey. Mavletova and Couper (2014, 2015) report that nonrespondents in two different mobile Web surveys tended to use their phones less frequently than respondents.

Relationship between stated willingness and response

In the above section, I considered the willingness to participate and response separately. But some factors may influence both processes. For example, the feeling of obligation to acquiesce to a request may increase willingness propensity and response propensity. Other factors that might influence both processes include busyness, attitudes about surveys, and smartphone use habits.

One area of psychological research that is relevant to this linkage is the theory of planned behavior (Ajzen 1985), which posits that a key predictor of behavior is an individuals' intention to perform that behavior. In line with this theory, Bosnjak and colleagues (2010) report that those who are more likely to express the intention to participate in a mobile Web survey are the most motivated to actually complete it.

In this chapter, I report on a study where one process is conditional on the other. Being invited to the mobile Web survey required having first expressed a willingness to do so. Several different theoretical frameworks can be used in these circumstances to show the influence of predictors on willingness and participation. Three of them are described in Figure C.1 in Appendix C. These frameworks vary in whether one assumes independence between the two processes and whether the processes are modeled together or separately. I use multiple approaches and first model the two processes separately and then fit a two-stage model that takes into account the fact that one process may be related to the other.

Several of the sample member characteristics hypothesized to influence the nonresponse process were measured in prior waves of the panel that conducted this study, and so the models are fit using data from a number of these waves. The modeling approaches and variables of interest will be further described later on.

Methods

Data collection effort

The data from this study come from the LISS panel, which was described in Chapter 2. In July-August 2013, the 6340 active panel members were invited to complete a five-item questionnaire which contained an item about their willingness to participate in a survey on a smartphone. The first question read:

> Bij dit onderzoek bestuderen we het invullen van LISS panel vragenlijsten met een smartphone. Het onderzoek start in september met een korte vragenlijst, in oktober en december wordt er voor dit onderzoek nogmaals een vragenlijst afgenomen. Als u zelf geen (geschikte) smartphone hebt, krijgt u er een in bruikleen. Hebt u interesse om aan dit onderzoek deel te nemen? Hebt u interesse om aan dit onderzoek deel te nemen?

The English translation is:

> In this study we examine filling out the LISS panel questionnaires with a smartphone. The study will start in September with a short questionnaire, and include other questionnaires in October and December. If you have no (suitable) smartphone, you will get a borrowed phone. Are you interested in participating in this research?

Although the literal translation is about "interest," I use the word "willingness" throughout this chapter because in this context it appears to be a polite, colloquial way of asking about panelists' willingness to do a Web survey using a smartphone[8]. Those who answered "yes" are considered eligible for the mobile Web study under the condition that they either owned a "suitable" smartphone (iPhone or Android, in this case) or were willing to except one. A subset of those who expressed willingness to participate were selected in a stratified sample that oversampled iPhone and Android phone owners compared to those who owned a different model or did not own a smartphone at all.

[8] Though in a different context, "willingness" might be considered to be a weaker indication of potential behavior than "interest."

One month later, the selected participants were instructed to complete a brief "baseline" survey in PC Web that collected a number of measures related to smartphone use.

One month after that, in October 2013, all of the selected panelists (regardless of whether they completed the baseline survey) were invited to take a Web survey on either their smartphone or their PC. Then, in December 2013, they switched devices; all of those panelists who were invited to use a smartphone were then instructed to use their PC and vice versa. The same questions were asked in both waves. For the analysis of response behavior, I look at participation in mobile Web regardless of wave and in conventional Web regardless of wave. Period effects were accounted for in the response models by using wave (1 vs. 2) as a control variable. The full data collection process is shown in Figure 3.1.

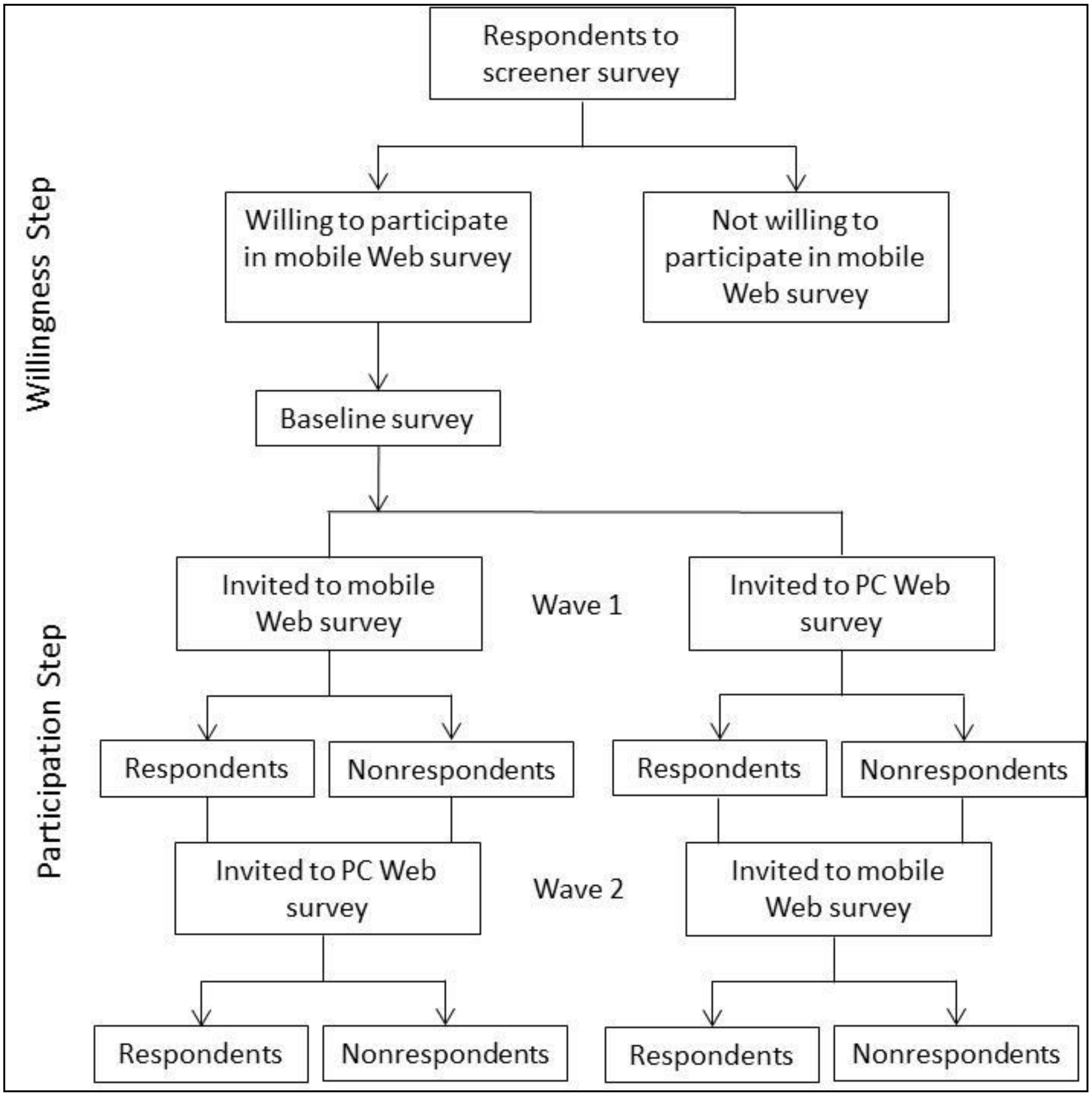


Figure 3.1: Participation process

As described in the previous Chapter, participants using a borrowed phone were sent their device (Samsung Gio) approximately one week before they were invited to take the mobile Web survey. All invitations to the survey were sent by e-mail. Two reminder emails were sent to all nonrespondents near the end of each month of data collection. The normal cash incentive for the LISS panel (15 Euros per hour) was provided as payment for participation in the study. The surveys contained 46 questions. It was browser-based and the mobile version was programmed to be optimized for small devices. The observed willingness rates and response rates are part of the focus of this chapter and will be reported in the Results section.

Predictors

For the nonresponse models, I use measures from LISS "core" studies which are surveys carried out annually in the panel, each during a different two-month field period. Several predictors were available from these surveys which are expected to be correlates of stated willingness or participation. Table 3.1 shows these predictors, how they were defined for this analysis and the surveys where they were collected. Measures from the wave of each core survey that panelists participated in most recently were used as long as it occurred within two years of the start of the data collection effort for the mobile Web study. This length of time was assumed to be both short enough to minimize the risk of using values that had changed over time and long enough to access values from earlier waves (for panelists who did not participate in later ones) to avoid dropping these panelists from my analysis all together.

Table 3.1: Attributes hypothesized to affect participation decision, their definition for this analysis, and the surveys where they were collected

Predictor	Definition/Notes	Survey[a]
age	Panelists were age 16 to 93.	1
% male	--	1
education	Six categories: 1. primary school, other; 2. junior high school; 3. high school degree; 4. vocational degree; 5. higher vocational degree; and 6. university degree or more.	1
imputed household income	Because the distribution was highly skewed, it was divided into five categories (1-5) based on quintiles. Missing values were coded as "1". The same pattern of results emerged in the multivariate models when more categories were used and when missing values were excluded.	1
urbanization level	Five categories: 1. not urban; 2. slightly urban; 3. moderately urban; 4. very urban; and 5. extremely urban.	1
civic engagement	Score is based on behavioral indicators of civic engagement; specifically, participation in 11 organizations with a 5-point scale for each (1 = no connection; 2 = donated money; 3 = participated in an activity; 4 = member; 5 = performed voluntary work). The 12 organization are: sports club; cultural association or hobby club; trade union; business organization; consumers' organization; organization for humanitarian aid; organization for environmental protection, peace or animal rights; church; political party; science, education, teachers' or parents' association; social society; other. Answers were summarized in an additive scale ranging from 12 to 60. High scores reflect increased civic engagement.	2
social isolation	6 items on three-point scales (1 = yes; 2 = more or less; 3 = no). The items are: I have a sense of emptiness around me; there are enough people I can count on in case of a misfortune; I know a lot of people that I can fully rely on; there are enough people to whom I feel closely connected; I miss having people around me; I often feel deserted. Recoded the reversed items and summarized in an additive scale ranging from 6 to 18. Higher scores reflect increased social isolation.	2
social trust	"Generally speaking, would you say that…") on a ten-point scale (0 = "You can't be too careful"; 10 = Most people can be trusted"). "Don't know" answers were coded as "5".	3
% married	--	1
% renting	--	1
satisfaction with leisure time	How satisfied are you with the way in which you spend your leisure time? (0 = "all satisfied"; 10 = completely satisfied"). "Don't know" answers were coded as "5".	2
children in household	Number of children ranged from 0 to 6.	1
% employed	This includes working or studying.	1
need for cognition	18 items on seven-point scales ("1-strongly inaccurate" to "7- strongly agree). Recoded the reversed items and summarized in an additive scale ranging from 18 to 125. The 18 items are from Cacioppo and Petty (1982).	3
openness to experience	Scale for each personality trait: 0 items on five-point scales ("1-very inaccurate" to "5- very accurate). Recoded the reversed items and summarized in an additive scale ranging from 10-50. The 10 items are from Goldberg's 'IPIP' Big-Five factor markers (Saucier and Goldberg 2002).	3
extraversion		3
agreeableness		3
conscientiousness		3
emotional stability		3
survey enjoyment	3 items on 7-point scales (1 = totally agree; 7 = totally agree). Reversed item was recoded and items were summarized in an additive scale ranging from 3 to 21. "I really enjoy responding to questionnaires through the mail or Internet"; "I really enjoy being interviewed for a survey"; "Surveys are interesting in themselves"	3
survey value	"Surveys are important for society"; "A lot can be learned from information collected through surveys"; "Completing surveys is a waste of time".	3
survey burden	"I receive far too many requests to participate in surveys"; "Opinion polls are an invasion of privacy"; "It is exhaustive to answer so many questions in a survey"	3
smartphone user	"Do you sometimes use a smartphone, besides when completing the questionnaires of this panel?"	2
tablet user	"Do you sometimes use a tablet, besides when completing the questionnaires of this panel?"	2
social media user	Can you indicate whether you ever spend time on the following online activities?... social network sites (like Facebook, Hyves, Myspace, Sugababes, or others)"	2
Internet use on computer (weekly hours)	"Can you indicate how many hours you use the Internet on a computer or laptop per week, on average (including emailing), besides when completing the questionnaires of this panel?" Values range from 0 to 110.	2
e-mail use (weekly hours)	"Can you indicate how many hours per week, on average, you spend on… email?" Values range from 0 to 168.	2

[a]1. Background Variables; 2. Social Integration and Leisure; 3. Personality

Five demographic variables were chosen for this analysis that are commonly used for weighting adjustments in order to determine the effect of other factors over and above demographic controls: age, gender, education, imputed family income, and urbanization level. Other demographic variables are used as proxies for the constructs of interest.

One composite measure of civic engagement derived from twelve items was chosen.

Four measures of social integration were included: a social isolation score derived from six items, social trust, marital status, and renting status. I assume that those who are married and not renting to be better integrated into their communities.

Following Abraham, Maitland, and Bianchi (2006), the number of children in the household and employment status were used as measures of busyness along with another measure related to perceived busyness: one’s satisfaction with their amount of leisure time. I expect that panelists who have low satisfaction with their amount of leisure time, have children in the household, and are employed to be busier.

Six personality measures based on multi-items scales were included: need for cognition and the big five personality traits of extraversion, agreeableness, conscientiousness, emotional stability, and openness to experience.

Three attitudes toward survey research were chosen: survey enjoyment, value, and burden. Finally, five measures of Internet use were included in this analysis: smartphone use, tablet use, social media use, weekly hours of PC Internet use, and weekly hours spent using e-mail. The possible connection between these technology measures and comfort with technology, particularly mobile devices, is the reason for their inclusion.

For a second model, I use measures collected from the “baseline” survey which are expected to be correlates of survey response. Table 3.2 shows these predictors and how they were defined for this analysis.

Table 3.2: Measures employed to predict participation decision and their definition for this analysis

Predictor	Definition
perceived usefulness of mobile Web surveys	“I would find a mobile survey useful because I could choose the time and place to respond” (“1-strongly disagree” to “5-strongly agree).
perceived enjoyment of mobile Web survey	“Filling out questionnaires using a mobile phone would be enjoyable for me” (“1-strongly disagree” to “5-strongly agree).
perceived trustworthiness and data security of mobile Web surveys	2 items on five-point scales (“1-strongly disagree” to “5-strongly agree). Items were summarized in an additive scale. The two items were: “My data would be well protected if I were to fill out a questionnaire using a mobile phone” and “Mobile surveys are trustworthy.”
compose e-mails	Do ever use your cell phone to do any of the following things? (yes vs. no)
post content on social media	
make online bank transactions	
post videos	
online shop	
fill out online questionnaires	
frequency of mobile Internet use	‘How often do you use the internet on a cell phone?” (1=never; 2= rarely, 3=some days; 4=most day; 5=every day)
duration of browsing sessions	“When you use the internet on your phone, how long is a typical browsing period before you take a break or move on to something else?” (1=Less than 2 minutes; 2=2 to 5 minutes; 3=5 to 10 minutes; 4=10 to 20 minutes; 5=More than 20 minutes)

Three attitudes toward mobile Web surveys were chosen: survey enjoyment, usefulness, and trustworthiness. Six types of smartphone activities were included: composing e-mails, posting content on social media, making online bank transactions, posting videos, online shopping, and filling out online questionnaires. Two measures of general smartphone use were selected: frequency and duration of browsing sessions.

Data analysis

A simple logistic regression model was fit to predict stated willingness (yes. vs. no) using the variables above.

Because a stratified sample of willing participants (rather than the full sample of willing panelists) was selected to participate in the subsequent surveys, selection weights were generated to be equal to the inverse of the probability of selection for each strata (stratum 1=phone owners; stratum 2=phone borrowers). These weights were used in the response models in order to make inference to the full sample of willing participants. No population weights were used, and so inference is not being made to the general population, but rather to the sample who responded to the screener survey.

Three types of response models were fit using the variables described above.

- A logistic regression model was fit to predict response (yes. vs. no) among those who expressed willingness. This aim of this model is to assess the effect of the predictors on response conditional on willingness.
- A two-stage instrumental variable approach was used to predict stated response (yes vs. nonresponse or unwillingness) using the full sample. The propensities of expressing willingness were included as an instrumental variable. The aim is to assess the effect of the predictors on response after controlling for their effects on willingness.
- A logistic regression was fit to predict response (yes vs. nonresponse or unwillingness) using the full sample. The aim is to assess the nets effects of the predictors on the full participation process (willingness and response).

For each model, two measures of fit are reported, a coefficient of determination labeled (pseudo) R^2 and a max-rescaled (pseudo) R^2 (Nagelkerke 1991), as well as a Hosmer-Lemeshow goodness of fit test (Hosmer and Lemeshow 2013).

Results

I first report participation rates. I then investigate the influence of predictors on willingness and participation using bivariate analysis and multivariate analysis.

Expressed willingness rate and response rate

Of the 5486 LISS panelists who completed the recruitment survey, 41.3% of panel members expressed willingness to participate in a mobile Web survey, while 58.7% did not want to participate[9]. The willingness rate depended on smartphone ownership: 57.5% of owners expressed willingness to do a mobile Web survey whereas only 28.7% of non-owners expressed willingness even though they were offered a phone ($\chi^2(1) = 439.9$, $p < .01$).

Of the 2263 panelists who indicated willingness, 1388 were selected randomly and sent e-mail invitations. From this group, 1024 participated in the mobile Web survey for a response rate of 73.8%. For the nonrespondents, 189 participated using an unassigned device like a PC or tablet (13.6%) and 175 failed to respond at all (12.6%). The response rate depended on smartphone use but in an unexpected way: 69.4% of those who use their own iPhone or Android phone responded to the mobile Web survey whereas 89.9% of those who used a borrowed phone responded, even though this group can be characterized as less familiar with smartphones ($\chi^2(1) = 63.4$, $p < .01$). Possible explanations for this are that those using borrowed phones were eager to try out their new device or they felt obligated to participate because the panel took the extra step of sending out devices. In addition, starting the survey may have been easier for these panelists because they had the option to automatically login by selecting a bookmark on their phone's home screen (containing an encrypted version of their login credentials). In any case,

[9] From the latter group, 2986 answered "no" to the request to participate and 125 answered "yes" but did not own an iPhone or Android phone and were unwilling to accept a loaned phone.

this suggests that providing phones is an effective, albeit expensive, way to motivate panelists to participate in a mobile Web survey.

Complete information on all predictors was available for 5,265 of the 5,486 panelists (or 96.0%) who answered the willingness question. Further analysis revealed the compliance rates among those with and without missing data (52.5% vs. 40.8%). Although the presence of any missing data was positively correlated with willingness ($\chi^2(1) = 12.0$, p <.01), cases with missing data were not included in further analyses because multivariate models are used to investigate the effects of each predictor.

Bivariate associations

As shown in the bivariate analysis in Table 3.3, there were several differences between those who did and did not express willingness and between respondents and nonrespondents (conditional on willingness). Those who are more civically engaged have higher likelihoods of expressing willingness and responding. As for the social integration variables, those who have high social trust have higher likelihoods of expressing willingness and higher likelihoods of responding. Contrary to my expectations, those who are married have lower likelihoods of expressing willingness but as expected they do have higher likelihoods of responding. The busyness variables had different effects on willingness and response; busy people (according to these indicators) have higher likelihoods of expressing willingness but lower likelihoods of responding. Several of the personality variables were significant predictors of willingness or response. For the willingness stage, those with high need for cognition, openness to experience, and extraversion have higher likelihoods of expressing willingness. Meanwhile for the response stage, those who are more agreeable and those with high emotional stability have higher likelihoods of responding. Conscientiousness has different effects on willingness and response;

conscientious people have lower likelihoods of expressing willingness and higher likelihoods of responding. Positive attitudes about surveys have positive effects on willingness to participate but no effects on responding. As might be expected, all of the technology use variables are positively associated with willingness, but unexpectedly, only one (social media use) was related to response and it had a negative association.

There were more differences at the willingness stage than the response stage. One explanation is that the power to detect such differences is enhanced in the former stage because of the larger sample size. Another explanation is that the first stage filtered out those who did not want to participate, leaving only highly motivated individuals for the second stage.

Table 3.3: Bivariate analysis for predictors of willingness to do a mobile Web survey and participation in the survey, given willingness

	Stated willingness			Survey response (given willingness)		
	Mean/%			*Mean/%*		
	Yes N=2254	No N=3011	Difference	Yes N=991	Yes N=320	Difference
Sociodemographic						
age	44.1	55.3	-11.2***	44.7	38.2	6.6***
male	48.3%	44.6%	3.6%**	50.1%	42.5%	7.6%*
education	2.8	2.3	0.5***	2.9	2.6	0.2*
imputed household income	3.0	2.9	0.1**	3.0	2.8	0.2*
urbanization level	3.1	2.9	0.1**	3.1	3.2	-0.1
Civic engagement						
civic engagement index	18.6	17.8	0.8***	18.6	17.9	0.9*
Social integration						
social isolation	7.9	7.8	0.1	7.8	8.0	-0.2
social trust	6.1	5.9	0.2**	6.1	5.8	0.3*
married	52.1%	61.0%	-8.9%***	51.2%	44.4%	6.8%*
renting	23.4%	28.5%	-5.1%***	23.9%	28.1%	-4.2%
Busyness						
satisfaction w/ leisure time	6.9	7.5	0.6***	7.1	6.6	0.5***
number of children in household	1.0	0.7	0.4***	1.0	1.4	-0.4***
employed	71.8%	50.5%	21.3***	70.4%	84.1%	-13.6%***
Personality						
need for cognition	80.5	74.8	5.7***	80.6	80.3	0.3
openness to experience	35.4	33.9	1.4***	35.5	35.1	0.4
extraversion	32.7	32.1	0.6**	32.7	33.4	-0.7
agreeableness	38.5	38.5	0.1	38.7	37.9	0.8*
conscientiousness	36.7	37.5	-0.8***	36.9	35.7	1.2***
emotional stability	34.7	34.8	-0.1	35.1	34.1	1.0*
Attitudes about surveys						
survey enjoyment	15.3	14.2	1.1***	15.3	14.9	0.4
survey value	13.5	13.2	0.3***	13.5	13.6	-0.1
survey burden	8.2	8.6	-0.5***	8.1	8.5	-0.4
Internet use						
smartphone use	59.0%	30.0%	29.0%***	63.6%	63.1%	0.5%
tablet use	47.5%	34.0%	13.1%***	47.7%	50.0%	0.2%
social media use	69.5%	42.9%	26.6%***	69.8%	77.2%	-7.4%*
computer Internet use (weekly hours)	14.0	9.2	4.8***	14.1	14.3	-0.2
e-mail use (weekly hours)	4.3	3.1	1.1***	4.0	4.6	-0.6

**p*<0.05, ***p* < 0.01, ****p* < 0.001

Multivariate models

As shown in Model 1 of Table 4, several effects became insignificant in the multivariate model. I first consider stated willingness to do the mobile Web survey. An examination of the goodness-of-fit indicated that the model fits the data adequately (Hosmer-Lemeshow test: $\chi^2(8) = 9.63$, $p = 0.29$; $R^2 = 0.21$; max-rescaled $R^2 = .28$). Further analysis detected no serious collinearity among the predictor variables: the variance inflation factor (VIF) never exceeded the cutoff value of 10 (Neter et al. 1996) in a linear regression form of the model (max VIF = 3.1) and the maximum absolute correlation between regression coefficients in the final model was .48.

In the model, willingness, remains associated with civic engagement, technology use, personality traits, and sociodemographic measures, and weakly associated with measures of social integration. Regarding the personality traits, those with higher need for cognition have higher likelihoods of expressing willingness, perhaps because they expected the mobile Web survey to be a stimulating task. Panelists who are more conscientious have lower likelihoods of expressing willingness, perhaps because they value doing what they say they will do, and in this case they were not confident that they would follow through on their commitment to participate in the mobile Web survey upon being invited. In addition, more extraverted panelists have lower likelihoods of expressing willingness. One measure of busyness is positively associated with willingness, while the other is negatively associated with it.

Some different factors may predict stated willingness among smartphone users and nonusers (e.g., openness to experience might only play a role for those who are uninitiated to smartphones). To test this, I assessed if there were two-way interactions between any of the predictors and smartphone use. I found seven significant interaction effects at the .05 level, three of which persisted when they were added to the model together. Among non-smartphones users, those who use tablets, those who are more open to experience, and those with more children living at home have higher likelihoods of expressing willingness to participate in the mobile Web survey.

To model response, I fit two different models, both of which utilized selection weights to make inference to full sample of participants. One is a model predicting response to the mobile Web survey, conditional on willingness (unweighted n = 1311). In this model, I control for wave

(1 vs. 2), which refers to the period when a participant was invited to complete the mobile Web survey. Another variables related to the design of the study, device type (loaned vs. owned), was not included as a predictor because it was found to be highly correlated with the measure of smartphone use (r =.50, p < .01). The other modeling approach is a two-stage logistic model that uses propensities of expressing willingness estimated from Model 1 as an instrumental variable to predict response versus nonresponse or unwillingness (unweighted n = 4318). This model makes use of the larger sample of panelists who completed the screener survey, even though most of them were not subsequently invited to participate in the experiment. Because the results are largely unchanged in the two-stage model, the simpler approach is presented as Model 2 in Table 4 and the two-stage model that uses willingness as a predictor variable is presented in Table D.1 in Appendix D.

Examination of the goodness-of-fit indicated that an unweighted form of the model (Model 2 in Table 4) fits the data adequately (Hosmer-Lemeshow test: $\chi^2(8) = 12.01$, $p = 0.15$). As for collinearity, the maximum VIF in an unweighted linear regression form of the model was low (2.8) and the maximum absolute correlations between regression coefficients in the final model was .52. Model fit, as measured by R^2, was lower at for the response model than the willingness model (.11 vs. .21). One explanation is that there is not much variation left to explain in the response model since the willingness stage filtered already out those who did not want to participate.

The model reveals that response propensity is moderately related to busyness, personality traits, and technology use. Specifically, those who are not employed, have fewer children living at home, are more agreeable, are less extraverted, and use smartphones have higher likelihoods of responding to the mobile Web survey. No associations were found for measures of civic engagement, social integration, or attitudes about surveys. Participants were more likely to respond when they were invited to participate in the first wave of the experiment rather than the second wave. Different factors may drive response among smartphone users and nonusers. To test this, I also assessed if there were two-way interactions between any of the predictors and smartphone use and found no significant interactions.

Some factors appear to have different effects on the two participation stages. Models 1 and 2 from Table 4 indicate that measures of civic engagement, busyness, technology use, and personality traits are related to both stated willingness and response (given willingness). By contrast, civic engagement, social integration, and attitudes about surveys are significant predictors in the willingness model but not in the response model.

I also assessed the net effect of the predictors on the full participation process (willingness and response) in a single weighted model predicting response versus nonresponse and unwillingness. It is presented as Model 3 in Table 3.4. A goodness-of-fit test for an unweighted version of this model indicated that it fit the data adequately (Hosmer-Lemeshow: $\chi^2(8) = 4.09$, $p = 0.85$). As with the other models, no serious collinearity among the predictor variables was detected; the maximum VIF in an unweighted linear regression form of the model was 3.1 and the maximum absolute correlations between regression coefficients in the final model was .49.

The moderate association between the social integration measures (marital status) and stated willingness does not persist in this final model. Several other associations do persist. Those who are more civically engaged, not employed, less extraverted, have positive attitudes about surveys, use smartphones, use social media, and spend more time online have higher likelihoods of participating in the mobile Web survey.

Table 3.4: Logistic regression models predicting willingness to do a mobile Web survey, participation in the survey (given willingness), and overall participation in the survey

	Model 1: Stated willingness		Model 2: Survey response (given willingness)		Model 3: Survey response (not controlling for willingness)	
	Est	*(SE)*	*Est*	*(SE)*	*Est*	*(SE)*
	N=5265		N=1311 (unweighted)		N=4318 (unweighted)	
Intercept	-1.399**	(0.520)	-0.210	(1.141)	-2.602***	0.665
Sociodemographic						
age	-0.039***	(0.003)	0.014*	(0.007)	-0.026***	0.004
male	0.114	(0.076)	0.272	(0.166)	0.223*	0.096
education	0.080**	(0.025)	0.095	(0.056)	0.116	0.032
inputted household income	0.112***	(0.029)	-0.029	(0.068)	0.071	0.037
urbanization level	0.022	(0.026)	-0.071	(0.057)	-0.017	0.033
Civic engagement						
civic engagement index	0.034***	(0.006)	0.012	(0.013)	0.035***	0.008
Social integration						
social isolation	0.020	(0.015)	0.009	(0.033)	0.017	0.020
social trust	0.018	(0.016)	0.050	(0.035)	0.022	0.020
married	0.191*	(0.075)	-0.047	(0.165)	0.092	0.097
renting	-0.126	(0.081)	-0.137	(0.163)	-0.110	0.104
Busyness						
satisfaction w/ leisure time	-0.048*	(0.019)	0.028	(0.037)	-0.010	0.024
number of children in household	0.007	(0.032)	-0.164**	(0.059)	-0.061	0.041
employed	-0.293**	(0.093)	-0.460*	(0.221)	-0.291*	0.124
Personality						
need for cognition	0.009***	(0.003)	-0.009	(0.006)	0.004	0.003
openness to experience	0.003	(0.009)	0.019	(0.018)	0.004	0.011
extraversion	-0.018**	(0.006)	-0.026*	(0.011)	-0.025***	0.007
agreeableness	0.003	(0.008)	0.032*	(0.015)	0.013	0.010
conscientiousness	-0.034***	(0.007)	0.014	(0.013)	-0.019*	0.009
emotional stability	0.005	(0.005)	0.005	(0.011)	0.010	0.007
Attitudes about surveys						
survey enjoyment	0.136***	(0.012)	0.025	(0.024)	0.119***	0.015
survey value	0.020	(0.020)	-0.069	(0.042)	-0.006	0.026
survey burden	-0.037***	(0.010)	-0.005	(0.023)	-0.027*	0.013
Internet use						
smartphone use	0.650***	(0.073)	0.359*	(0.160)	0.716***	0.090
tablet use	0.103	(0.069)	-0.202	(0.146)	0.003	0.088
social media use	0.475***	(0.073)	-0.181	(0.178)	0.416***	0.098
computer Internet use (weekly hours)	0.007**	(0.003)	0.005	(0.005)	0.008*	0.003
e-mail use (weekly hours)	-0.003	(0.005)	-0.010	(0.008)	-0.010	0.006
Experimental variables						
wave (=1)			0.329*	(0.138)		

*p<0.05, **p < 0.01, ***p < 0.001
Model 1: R^2 = .21; Max-rescaled R^2 = .28; Model 2: R^2 = .11, Max-rescaled R^2 = .14; Model 3: R^2 = .16, Max-rescaled R^2 = .21
Model 2: weighted n = 2147; Model 3: weighted n = 5147

Measures directly related to mobile Web use

Next, I investigate another set of measures that were collected from smartphone users that are directly related to mobile Web use. Unfortunately, I cannot test the association between these measures and stated willingness, as these variables were only available for respondents to the baseline survey. Complete information on these measurers was available for 973 of the study participants.

Bivariate associations

As shown in the bivariate analysis in Table 3.5, few of the measures appear to be strongly associated with response (conditional on willingness) and the ones that were had unexpected effects. Having ever posted content on social media, posted videos online, or shopped online using a smartphone are associated with lower response propensities. One explanation for this counter-intuitive result is that these activities are correlated with demographic characteristics (e.g., age, education) that actually account for the differences in response rates. This appears to be the case, because in the multivariate model investigating the effects of these predictors on response, all of these effects were eliminated (right-hand side of Table 3.5). The one significant predictor of response (over and above demographic characteristics) is mobile e-mail use: those who compose e-mail on their smartphones have higher likelihoods of responding to the mobile Web survey than those who do not.

Table 3.5: Bivariate and full model for predictors of participation in a mobile Web survey

	Bivariate			Full Model	
	Mean/% (SE)			*Est (SE)*	
	Yes N=716	No N=257	Difference	N=973	
Sociodemographic					
age	40.7	35.5	5.2***	0.018**	(0.007)
% male	51.0%	44.8%	6.2%	0.065	(0.168)
education	2.9	2.6	0.3**	0.120*	(0.060)
inputted personal income	3.1	2.7	0.3	-0.009	(0.071)
urbanization level	3.1	3.2	-0.1	-0.058	(0.059)
Perceptions about mobile Web surveys					
usefulness	3.5	3.5	0.0	0.021	(0.103)
enjoyment	3.4	3.4	0.0	0.066	(0.111)
trustworthiness	7.3	7.4	-0.1	-0.054	(0.058)
Activities completed using smartphone					
compose e-mails	69.7%	66.9%	2.8%	0.369*	(0.180)
post content on social media	53.8%	66.9%	-13.6%***	-0.286	(0.182)
make online bank transactions	51.1%	53.3%	-2.2%	0.148	(0.173)
post videos	17.1%	9.8%	-7.3%**	-0.406	(0.235)
online shop	16.2%	22.2%	-6.0%*	-0.263	(0.216)
fill out online questionnaires	21.2%	23.4%	-2.1%	0.083	(0.202)
General mobile Web use					
frequency of use	4.4	4.5	-0.1	0.151	(0.111)
duration of browsing sessions	2.6	2.7	-0.1	-0.050	(0.075)
Experimental variables					
wave (=1)				1.254	(0.313)
device type (=loaned)				0.297***	(0.153)
Intercept				-0.595	(0.765)

*$p<0.05$, **$p < 0.01$, ***$p < 0.001$

R^2 = .07; Max-rescaled R^2 = .10; Hosmer-Lemeshow test: $\chi^2(8) = 5.36$, $p = 0.72$

One remaining question is whether the significant predictors of response from the models presented in Table 3.4 are unique to mobile Web or not. To investigate this, I compared the response behavior of the invited panelists in mobile Web to the response behavior of those same panelists in the PC Web survey that was part of the current study, which achieved a slightly higher response rate than the mobile Web survey (81.6% vs. 73.8%, $p < .01$); this difference is due to the fact that fewer panelists used an unassigned device in the PC Web survey than in the mobile Web survey (4.1% vs. 13.6%, $p < .01$). My approach was to fit a combined (mobile Web and PC Web) model which included a random effect of subject to account for the fact that panelists were invited to take both surveys and then assess if there were two-way interactions between the predictors and survey mode. Of the 28 predictors that were considered, three of them significantly interacted with mode and can be characterized as device-specific. According to these interactions, those who use smartphones have a higher likelihood of responding in mobile Web but not PC Web. Those who use tablets, on the other hand, have a higher likelihood of responding in PC Web but not mobile Web. Younger respondents were less likely to respond in both modes, but the effects of age were smaller in mobile Web than PC Web even after accounting for other predictors, suggesting that mobile Web brings in a larger proportion of younger participants into the sample than PC Web.

Discussion

The focus of this chapter was on nonresponse that occurred in a Web survey in an online panel for individuals who were prompted to use smartphones. I separated the participation process into two components, willingness to participate and the actual participation, because of the expectation that different factors may affect each step. I then modeled stated willingness and response using social, psychological, attitudinal, and behavioral measures.

As one might expect, the results show that several factors that are known to influence the nonresponse process in other survey modes also play a role in mobile Web surveys. These include personality traits, civic engagement, and attitudes about surveys. However, other factors that are not considered in the classical nonresponse literature are important factors. These include smartphone use, social media use, and smartphone e-mail use. Still others factors, like need for cognition, appear to have a larger effect on participation in this particular mobile Web survey than has been shown for other modes, though this cannot be formally tested in this study.

The results revealed several differences between those who were willing to participate and those who were not but few differences between respondents and nonrespondents (conditional on expressing willingness). This suggests that the crux of the participation decision was at the first stage because it effectively filtered out those people who knew they were less likely to respond, leaving only highly motivated individuals to be invited to the survey. That said, measures of busyness, personality traits, and technology use were still associated with survey response. Notably, those who use smartphones, use social media, are more civically engaged, and report enjoying surveys had an especially high likelihood of responding to the mobile Web survey.

Collectively, these findings have different implications for Web surveys that prompt smartphone use and generic Web surveys that do not prompt smartphone use. In the case where smartphone use is prompted (mobile-only surveys), these results show that convincing people to participate may be a challenge, as only 58% of smartphone owners expressed willingness to do so. This highlights the fact that just because an individual uses a smartphone does not mean that they will agree to use it to take surveys; people may have very particular online tasks that perform on their phones and for some people surveys may not be one of them. Among non-owners, only 29% expressed willingness to do a mobile Web survey even though loaned phones were offered in an effort to broaden the sample base. These non-owners were also less likely to respond to the mobile Web survey. Some factors increased the likelihood that non-users would participate. For instance, those already using touchscreen mobile devices (tablets) and those who were more open to new experiences had higher chances of expressing willingness to participate.

Another implication of these results for mobile-only surveys is that a pre-screened sample (on willingness) cannot be characterized as representative of the general population. This is because of the key differences between those who were willing and unwilling to participate on such measures as civic engagement, technology use, and personality traits. Weighting adjustments on a pre-screened mobile sample would need to account for these measures in addition to sociodemographic characteristics.

In the case where smartphone use is not prompted (e.g., conventional Web surveys), the same results are more encouraging from a researcher's perspective. The fact that a minority of people are interested in completing surveys using smartphones is evidence that most of them still prefer larger-screened devices that present fewer measurement challengers (i.e., fewer breakoffs, shorter completion times). And to the extent that those who are interested in participating using a smartphone in this study are similar to those who intend to use smartphones for generic Web surveys, my results suggest the people in this group who respond may not be so different from those who don't. That said, device-specific response mechanisms like smartphone use might generate relatively large biases for some estimates.

Several other results shed light on the nonresponse process in mobile Web surveys. Mobile e-mail use was significantly associated with response. This is consistent with the findings from de Bruijne and Wijnant (2014) in the LISS panel that respondents who read e-mails on their phones are more likely to respond to a mobile Web survey. One potential reason for this is that individuals who do not use e-mail on their phone are forced to type a survey URL into a mobile browser. To prevent this, researchers might consider sending out an invitation with a clickable link via SMS. Early evidence suggests that this is an effective strategy (De Bruijne and Wijnant 2014; Mavletova and Couper 2014).

Those who are more civically engaged had a higher chance of expressing willingness and those who are more agreeable had a higher chance of responding. This suggests that some form of altruism motivated respondents because they viewed the survey as a chance to help researchers or society at large. The effectiveness of recruitment approaches for mobile Web surveys might depend on whether they can effectively appeal to this motive.

While a large proportion of individuals who are busy with other activities (according to the measures used in this analysis) expressed willingness, a smaller proportion of them participated when invited. This might be because they expect mobile Web surveys to take a relatively long time to complete due to slow transmission times, increased scrolling, usability problems, and so forth (Couper and Peterson 2015).

There are several limitations in the present study. The measures used in this analysis were subject to missing data that were correlated with both stated willingness and survey response. Since the measures were collected up to two years prior to the mobile Web survey, some of them may have become outdated. The measures directly related to smartphone use (shown in Table 5) may have influenced willingness, but the study design prevented me from testing this. Some of the measures (e.g., marital status, renter/homeowner) were only weak proxies for the constructs of interest. Finally, the results come from a crossover experiment, and so when I combine data from waves one and two, I make the assumption that response mechanisms are the same in both waves.

There is still much to be learned about the response process in mobile Web surveys. One direction for future research is to investigate the extrinsic and intrinsic motivations for responding to surveys via smartphones such as materialism, altruism, and curiosity. While researchers have several ideas about what motivates respondents to join online panels and participate in them (Brüggen et al. 2011; Keusch, Batinic, and Mayerhofer 2014), little is known about respondents' motives for participating in mobile research. Motivating factors for participation in smartphone surveys may be the incentive (especially if it is larger than for other surveys), the chance to help researchers, or curiosity about this new type methodology. It also would be worth considering how these motives change once panelists become more familiar taking mobile Web surveys.

Another topic for future research is to consider how situational factors influence participation in a mobile Web survey. For instance, does it matter if someone views the invitation to participate when they are away from home, on-the-go, or around other people? Additional research could also focus on practical approaches to either reduce nonresponse – e.g., by appealing to those who do not use smartphones – or adjust for it by implementing different weighting techniques. Finally, one could estimate nonresponse errors for several variables of interest in a mobile Web survey, as is done in the next Chapter.

Chapter 4: Decomposing Mobile versus PC Web Mode Effects into Different Error Components in a Probability Web Panel

Summary

Researchers are looking to conduct surveys that require a smartphone to participate in order to take advantage of their advanced features. But it is still an open question whether such surveys allow for inference to general populations, given that non-coverage of the population, survey nonresponse, and measurement errors can result in bias. This study is among the first to use a Total Survey Error (TSE) approach to estimate multiple sources of error simultaneously in a mobile Web survey. It reports on a repeated measures experiment conducted in a Dutch online probability panel that allows for estimation of errors as a mode effect against a PC Web survey, which serves as the benchmark. Total error (coverage, nonresponse, and measurement errors) is estimated for a range of non-demographic variables and then decomposed into its underlying components. The study finds few overall mode effects, suggesting that mobile Web is no more error-prone than PC Web for most variables. Additionally, I found no evidence whatsoever of measurement effects. However, non-coverage was a concern for several reasons: it was the largest contributor to total error, it biased over one-third of the estimates used in this analysis, and it was not consistently canceled out by nonresponse or measurement errors. This suggests that non-observation errors, not measurement errors, are the largest obstacle to the adoption of mobile Web surveys for population-based inference.

Introduction

As discussed in earlier chapters, mobile Web surveys are becoming widely used in market and social-scientific research. This is not only because respondents are choosing to use their smartphones rather than their PCs to complete Web surveys, but also because researchers are looking to conduct surveys that require a smartphone to participate. Indeed, some panels are inviting their members to complete surveys designed to take advantage of the "enhanced data collection" opportunities of smartphones (Couper 2013, p. 150), like diary apps that prompt respondents (e.g., Fernee and Sonck 2013). While most mobile-only surveys have been part of small-scale feasibility studies, others have been part of larger population-based surveys. For example, in a recent survey using address-based sampling (ABS) to construct a probability sample, Research Triangle International (RTI) sent invitations by mail and prompted respondents to download a research app (Roe et al. 2013).

But it is still an open question whether the data collected from mobile-only surveys allow for inference to general populations. Coverage, nonresponse, and measurement issues may limit who can be reached for such a survey, who will respond, and whether respondents will make reporting errors, all of which might affect the accuracy of estimates based on mobile Web surveys. Researchers have begun evaluating the impact of mobile administration on the selection process (e.g., Fuchs and Busse 2009; Chapter 3) and measurement process (e.g., De Bruijne and Wijnant 2013a; Chapter 2). Studies that investigate multiple sources of error simultaneously to identify the largest contributor to total error and inform efforts to reduce it are less common.

In this chapter, I estimate mode-specific error in mobile Web survey estimates of behavioral and attitudinal measures by using a parallel PC Web survey as a benchmark. To minimize the effect of true change between surveys (i.e. period effects), the benchmark survey was conducted first for half of the sample and second for the other half (rather than second for the full sample). I then decompose total error into mode-specific coverage, nonresponse, and measurement error components to determine whether the relative effects of mobile administration on each component are different.

Approach for evaluating errors

There are a number of commonly used designs to evaluate error in mobile Web surveys in the absence of true values. But most of them face problems disentangling selection bias from measurement error. In observational studies where respondents select the device they use, measurement effects can be estimated but selection errors cannot, because information about who was covered (i.e., who had access to smartphones) and who intended to use them but failed to respond is not readily available. In split-ballot experiments in which respondents are randomized to a device, measurement effects can be estimated but coverage errors cannot, because only smartphone users are invited to participate. Furthermore, low response rates in the mobile condition make it difficult to entirely separate selection effects from measurement effects. In record-check studies, errors can be decomposed (see e.g., Sakshaug and Kreuter 2012; Sakshaug, Yan, and Tourangeau 2010), but this approach has not yet been used in a mobile Web survey since records are not readily available or may not even exist for some measures. In test-retest studies where respondents answer the same questions in the same mode twice, the focus is entirely on measurement effects (specifically simple response variance) and not selection effects.

Another approach, and the one utilized in this study, is the reinterview survey. This type of study has a long history of use in methodological research (e.g., see Hansen, Hurwitz, and Bershad 1961). In this approach, respondents are reinterviewed in a different mode that is considered optimal or at least preferred. After conditioning on those in the sample of people who completed the benchmark survey, both selection errors and measurement errors can be estimated. For example, in a U.S. national agricultural survey, Fecso and Pafford (1988) were able to estimate error in telephone survey estimates of livestock and crop inventories because they used a face-to-face reinterview as a benchmark (as cited in Biemer and Lyberg 2003). More recently, in an experiment tied to the Dutch Crime Victimization study, Schouten et al. (2013) were able to estimate total error in a telephone, mail, and Web survey because of a face-to-face reinterview of the full sample. They then decomposed total error into components arising from selection and measurement based on models fit using frame and population registry covariates.

One limitation of the reinterview approach is related to true change between surveys. Such change makes it difficult to entirely separate period effects from measurement effects unless true values are stable over time. The current study addresses this limitation by using a balanced crossover design, wherein the benchmark survey was conducted first for half of the sample and second for the other half, rather than second for the full sample. This way, I am not forced to assume that true values are stable over time, only that change is balanced in the two sequences.

Another limitation of the reinterview approach is related to errors in the reference surveys. Such errors force one to focus on mode-specific errors rather than absolute errors. The current study, which uses a PC Web survey as reference survey, cannot address this limitation. So how do we know which estimate is better? In this study, there are several reasons to assume that PC estimates are more accurate. First, coverage errors are not a concern in PC Web in the Web panel where this research was conducted because computer Internet access is universal. Second, as reported in Chapter 3, nonresponse rates were significantly lower in the PC Web survey than in the mobile Web version of this study. Finally, measurements collected in PC Web surveys are considered valid due to the consistent finding of near-comparability between this and face-to-face surveys (Heerwegh and Loosveldt 2008; Kreuter et al. 2008), which is the mode commonly used for benchmark surveys. Of course, this is not to say that PC Web surveys are error-free, but only that statistically significant mode-specific deviations can be interpreted as a reflection of worse rather than better data quality in the mobile Web survey.

Research questions

The estimation of errors in a mobile Web survey allows me to explore a series of research questions. The first one relates to total error.

RQ1. What is the overall effect on estimates of using mobile Web as a mode of date collection relative to PC Web?

One aim is to assess whether the mobile Web survey produces different estimates than the parallel PC Web survey. Obtaining similar estimates would suggest that mobile Web surveys are becoming a viable way to augment or replace PC Web surveys. On the other hand, obtaining different estimates would raise some doubt about using mobile Web for general population surveys.

Other questions relate to individual error sources. According to Biemer and Lyberg (2003), total error under the Total Survey Error (TSE) framework can be divided into five major sources: specification error, frame or coverage error, nonresponse error, measurement error, and processing error. In this study, I focus on three of these sources – coverage, nonresponse, and measurement – because I expect their relative impact to be larger in mobile Web than PC Web surveys. By contrast, I do not expect the size of specification and processing errors to be any different for mobile Web than regular Web surveys.

RQ2. Which error source -- coverage, nonresponse, or measurement -- tends to be the largest contributor to total error?

One possibility is that coverage errors contribute the most to total mode effects. Non-coverage is a concern because there are still those who do not use smartphones and who could never participate in mobile Web surveys that require such a device to participate. Evidence suggests that coverage is related to demographics characteristics. For example, using a face-to-face survey conducted in 33 European countries (Eurobarometer), Fuchs and Busse (2009) report that mobile users are younger, more likely to be male, and more likely to be single. Such characteristics might in turn be related to important survey variables.

Another possibility is that nonresponse errors are the largest of the three components. As mentioned earlier, a consistent finding is that response rates tend to be lower in mobile Web than PC Web (Buskirk and Andrus 2014; De Bruijne and Wijnant 2013a; Mavletova 2013; Mavletova and Couper 2013; Wells et al. 2013). Nonresponse rates by themselves are not an indicator of bias. But for mobile Web surveys, respondents have been shown to differ from nonrespondents on several variables, including civic engagement (Chapter 3), attitudes about surveys (Chapter 3), mobile e-mail use (Bruijne and Wijnant 2014; Chapter 3), and frequency of mobile phone use (Mavletova and Couper 2013, 2015).

A final possibility is that mode effects will be driven mostly by measurement errors for the reasons described in Chapter 2 (technical features, use context, user characteristics). This may not turn out to be the case, however, given my findings from Chapter 2 and other recent research about response quality suggesting that mobile Web surveys do not necessarily produce lower quality responses than PCs.

An advantage of looking at these errors separately, besides the ability to compare their size, is to look at the direction in which individual errors move for a given variable.

RQ3. Do the three errors move in different directions to offset one another or move in the same direction to compound error?

One possibility is that the errors will have the same sign and compound error. There are reasons to expect this, because coverage and nonresponse errors are influenced by some of the same factors. For example, attitudes related to the perceived ease of use and usefulness of smartphones have been shown to influence actual use (Park and Chen 2007) and the intention to participate in smartphone surveys (Bosnjak et al. 2010). More general personality traits related to technology acceptance may influence both processes. In addition, social media use is likely to be correlated with both smartphone ownership (Smith 2015) and willingness to participate in mobile Web surveys (Chapter 3).

Another possibility is that coverage errors offset (or cancel out) the errors associated with nonresponse, suggesting that the factors leading to not having a smartphone are different from the factors that influence participation. There is a precedent for this in other survey modes. For instance, Peytchev, Carley-Baxter, and Black (2011) estimated both coverage and nonresponse error in a telephone study and found that they moved in mostly opposite directions. In addition, several studies have identified distinct correlates of having internet access and responding to PC Web surveys (e.g., Bosnjak et al. 2013b; Couper et al. 2007; Lee, 2006; Chapters 2-3 of Tourangeau, Conrad, and Couper 2013).

Since errors are “a property of a statistic, not a survey” (Groves 2004, p. 85), a final consideration is the relationship between individual error sources and individual survey variables. The questionnaire administered in this study contained a range of questions about politics, technology use, and behaviors and attitudes related to health and social life. Some of the items are sensitive in that they ask about potentially embarrassing behaviors (e.g., binge drinking) while others are not.

RQ4. Which estimates tend to be most affected by mobile administration?

Since this is an exploratory study, and little is known about the error profile of mobile Web surveys, my expectations are quite general. One organizing framework for a particular variable's relationship with bias is a three-group framework offered by Groves (2006). These groups were intended to describe nonresponse errors but also apply to other selection errors like coverage error. The first group is called "survey variable cause" and contains variables that directly influence selection into the sample. Estimates from variables in this group will suffer from bias, and the association between the attribute and selection will persist even after controlling for demographic characteristics.

For coverage error, a variety of variables could fall into this "survey variable cause" group. The likely causes of smartphone use include attitudes about the perceived ease of use and usefulness of smartphones (Park and Chen 2007), the financial burden of using them, and the makeup of one's peer group (Lee 2014). The current study looks at a limited set of measures but includes a few that may directly influence smartphone use. For example, those who prefer using tablets to go online or prefer using them to fill out questionnaires may be more likely to use smartphones because of their positive experience with touchscreen devices. Similarly, those who frequently eat out in restaurants or go shopping may be more likely to use smartphones because such devices enable online browsing while on-the-go.

Groves' (2006) second group is called "common cause" and contains variables that are influenced by the same factors that influence the selection mechanism. Postsurvey adjustments should be effective for reducing bias in this case, assuming that one has access to the auxiliary variables (e.g., age or education) that are related to both the selection mechanism and survey variable of interest.

The likely correlates of smartphone use include socio-demographic characteristics like age, education, race, and marital status (Fuchs and Busse 2009; Smith 2012) as well as a range attitudinal and behavior measures. Again, the variables in this study are limited but some may fall into this group because they are influenced by some of the same factors that influence smartphone use. For example, binge drinking is associated with age (Naimi et al. 2003), and age in turn is associated with smartphone use (Fuchs and Busse 2009). Similarly, attitudes about immigrants (Hainmueller and Hiscox 2007) are correlated with education, which is turn is correlated with smartphone use (Fuchs and Busse 2009). Political attitudes that are associated with age or education may also fall into this group.

The third grouping described by Groves is called “separate causes” and contains variables that are influenced by distinct factors from the selection mechanism. In the current study, estimates for variables in this group will not suffer from any coverage bias at all because the mobile Internet population and full population do not differ on the attribute being measured.

This framework can be applied to nonresponse errors. The likely causes of nonresponse include frequency of mobile phone use (Mavletova and Couper 2013, 2015), mobile e-mail use (Bruijne and Wijnant 2014; Chapter 3), attitudes toward surveys (Bosnjak et al. 2010; Chapter 3) and civic engagement (Chapter 3). Unfortunately, none of these variables are measured in the current study. However, a number of variables were collected that may be associated with variables that in turn influence response, such as household income (Mavletova and Couper 2013), age, and education (De Bruijne and Wijnant 2014). For example, political views may be affected by age which in turn affects response propensities. Satisfaction with safety from crime may be influenced by income which in turn influences survey response (Chapter 3). Or satisfaction with one's social life may be affected by busyness which also influences response propensities (Chapter 3).

Unlike the cases of coverage and nonresponse, the relationships between survey variables of interest and measurement error cannot be represented with this three-group categorization. Other frameworks may prove useful for dividing variables into groups based on their association with measurement effects. One possibility is that sensitive items will be most affected. This might be the case if when using smartphones respondents are more likely to be around other people who could potentially look over the respondent's shoulder or whose mere presence could heighten the sensitivity of certain questions. But the findings from Chapter 2 suggest that this will not be the case.

Another option is that some technology estimates will be biased in mobile Web because of psychological priming; the attributes of the input device may stimulate thoughts related to that device in a natural and instantaneous way. For example, in the LISS panel, Bruijne and Wijnant (2013) found that respondents were more likely to report that they preferred to go online using their smartphone when using such a device than when using PCs. Buskirk and Andrus (2014) found that respondents reported owning more apps when using an iPhone than when using a PC (39.1 vs. 30.1 apps). Yet another option is that measurement error related to respondents' level of motivation and effort will emerge. For example, when using smartphones, respondents are more likely to select the first acceptable response option rather than considering the full set (primacy effects) (Lattery et al. 2013; Stapleton 2013). For the questions of interest in this study, response options were not presented in random order, and so this type of satisficing behavior would affect response distributions.

Collectively, the findings from prior research suggest that mobile Web administration may impact survey errors such as coverage, nonresponse, and measurement. But more research is needed to understand the size of such errors, the relationship among errors, and their association with particular types of survey variables. While several researchers have conducted mode comparison studies involving mobile Web, no one to my knowledge has used a reinterview study to estimate total error and then decompose it into different components. This study tries to address this gap.

Methods

Research setting

As described in earlier chapters, a randomized two period, AB/BA crossover design experiment was conducted to compare two modes (mobile Web vs. PC Web). Panelists were eligible if in a screener survey they expressed willingness to participate in mobile research and either owned their own iPhone or Android phone or wanted to take a loaned phone.

The selected participants were randomized to one of two sequences of survey modes: approximately half of the participants were invited to complete the Web survey on their PC, followed by a smartphone; the other half of participants were invited to complete the Web survey on a smartphone first, followed by their PC. Data collection was carried out from October 7-October 29, 2013 for period 1 and from December 2-December 31, 2013 for period 2. Invitations to the surveys were sent by email. The same questions were asked in both modes and in both periods.

Items used

The items used in this analysis are listed in Table 4.1, where they are divided by topic into three groups: technology, lifestyle, and politics.

Each variable was dichotomized into two categories; this way, I could express errors as percentages without the need to rescale them into relative errors.

Table 4.1: Items used in this analysis, their wording, and their distribution in the benchmark sample

Relevant Response	Question	PC Web Benchmark across both periods (n = 1127)
Technology		
Prefer using tablet to go online	What is your preferred device for going online? (a mobile phone / smartphone; a personal computer; a tablet)	17.7%
Prefer using tablet to fill out questionnaires	If you could choose, which device would you prefer to use to fill out your next questionnaire? (laptop computer / desktop computer; mobile phone / smartphone; tablet)	15.4%
Hours watching TV (>3 hours)	On an average weekday, how much time, in total, do you spend watching television?	45.9%
Lifestyle		
Satisfied with social life	How satisfied are you with your social life? (not at all satisfied; not too satisfied; somewhat satisfied; very satisfied)	92.6%
Satisfied with family life	How satisfied are you with your family life?	93.2%
Satisfied with your pace of life	How satisfied are you with your pace of life?	85.8%
Satisfied with your safety from crime	How satisfied are you with the feeling of safety where you live?	94.9%
Exercise less than once per week	In a typical week, about how often do you exercise? (less than 1 time per week; 1 or 2 times per week; 3 times per week; 4 or more times per week)	82.2%
Binge drank in past 30 days	On how many days did you have 5 or more drinks of an alcoholic beverage on the same occasion?	41.6%
Ever driven while intoxicated	Have you ever, in your entire life, driven a car (or other motor vehicle) when you were (at least a little) intoxicated? (yes; no)	42.1%
Eat in restaurants at least once per month	During the past 12 months, how many times did you eat in restaurants?	32.9%
Go shopping at least once a week	During the past month, how many times did you go shopping?	32.6%
Politics		
Immigrants make country a worse place	In your opinion are the Netherlands made a worse or better place to live by immigrants coming to live here? (1 = much worse; 10 = much better)	67.3%
Feel favorably towards George W. Bush	Please indicate how favorable or unfavorable you feel toward the following person or organization by entering a number between **0 and 100 [**dichotomized as 0-49 (unfavorable) and 50-100 (favorable]. Former U.S. President George W. Bush	13.2%
Feel favorably towards Barack Obama	U.S. President Barack Obama	63.4%
Feel favorably towards Mark Rutte	Mark Rutte (current Prime Minister of the Netherlands)	35.7%
Feel favorably towards Jan Peter Balkenende	Jan Peter Balkenende (former Prime Minister of the Netherlands)	31.9%
Feel favorably towards CDA	Christian Democratic Appeal (Political party in the Netherlands on the center to center-right of the political spectrum)	22.6%
Feel favorably towards VVD	People's Party for Freedom and Democracy (Political party in the Netherland that is considered socially liberal and economically conservative)	27.2%

Decomposition of errors

As mentioned earlier, the two-mode, two-wave experiment made available PC Web answers that could be used to estimate error in the mobile Web survey for range of survey variables. Data are combined across the two waves and errors were estimated as a mode effect against the benchmark PC Web survey. Throughout this chapter, I purposefully refer to these errors as "deviations" rather than as “biases.” This is because a deviation does not necessarily indicate only bias. Rather these deviations are a function of both true bias and sampling variation. However, I can still investigate bias by conducting significance tests on the estimated deviations. If a deviation is significantly different from zero, this is evidence of bias over and above sampling variance, though the power to detect such differences in enhanced at the coverage stage (because of larger sample sizes) compared to the subsequent stages (because of smaller sample sizes). To simplify the challenge of isolating errors sources, my focus is on descriptive statistics rather than analytic ones (e.g., correlations, regression coefficients).

Coverage error is estimated by computing the difference between the proportion of interest measured in PC Web based on the covered sample and the proportion measured in PC Web based on the full sample:

$$Deviation_{coverage} = p^{PC}_{covered} - p^{PC}_{benchmark} \tag{1}$$

In this notation, the superscript refers to the mode in which the variable was measured [PC Web (“PC”) vs. Mobile Web (“mobile”)], and the subscript refers the subgroup that the statistic is based on [benchmark sample (“benchmark”), covered sample (“covered”), or responding sample (“respond”)].

The covered sample consists of those individuals who reported owning an internet-enabled smartphone in the screener survey conducted two months before the start of the crossover experiment. It should be noted that in the actual implementation of this experiment, those panelists who did not use smartphones were offered a loaned smartphone in order to achieve full coverage. While panelists who used such devices are considered in the analysis of response quality in Chapter 2 and nonresponse in Chapter 3, they are counted as being uncovered here (and therefore are not considered in subsequent estimation of nonresponse or measurement errors).

The benchmark sample consists of those who used PCs to complete the PC Web survey. Those who use a different device (detected from user agent strings) or did not respond at all are considered to be nonrespondents. There is no benchmark information available on these particular measures for nonrespondents to the PC Web survey, so this group cannot be included in the analytic sample, i.e., inclusion in the sample is conditional on responding to the PC Web survey (see Chapter 3 for an examination of nonresponse that uses prior wave LISS data). Error in this analysis, then, can be conceptualized as either a mode effect against the compliant PC Web sample or as a mode effect against the full PC Web sample under the assumption that PC Web nonrespondents are missing completely at random[10].

Nonresponse error is estimated by computing the difference between the proportion of interest measured in PC Web based on the responding sample and the proportion measured in PC Web based on the covered sample:

$$Deviation_{nonresponse} = p_{respond}^{PC} - p_{covered}^{PC} \quad (2)$$

[10] This presumes that the cause of missingness is separate from the survey variable of interest, and is not correlated with the errors that are estimated in the mobile Web survey. This is a strong assumption, and one that is not supported: nonrespondents to the PC Web survey were reliably older (*t*(1386) = -9.22, *p* <.01) and more likely to be male (*t*(1386) = -2.92, *p* <.01) than respondents.

The responding sample consists of those individuals who used smartphones to complete the mobile Web survey. Those who use a different device (detected from user agent strings) or did not respond at all are considered to be nonrespondents.

Measurement error is estimated by computing the difference between the proportion of interest measured in mobile Web based on the responding sample and the proportion measured in PC Web based on the same sample:

$$Deviation_{measurement} = p_{respond}^{mobile} - p_{respond}^{PC} \quad (3)$$

The overall mode effect can represented in two ways. One is the difference between the proportion of interest measured in mobile Web based on the responding sample and the proportion measured in PC Web based on the benchmark sample. Another is as the sum of the three error estimates described in Equations 1, 2, and 3.

Assumptions

In addition to assuming that the PC Web survey is the preferred data collection mode, I make two assumptions described by Klausch et al. (2014) that are necessary to estimate errors using a reinterview design. The first one relates to the stability of true values over time. If one mode followed the other for all cases, then I would have to assume that true values are time-stable. Otherwise change over time would confound the estimation of errors. The advantage of using a balanced crossover design, however, is that the assumption can be weaker. I only assume that true values change in the same way for participants in both mode sequences (those who take the PC Web survey first and those who take the mobile Web survey first). Since participants were randomly assigned to these sequences, this seems like a reasonable expectation.

In addition, I make an assumption about the stability of the response mechanism over time. Like before, I do not have to assume that there is stability over time. I only assume that response propensities change in the same way for participants in both mode sequences. No such assumption needs to be made about coverage because this indicator was measured a few months before the start of experiment; so in this analysis, it did not change between waves (even though it could have in reality, e.g., if someone purchased a smartphone during that time).

Implicit in these assumptions is that are also no carryover effects (Johnson 2010) on nonresponse or measurement due to mode (i.e., that the experience of answering in one mode did not affect participation or response in the other mode); the one month lag between waves is a way to minimize such effects at least for measurement errors, because it is long enough to reduce the chances that respondents remember and repeat their previous responses.

Significance testing

Error estimates are considered to be statistically significant if their 95% confidence intervals do not contain zero.

The confidence intervals were obtained using two different approaches. One was to use a closed formula to estimate standard errors and then add and subtract 1.96 standard errors from the estimates. More information about these formulas will be presented later on.

The other was to use bootstrap resampling. This involved drawing repeated with-replacement samples (n = 1000) of the same size as the full sample (n = 1127), repeatedly computing the bias measure of interest, and then assessing the distribution of these bias estimates (with the 2.5th and 97.5th percentiles serving as the upper and lower limits of the confidence interval).

For two variables that were found to have either relatively large coverage bias or nonresponse bias, I compared the resulting confidence intervals from the closed formula approach and bootstrap approach. As shown in Table 4.2, the two approaches yielded nearly identical results across all of the bias measures for these two variables. This suggests that the assumptions that underlie the closed formula approach, which will be described later on, are reasonable, and so for simplicity this approach was used rather than a bootstrap.

Table 4.2: Estimated 95% confidence intervals for different error components using closed formula approach and bootstrap approach.

Survey variable	*Approach for estimating SE*	Confidence interval for coverage deviation	Confidence interval for nonresponse deviation	Confidence interval for measurement deviation	Confidence interval for total deviation
Binge drank in past 30 days (%)	*closed formula*	**2.9 – 6.4**	-3.5 – 0.6	-0.6 – 6.4	**2.3 – 9.9**
	bootstrap	**2.9 – 6.3**	-3.5 – 0.7	-0.5 – 6.8	**2.6 – 10.1**
Immigrants make country a worse place (%)	*closed formula*	-1.0 – 2.5	**-4.6 – -0.1**	-4.2 – 2.2	-6.4 – 0.2
	bootstrap	-1.3 – 2.5	**-4.6 – -0.1**	-4.4 – 2.2	-6.6 – 0.3

Does not contain zero (**bolded**)

Closed formulas for standard errors

For coverage error estimates, standard errors were estimated using the following formula adapted from Lee (2006, p. 465):

$$se(p_{covered}^{PC} - p_{benchmark}^{PC}) = \left(\frac{n_{benchmark} - n_{covered}}{n_{benchmark}}\right)\sqrt{var(p_{covered}^{PC}) + var(p_{notcovered}^{PC})}\ , \qquad (4)$$

where there are $n_{benchmark}$ panelists in benchmark sample and $n_{covered}$ panelists in the covered sample, and $p_{notcovered}^{PC}$ is the estimated proportion based on those who are not covered. This assumes that $cov(p_{covered}^{PC}, p_{notcovered}^{PC}) = 0$ and that $n_{covered}$ is a fixed quantity rather than a random variable.

For nonresponse error estimates, standard errors were estimated using the same framework:

$$se(p^{PC}_{respond} - p^{PC}_{covered}) = (\frac{n_{covered} - n_{respond}}{n_{covered}})\sqrt{var(p^{PC}_{respond}) + var(p^{PC}_{notrespond})}\,, \tag{5}$$

where there are $n_{respond}$ panelists who responded to the mobile Web survey, and $p^{PC}_{notrespond}$ is the estimated proportion based on those who were covered but did not respond. This assumes $cov(p^{PC}_{respond}, p^{PC}_{notrespond}) = 0$ and that $n_{respond}$ is a fixed quantity rather than a random variable.

The variance of $p^{PC}_{covered}$, $p^{PC}_{notcovered}$, $p^{PC}_{respond}$, and $p^{PC}_{notrespond}$ are estimated using the variance formula for a simple random sample.

For measurement error estimates, standard errors were estimated using the formula associated with McNemar's test for comparing dependent proportions (McNemar 1947):

$$se(p^{mobile}_{respond} - p^{PC}_{respond}) = \sqrt{\frac{p^{PC}_{12}(1-p^{PC}_{12}) + p^{PC}_{21}(1-p^{PC}_{21}) + 2(p^{PC}_{12} \times p^{PC}_{21})}{n_{respond}}}\,, \tag{6}$$

where p^{PC}_{12} and p^{PC}_{21} are measures of response changes: p^{PC}_{12} is the proportion of respondents who recorded a *yes* answer in PC Web after recording a *no* answer in mobile Web, and p^{PC}_{21} is the proportion of respondents who recorded a *no* answer in PC Web after recording a *yes* answer in mobile Web[11].

For overall error estimates, standard errors were estimated using the same framework as in Equations 4 and 5:

$$se(p^{mobile}_{respond} - p^{PC}_{benchmark}) = (\frac{n_{benchmark} - n_{respond}}{n_{benchmark}})\sqrt{var(p^{mobile}_{respond}) + var(p^{PC}_{notrespond,notcovered})}\,, \tag{7}$$

[11] The subscripts refer to the off-diagonal elements in a 2 × 2 frequency table cross-classified according to responses in mobile Web and responses in PC Web for the same participants.

where and $p^{PC}_{notrespond,notcovered}$ is the estimated proportion based on those who did not respond and were not covered, assuming $cov(p^{mobile}_{respond}, p^{PC}_{notrespond,notcovered}) = 0$ and that $n_{respond}$ is a fixed quantity rather than a random variable.

Results

Of the 1127 LISS panelists in the benchmark sample, 800 (71.0%) owned a smartphone and were therefore counted as covered and 327 did not (29.0%). Out of the 800 panelists from the benchmark sample who were covered, 589 (73.6%) responded to the mobile Web survey upon being invited and 211 did not.

As shown in Table 4.3, I estimate proportions for each variable of interest in four different ways: the first column presents the proportions measured in PC Web based on the benchmark sample; the next column presents proportions measured in PC Web based on the covered sample; the third column presents proportions measured in PC Web based on the responding sample; and the fourth column presents proportions measured in mobile Web based on the responding sample.

I then estimate errors by computing the difference between different combinations of columns. The impact of non-coverage, which presented in the fifth column, is assessed by comparing columns 1 and 2. The impact of nonresponse, presented in the sixth column, is assessed by comparing columns 2 and 3. The difference between columns 2 and 3 reflects the impact of nonresponse, which is presented in the seventh column. The overall mode effect, presented in the last column, is estimated by comparing columns 1 and 4.

For example, according to the benchmark sample, 41.6% of respondents report binge drinking in the past 30 days. The corresponding estimate jumps up to 46.3% among the panelists who own smartphones, resulting in a coverage deviation of 4.6% (4.7% if using the rounded estimates). The estimate comes back down to 44.8% among the panelists who participated in the mobile Web survey, resulting in a nonresponse deviation of -1.4%. The final estimate jumps back up to 47.7% when it is based on respondents' answers recorded on smartphones rather than PCs, producing a measurement deviation of 2.9%. The estimated mode effect, which is based on the benchmark sample estimate (41.6%) and mobile Web survey estimate (47.7%), is 6.1%.

Noncoverage was the largest contributor of error for that particular estimate. Others were more affected by non-response. For example, based on the benchmark sample, 67.2% of respondents report that immigrants make the country a worse place. The corresponding estimate is 68.0% among the panelists who own smartphones, resulting in a coverage deviation of 0.7% (0.8% if using the rounded estimates). Based to the responding sample, 65.2% of respondents report that immigrants make the country a worse place, resulting in a nonresponse deviation of -2.8%. The final estimate is 64.2%, resulting in a measurement deviation of -1.0% and an overall mode effect of -3.1%.

Table 4.3: Distributions of survey variables across subgroups (and standard errors) and total error decomposed into coverage, nonresponse, measurement components (and standard errors)

	(1) Benchmark *n* = 1127	(2) Covered *n* = 800	(3) Responded: PC answers *n* = 589	(4) Responded: smartphones answers *n* = 589	Estimated coverage deviation	Estimated nonresponse deviation	Estimated measurement deviation	Total deviation
	(*SE*)	(*SE*)	(*SE*)	(*SE*)	(*SE*)	(*SE*)	(*SE*)	
Technology								
Prefer using tablet to go online (%)	17.7 (1.1)	19.1 (1.4)	18.0 (1.6)	20.2 (1.7)	**1.4** (0.7)	-1.1 (0.9)	2.2 (1.4)	2.5 (1.6)
Prefer using tablet to fill out questionnaires (%)	15.4 (1.1)	16.3 (1.3)	15.8 (1.5)	16.3 (1.5)	0.8 (0.7)	-0.5 (0.8)	0.5 (1.5)	0.9 (1.4)
Hours watching TV (>3 hours) (%)	45.9 (1.5)	42.0 (1.7)	42.1 (2.0)	43.1 (2.0)	**-3.9** (0.9)	0.1 (1.0)	1.0 (1.7)	-2.8 (1.9)
Lifestyle								
Satisfied with social life (%)	92.6 (0.8)	93.5 (0.9)	94.4 (0.9)	94.6 (0.9)	0.9 (0.5)	0.9 (0.6)	0.2 (1.1)	1.9 (1.0)
Satisfied with family life (%)	93.2 (0.8)	93.3 (0.9)	93.2 (1.0)	93.7 (1.0)	0.1 (0.5)	-0.0 (0.5)	0.5 (1.0)	0.6 (0.9)
Satisfied with your pace of life (%)	85.8 (1.0)	86.6 (1.2)	87.1 (1.4)	86.9 (1.4)	0.8 (0.7)	0.5 (0.7)	-0.2 (1.4)	1.1 (1.3)
Satisfied with your safety from crime (%)	94.9 (0.7)	95.0 (0.8)	95.1 (0.9)	94.9 (0.9)	0.1 (0.4)	0.1 (0.5)	-0.2 (0.9)	0.1 (0.9)
Exercise less than once per week (%)	82.2 (1.1)	81.8 (1.4)	81.7 (1.6)	82.3 (1.6)	-0.4 (0.7)	-0.1 (0.8)	0.7 (1.4)	0.2 (1.5)
Binge drank in past 30 days (%)	41.6 (1.5)	46.3 (1.8)	44.8 (2.1)	47.7 (2.1)	**4.6** (0.9)	-1.4 (1.1)	2.9 (1.8)	**6.1** (1.9)
Ever driven while intoxicated (%)	42.1 (1.5)	42.5 (1.7)	45.0 (2.1)	45.7 (2.1)	0.4 (0.9)	**2.5** (1.0)	0.7 (1.6)	3.5 (1.9)
Eat in restaurants at least once per month (%)	32.9 (1.4)	36.6 (1.7)	36.5 (2.0)	36.2 (2.0)	**3.7** (0.8)	-0.1 (1.0)	-0.3 (1.9)	3.2 (1.9)
Go shopping at least once a week (%)	32.6 (1.4)	30.3 (1.6)	30.2 (1.9)	33.8 (2.0)	**-2.3** (0.9)	-0.0 (1.0)	3.6 (2.2)	1.2 (1.8)
Politics								
Immigrants make country a worse place (%)	67.3 (1.4)	68.0 (1.7)	65.2 (2.0)	64.2 (2.0)	0.7 (0.9)	**-2.8** (0.9)	-1.0 (1.6)	-3.1 (1.7)
Feel favorably towards G.W. Bush (%)	13.2 (1.0)	14.6 (1.3)	14.8 (1.5)	13.8 (1.4)	**1.4** (0.6)	0.1 (0.7)	-1.0 (1.4)	0.5 (1.3)

Feel favorably towards B. Obama (%)	63.4 (1.4)	64.0 (1.7)	66.6 (1.9)	66.4 (1.9)	0.6 (0.9)	**2.6** (1.0)	-0.2 (2.0)	3.0 (1.9)
Feel favorably towards M. Rutte (%)	35.7 (1.4)	38.3 (1.7)	39.6 (2.0)	39.6 (2.0)	**2.6** (0.9)	1.3 (1.0)	-0.0 (1.9)	**3.9** (1.8)
Feel favorably towards J.P. Balkenende (%)	31.9 (1.4)	32.5 (1.7)	35.1 (2.0)	33.4 (1.9)	0.6 (0.9)	**2.6** (0.9)	-1.7 (1.9)	1.5 (1.7)
Feel favorably towards CDA (%)	22.6 (1.2)	22.3 (1.5)	23.3 (1.7)	23.9 (1.8)	-0.4 (0.8)	1.0 (0.9)	0.7 (1.6)	1.3 (1.6)
Feel favorably towards VVD (%)	27.2 (1.3)	30.1 (1.6)	30.9 (1.9)	34.0 (2.0)	**3.0** (0.8)	0.8 (1.0)	3.1 (1.8)	**6.8** (1.7)
Summary measures								
Number of significant bias estimates	--	--	--	--	8	4	0	3
Average absolute deviation (%)	--	--	--	--	1.5	1.0	1.1	2.3

Statistically significant (**bolded)**

I next use these results to answer my main research questions.

RQ1. What is the overall effect on error of using mobile Web as a mode of date collection relative to PC Web?

As shown in Table 4.3, the mobile Web survey produces significant mode effects (rightmost column) for a minority of questions. Out of the 19 questions, only 3 of them yield significant effects. Estimates of binge drinking, support for the current Prime Minister (Mark Rutte), and support for the VVD party appear to be inflated in mobile Web relative to the benchmark. The average absolute deviation for all variables (bottom row of rightmost column) is 2.3%, which is quite small.

RQ2. What error source -- coverage, nonresponse, or measurement -- tends to be the largest contributor to total error?

The average absolute deviation for each error source across all of the variables is shown in the final row of Table 4.3. Non-coverage appears to be the largest contributor to total error. The average absolute non-coverage deviation is 1.5% compared to 1.0% for nonresponse and 1.1% for measurement. Evidence of coverage bias over and above sampling variance comes from the number of deviations that are significantly different from zero. According to significance tests, 8 of the estimates (or 42% of them) suffer from non-coverage bias and 4 of the estimates (or 21% of them) are biased by nonresponse, whereas none of the estimates are biased by measurement effects.

RQ3. Do the three error sources move in different directions to offset one another or move in the same direction to compound error?

It is possible that errors would offset one another, resulting in a relatively small total error. While this did occur for some variables, there appears to be no systematic pattern. The error from non-coverage was offset by nonresponse for 8 variables (and compounded by nonresponse for the remaining 11 variables). Because nonresponse error failed to consistently cancel out the error associated with non-coverage, the factors that influence smartphone adoption are likely to be different from the factors that influence participation. There was little reason to expect that measurement effects would consistently cancel out error from either non-coverage or nonresponse. In line with this expectation, the error from coverage was offset by measurement for 12 variables and the error from nonresponse was similarly offset by measurement for 12 variables (and compounded for the other 7 variables). Since none of the measurement error deviations were significantly different from zero, their sign relative to other errors may be coincidental (a function of sampling error) and not a systematic pattern related to measurement effects.

RQ4. Which estimates tend to be most affected by mobile administration?

As mentioned earlier, the relation between the survey variables of interest and selection errors (i.e., coverage and nonresponse) can be represented by three models described by Groves (2006).

For coverage, the variables associated with the 11 estimates that were not biased according to significance tests would fall under the *separate causes* group while the 8 variables that were biased would fall in either the *common causes* or *survey cause* groups. To further investigate, I fit a model predicting coverage (having mobile Web access vs. not having access) with the 8 affected variables and the following demographic controls commonly used in postsurvey adjustments: age, gender, education (1. primary school, other; 2. junior high school; 3. high school degree; 4. vocational degree; 5. higher vocational degree; and 6. university degree or more), marital status, presence of children living at home, renting status, and urbanization level (1. not urban; 2. slightly urban; 3. moderately urban; 4. very urban; and 5. extremely urban.). It is presented as Model 1 in Table 4.4. Two of the associations between survey variables and coverage persist in the multivariate model, suggesting that they fall into the *survey cause* group because bias for these two attributes cannot be accounted for by statistical adjustments using demographic variables. These variables are related to tablet use and eating out. In line with my expectation, the affected variables could conceivable influence smartphone usage in a direct way.

Similarly for nonresponse, the 15 variables that were not biased according to significance tests would fall under the *separate causes* model while the 4 variables that were biased would fall in either the *common causes* or *survey cause* model. Using the same strategy that was used to model coverage, I fit a model predicting response to the mobile Web survey with demographic controls along with the 4 variables that suffered from nonresponse bias. It is presented as Model 2 in Table 4.4. Two of the variables – attitudes about immigrants and support for former Prime Minister Jan Peter Balkenende – remain significant after controlling for demographic differences. It is unclear why these variables would influence nonresponse directly, but they may be correlated with auxiliary variables that were not included in the model.

Table 4.4: Logistic regressions to predict smartphone ownership (Model 1) and participation in a mobile Web survey conditional on smartphone ownership (Model 2)

	Model 1: Coverage (yes=1; no=0)		Model 2: Survey response, conditional on coverage (yes=1; no=0)	
	Est	*SE*	*Est*	*SE*
Intercept	3.926***	0.463	-0.700*	0.298
Demographic characteristics				
Age (continuous)	-0.083***	0.007	-0.150	0.080
Gender: male vs. female	0.211	0.160	-0.180	0.171
Highest degree (low to high)	0.096	0.055	-0.302	0.179
Married: yes vs. no	0.208	0.178	0.141	0.231
Children in household: yes vs. no	-0.286	0.183	0.126	0.168
Renter: yes vs. no	-0.076	0.195	-0.150	0.080
Urbanization (low to high)	0.083	0.062	-0.180	0.171
Survey variables biased by non-coverage				
Prefer using tablet to go online	0.480*	0.207		
Hours watching TV (>3 hours)	-0.007	0.160		
Binge drank in past 30 days	0.268	0.170		
Eat in restaurants at least once per month	0.448*	0.179		
Go shopping at least once a week	-0.134	0.163		
Feel favorably towards George W. Bush	0.122	0.262		
Feel favorably towards Mark Rutte	0.061	0.203		
Feel favorably towards VVD	0.350	0.230		
Survey variables biased by nonresponse				
Binge drank in past 30 days			0.215	0.174
Immigrants make country a worse place			0.418*	0.190
Feel favorably towards Barack Obama			-0.197	0.180
Feel favorably towards Jan Peter Balkenende			-0.391*	0.197
n	1126		799	
R^2	0.226		0.038	
Max-rescaled R^2	0.323		0.056	

*p<0.05, **p < 0.01, ***p < 0.001
Note: one case has a missing value for age and was excluded from both models.

Lastly, I consider measurement error. According to significance tests, none of the estimates are biased by measurement effects. It seems that when using smartphones, respondents could give similar answers to these survey questions, whether sensitive or not and whether related to the topic of technology or not.

Discussion

This chapter assessed the mode-specific errors in a mobile Web survey conducted in a Dutch probability panel using a TSE perspective. This was done by way of a parallel

PC Web survey that served as a benchmark so that deviations from it could be regarded as error. A balanced crossover design was used (rather than conducting the PC Web survey after the mobile Web survey for all respondents) so that effects due to the time periods for conducting the surveys did not contaminate the mode comparison. Total error was estimated and then decomposed into three different parts – non-coverage, nonresponse, and measurement – to evaluate the size of these components, the relationship among them, and their association with particular types of survey variables.

The main contribution of this study to the growing literature about mobile Web surveys is to evaluate multiple sources of error simultaneously. This produced results that were both encouraging and discouraging about the viability of mobile Web for making inference to general populations. I found few overall mode effects, suggesting that mobile Web-based methods are no more error-prone than PC Web methods for most variables. Additionally, I found no evidence whatsoever of measurement effects. In other words, conditional on coverage and response to the survey, respondents appeared to have no difficulty recording their answers on smartphones rather than PCs. This finding of measurement equivalence between smartphone and PCs is largely consistent with the results from observational studies (Bosnjak et al. 2013a; Lugtig and Toepoel 2014; Toepoel and Lugtig 2014), split-ballot experiments (Baker-Prewitt 2013; Mavletova 2013; Zahariev et al. 2009) and Chapter 2 of this dissertation.

On the other hand, these results could serve as a warning about the potential for selection effects in mobile Web surveys. In spite of high coverage rates (71%), I still find eight variables that are biased by non-coverage; two of which continue to be associated with non-coverage over and above demographic control variables. Furthermore, such errors are not consistently canceled out by the other two error sources. While non-coverage appears to be the most pernicious error source, nonresponse also poses a problem for some estimates. In spite of high response rates (74%), I still find four variables that are biased by non-response; two of which continue to be associated with nonresponse over and above demographic control variables. Collectively, this evidence of selection errors is likely to generate some doubt among panel researchers about the viability of using mobile Web for general population surveys because even if weighting adjustments are used one is likely to obtain different estimates for a subset of variables.

The TSE framework is useful, in part, because of its implications for survey practice – it can help to identify the largest contributor to total error and inform data collection strategies to minimize such error. In this application of the TSE framework, survey researchers doing mobile research with limited resources should focus their efforts on reducing coverage errors as opposed to improving response rates (e.g., increasing the incentive or shortening the survey) or improving measurement (e.g., eliminating sensitive questions, sending out tablets). One obvious design change to reduce non-coverage is to provide smartphones for panelists who don’t already have them, which is analogous to the commonly used strategy in probability Web panels of providing PCs to those who don’t have them. While this would be costly, it would eliminate coverage errors by definition and enable panels to at least occasionally invite their members to complete mobile Web surveys that are designed to take advantage of the advanced features of smartphones. That said, it could increase measurement error by bringing in those panelists who are unfamiliar with smartphones (though I found no evidence to support this in Chapter 2). Another option outside of a panel setting is to only conduct mobile

Web surveys on samples where larger proportions of people own smartphones like undergraduate college student samples.

There are several limitations of this study. The survey contained a narrow set of topics, and some of the selected variables were positively correlated (e.g., general tablet use and tablet use to fill out questionnaires; satisfaction with social life and satisfaction with family life). The PC Web nonrespondents do not appear to be missing completely at random and so it is more appropriate to the conceptualize the deviations from this benchmark as a mode effect against the compliant PC Web sample rather than the full PC sample. The error estimates are mode-specific and not absolute, and so errors estimated against a survey that used a different data collection mode (e.g., face-to-face interviews) would yield different error estimates. Because of sample size differences, the power to detect bias via significance testing of the error estimates is enhanced at the coverage stage compared to the subsequent stages. The significant deviations are interpreted as bias because of the preferred coverage and response properties of PC Web surveys, though it is possible that some deviations actually reflect improved data quality in the mobile Web survey. Finally, the sample being studied was self-selected based on their willingness to participate in this experiment and cannot be characterized as representative of the full panel (see Chapter 3). This is apparent by looking at the rates of smartphone use in the sample. The observed coverage rate for this sample (71%) is substantially higher than that for the general population in the Netherlands (60%) (Statistics Netherlands 2013). Nevertheless, the main finding that selection errors were the largest contributor to total error should generalize to other non-panel settings where selection errors may have an even larger impact due to lower coverage rates and higher nonresponse rates than the ones observed in this setting.

Several key questions regarding the accuracy of estimates from mobile Web surveys remain unanswered. The impact of weighting adjustments, not the impact of controlling for demographics in multivariate models, is unexplored. The impact of mobile Web on other statistics such as correlations and regression coefficients has not been explored, nor has its impact on variance or mean-square error. There is little information about the accuracy of mobile Web surveys conducted outside of online panels, though one can imagine a wave of research being conducted on this topic in the near future. Finally, the association of errors with different question topics is still ambiguous. There would be value in future research that considers a wide range of topics and then attempts to categorize them into groups based on their relationship with difference error sources. This would allow researchers to anticipate which variables in a future survey might be biased.

In sum, the findings in this chapter suggest that moving beyond split-ballot experiments and towards efforts to evaluate bias using the TSE framework will be useful for evaluating the accuracy of mobile Web surveys. Given the rapid growth in smartphone ownership and the on-going changes in how people use such devices, more research is needed on the size and interplay of different error sources in mobile Web surveys. This will guide the choice of computing device for researchers interested in collecting high-quality data who now must consider the trade-offs of the enhanced data collection opportunities of smartphones and the selection errors that such devices may introduce.

Chapter 5: Conclusion

Summary

This dissertation represents the first attempt to take a comprehensive look at the error properties of mobile Web surveys using a TSE perspective. There were several findings that shed light on how smartphones in surveys affect results.

When using smartphones, respondents who participated in this study were at least as likely to provide conscientious answers and at least as likely to disclose sensitive information as when using PCs. And they did this while doing more other things and while around more other people than when using PCs. They also typed longer answers to an open-ended question than when using PCs; additional research is needed to determine whether (1) respondents inferred the requested amount of information not from the absolute size of the text box but from its size relative to their device's screen size and (2) whether they are starting to use features that facilitate text entry like auto-complete and voice-to-text typing.

While I saw no evidence that respondents' context of use had an impact on the quality of their answers, I did see an effect of the technical features of phones (small screen size and touchscreen input). When using smartphones, respondents had trouble using their fingers to accurately move a small-sized slider bar and date picker wheel to the intended values.

While measurement differences were small, nonresponse was a larger concern because convincing panelists to participate in this study was a challenge. About 58% of smartphone owners and only 29% of non-owners expressed willingness to participate, even though loaned phones were offered. Those who are more civically engaged, have positive attitudes about surveys, and are high in need of cognition, among other things, had a higher chance of expressing willingness to participate in “mobile research.” These differences might be easier to address, or at least identify, if they were the same ones observed in other surveys that use different data collection modes. However, some differences appear to be specific to mobile Web. Those who use smartphones, mobile e-mail, and social media were also more likely to participate.

As for coverage, smartphone owners were different from non-owners in ways that would produce bias in mobile-only surveys for several of the variables measured in this study. This coverage bias was not confined to a specific set of topics. It affected variables ranging from political attitudes to binge drinking to TV viewing.

Practical implications

These results have practical implications for researchers conducting mobile Web surveys. My finding of near-comparability in the quality of responses between mobile and PC users, even for sensitive and burdensome questions, suggest that smartphone users should be accommodated by using mobile-optimized software for Web surveys. Otherwise, those who only use mobile devices to go online may participate at lower rates. That said, researchers might want to refrain from using some interactive question formats in order to prevent mode effects. It is yet to be seen whether the negative effects of slider and pickers will occur for other input tools.

The finding that several sample person characteristics were significantly associated with the decision to participate raises questions about the representativeness of the pre-screened samples used for methodological research in this area.

Related to this, the finding of substantial coverage and nonresponse errors for some estimates in a mobile-only survey suggest that such surveys cannot be used to make inference to general populations unless special efforts are made. These efforts might include providing phones to those who don't have them, developing approaches to reduce nonresponse (e.g., decreasing the length of survey, offering a larger incentive), or adjusting for nonresponse using the response mechanisms identified in Chapter 3.

The future of mobile Web

Looking across the landscape of mobile research and thinking about the state of the world five years from now, it is my view that a distinction should be made between the types of research now being conducted and the next wave of potential mobile research. The former involves administering a traditional Web questionnaire containing many questions in a mobile browser. It also involves treating PC Web as the primary mode of data collection and mobile Web as a complementary mode that is adapted to the main mode and not necessarily used to its fullest potential. This type of research was the focus of this dissertation research.

In the next wave of research, by contrast, I anticipate that mobile Web will be treated as the primary research method, and researchers will try to leverage the advanced features of smartphones and the unique ways that people use them. Some of these opportunities are described below:

- *Surveys in non-research apps*. This will involve administering surveys in the screen location where advertisements are displayed in general mobile applications (gaming apps, news apps, etc.). The general advantage of this approach is that it avoids the nonresponse issues related to asking respondents to download special research apps.
- *New types of incentives.* Rather than motivating people to participate via a cash or lottery incentive, new incentives might motivate people to participate by allowing them reach their goal in a particular mobile app. For example, respondents might be offered the chance to answer survey questions rather than having to wait or pay money when trying to do something like watch an online video or move to the next level of a game.
- *Short questionnaires.* Questionnaires with relatively few questions may be better suited for the different ways that people use smartphones compared to other devices, which is in short bursts while doing things like waiting in line or taking a short break from the task at hand. To supplement data collected from short questionnaires, researchers can add passively collected data (e.g., GPS coordinates, health metrics). In addition, researchers can increase sample sizes by inviting many more people to participate. For example, there is the possibility that an in-app survey request can be distributed across a network of apps to be viewed by hundreds of thousands of potential respondents in a matter of minutes. This presents new possibilities and new limitations. While the old constraint was on sample size – it was expensive to contact and convince new sample members to participate, so once someone agreed to do so, it was cost-effective to administer

many questions at once – the new constraint in mobile research will be on questionnaire length because most people use their phones for brief periods at a time.

- *Utilizing passively collected data for survey research.* Auxiliary data like GPS coordinates could dictate the timing of a survey invitation. In Chapter 2, I treated mobility as potential threat to data quality, but it could also be viewed as an opportunity to initiate location-based surveys. In consumer research, for example, a restaurant might use an app to trigger a question about food quality when someone uses that app inside of their restaurant; or a hairspray company might use an app to ask whether someone decided to use hair products before leaving the house when they use that app outside of their home location. Auxiliary data could also be used to modify the content of a question. Finally, auxiliary data (e.g., whether someone is a light or heavy app user, whether someone travels frequently) could potentially be used to make weighting adjustments to survey data collected via a smartphone.

Comparability between little and big devices

Because the next wave of mobile research will involve not only in-app surveys but also traditional Web surveys that are completed using the full gamut of computing devices, an important question relates to how to achieve comparability between little devices and big devices. Dillman's *universal mode design* or *unimode* design principles (Dillman 2000), which attempt to minimize differences across modes, provide one notion of comparability. To implement these principles for Web surveys, one must display the exact same question formats across devices (i.e., which means an "unoptimized" mobile questionnaire). Researchers who are in favor of this approach make two arguments: (1) any change in question formatting will introduce mode effects (2) mobile optimization will encourage people to use their phones, and when using phones, respondents provide lower quality responses than when using PCs. However, both of these arguments run contrary to the results of this dissertation research. I found no evidence of mobile vs. PC Web mode effects, even though the mobile questionnaire used different question formatting because it was optimized for small screens. In addition, for conventional question formats (i.e., not widgets), I found no evidence that when using phones respondents provide lower response quality than when using PCs.

Another notion of comparability, and one that I would argue for, is a *best practices* approach (Tourangeau et al. 2013). This is a variant of de Leeuw's "generalized mode design" (2005, p. 248). Here, the emphasis is not on presenting the exact same surface-level features, but on utilizing the best practices of each mode in order to present the same stimulus to the respondent and minimize error within each mode. To carry out these principles for mobile Web research, one must play to the strength of smartphones by utilizing new input tools that are touchscreen friendly (e.g., wide buttons, page turns via swiping) and avoid using any formats that are not user-friendly on small screens (grids, horizontal displayed response options). In addition, one must avoid certain features from PC Web questionnaires whose analog in mobile Web may produce suboptimal results because it is not easily optimizable. There are several formats that if optimized for small screens may introduce mode effects. For example, a drop box in a PC Web survey that is displayed as a spin wheel in a mobile Web survey may lead to input errors; a grid in a PC Web survey that is displayed as single items in a mobile Web survey may weaken the correlation between responses to the items in a scale; and long horizontal response scales in a PC Web survey that are displayed vertically in a mobile Web survey could lead to differential question order effects. To treat both modes as equal, then, one must create the PC Web survey with the mobile survey in mind, and vice versa.

In order to create mobile Web questionnaires according to best practices, one must of course know the attributes of question formats that make them effective on small touchscreens. More empirical research is needed in this area. In addition, with the growing use of large-sized smartphones (iPhone 6 plus, Samsung Galaxy S6) and tablets, a "fluid design" that adapts a questionnaire to the continuum of different screen sizes by wrapping content from the right side of the screen to the bottom of the screen is more appropriate than binary optimization. An interesting question relates to the nature of fluid design parameters – should they be based on what is considered subjectively optimal by respondents or what is considered optimal in objective terms (e.g., buttons should always be larger than two centimeters wide to avoid input errors). Another question relates to whether researchers should use one general layout with scalable features (e.g., radio buttons that become larger so that they are easier to touch on small screens but are still radio buttons) or use varying layouts with different design elements (e.g., radio buttons that turn into wide touchscreen-friendly buttons for screens that are smaller than a particular threshold or cut point). More research is needed to understand the tradeoffs associated with different fluid design principles.

Will mobile research continue to grow?

As we move to the next wave of mobile research, another question is whether the growth in smartphone ownership will translate into the same level of growth in people's willingness to complete surveys on such devices. It is too early to say and there are competing views on the matter. One is that willingness will remain at relatively low levels. This is because large proportions of people perceive mobile Web surveys to be burdensome based on their experience doing other online tasks that they equate with surveys (e.g., using a smartphone to fill out online forms). In addition, the vast majority of Internet users have access to other devices that may be viewed as better suited for such tasks. Under this view, mobile penetration rates could reach 100%, but the proportion of mobile survey starts will continue to lag far behind.

The other view is that the majority of Web survey completes will soon come from mobile users. The reasoning is that as mobile interfaces improve (more user-friendly, easier text input), mobile devices improve (lighter, better connectivity, longer battery life), and mobile questionnaires improve (better optimization, shorter), more people will want to participate in mobile surveys. In addition, the rate of mobile-reliant users may continue to grow. Whether mobile research ever becomes the predominant method for data collection, then, depends on technological advances, people's perceptions about mobile surveys, their access to other devices, and whether researchers can find more effective ways to design mobile questionnaires.

In any case, it is reasonable to assume that mobile Web surveys are here to stay and are unlikely to fall out of favor and go the way of e-mail surveys and personal digital assistant (PDA) surveys.

Further research/Research agenda

Given all of this, there are several topics that I want to investigate in my future research.

- *In-app survey invitations*. Is it feasible to administer surveys in a non-research app? Is it possible to increase motivation by asking them to complete a survey in order to reach their goal in a particular site?
- *Best practices for mobile questionnaire design*. What are the attributes of question formats that make them effective on small touchscreens? And what are the broad design parameters (e.g., size and position of navigation buttons, paging vs. scrolling layout, visual design features) that make questionnaires easier to complete on smartphones?
- *Comparability between devices*. Which PC Web question formats are *optimizable* in that their mobile Web analog constitutes the same stimulus to the respondent?
- *Fluid design*. Should fluid design be used to create one questionnaire that is scalable or several questionnaires that contain different design elements? How can fluid layout parameters be informed by not only objective design principles but also by user experience?
- *Situational factors and nonresponse.* Does respondents' context of use (location, time of day, busyness), in addition to their fixed characteristics, affect their likelihood of responding to mobile Web surveys?
- *Enhanced surveys and nonresponse*. What factors influence the decision to download a research app or allow researchers to passively collect data from one's phone?

- *Multitasking.* In this dissertation research, respondents were more likely to multitask when using smartphones than when using PCs. But I relied on self-reports of multitasking rather than objective measures and had no knowledge about the type of secondary tasks completed, which research conducted in a lab setting might better be able to address. Several unanswered questions remain about the nature of multitasking on smartphones. While completing surveys, what is the prevalence of multitasking on the device versus not on the device? What impact does each type of multitasking have on response effort?
- *Mobility.* For researchers who conduct either Web surveys on smartphones or telephone surveys by calling people on their mobile phones, the issue of mobility is unlikely to go away. In this dissertation research, respondents were more likely to be away from home when using smartphones than when using PCs but I could only look at the effect of mobility on response quality in a crude way (by comparing when respondents who were away from home vs. at home based on self-reports). Future research is needed to look at how different away-from-home locations, which could be detected based on passively-collected data, affect response quality.
- *Modular design*. When answering surveys using a smartphone, what is the length of respondents' attention span and does this depend on the complexity of the survey task? Will people agree to complete a series of short surveys on their smartphones?

- *Response time.* What factors (respondent characteristics, item level characteristics, context of use) are responsible for the increase in response time in mobile Web relative to PC Web?
- *Responsive e-mail design*. What are the attributes of e-mail invitations that encourage participation in mobile Web surveys?
- *Mixed mode designs.* What is the feasibility of using cell random-digit-dial (RDD) sampling methods to generate a sample and then invite smartphone users to complete a Web survey in lieu of or in addition to a telephone interview? In one implementation of this strategy, Hu and Dayton (2014) administered some questions using computer-assisted telephone interviewing (CATI) and then switched to mobile Web for questions about alcohol and drug use by sending respondents a text message invitation.
- *Mode choice.* Is it possible to increase motivation by offering respondents a choice of several modes that operate through their smartphone including mobile Web, SMS text, and telephone? Because this approach allows respondents to choose a mode to fit their situational needs, it has the potential to improve data quality compared to when mode is assigned (Conrad et al. 2013).
- *Tablets*. How does tablet use affect response quality in surveys relative to mobile Web use? What are the general causes of nonresponse among tablet users?

Final discussion

The reaction among survey researchers to the mobile revolution has been mixed. On the one extreme, researchers have been quite wary of mobile Web as a tool for gathering information. As Callegaro (2010) and Macer (2012) point out, some researchers have ignored respondents' mobile Web use, requested that they switch devices before starting their Web surveys, or even blocked them.

At the other end of the continuum, researchers are embracing this shift by actively accommodating mobile users and trying to take advantage of the advanced capabilities of smartphones for collecting new types of data. In just the past few years, researchers have use them to collect location data (e.g., Olson and Wagner 2013; Roe et al. 2013), visual data (e.g., Link 2013; Jones et al. 2013), and health data via bluetooth sensors (e.g., De Nazelle et al. 2013). And social scientists will likely find other innovative ways to use these devices for social measurement in the future.

In my judgment, the moderate view may be the most tenable. It is clear that the Luddite vision of turning back the clock to a time of PC-only use will simply not work given that smartphones play an increasingly important role in many people's lives. But the survey landscape has not changed so much, and smartphone use is not so universal, that we can entirely replace traditional Web methods with smartphone-based ones either. Thus, rather than having an impulsive reaction towards either the old or the new, mobile Web surveys should be viewed as another tool for data collection whose effectiveness depends on a variety of factors including the particular target population being studied, the relationship between the survey variables of interest and selection errors, and the design of the questionnaire.

In addition, we cannot rely on tips for how to conduct mobile Web based on intuition, as sometimes the very features that seem well-suited for smartphones (e.g., sliders) turn out to be problematic. I believe the path forward requires continued efforts to empirically investigate the error properties of this emerging mode. As an overarching goal, these efforts should be both forward-looking and cautious, and should seek to inform new strategies that take advantage of the opportunities provided by smartphones but still generate high-quality data.

Appendix A: Questionnaire

1 A – D. First we would like to know how satisfied you are with different aspects of your life.

How satisfied are you with your social life?
How satisfied are you with your family life?
How satisfied are you with your pace of life?
How satisfied are you with the feeling of safety where you live?

- not at all satisfied
- not too satisfied
- somewhat satisfied
- very satisfied

The following statements are about common, daily activities.

2. During the past 12 months, how many times did you eat in restaurants?

3. During the past month, how many times did you go shopping?

4. In a typical week, about how often do you exercise?

- Less than 1 time per week
- 1 or 2 times per week
- 3 times per week
- 4 or more times per week

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

6. Do you have any hobbies? If so what are these?
If you do not have any hobbies, then leave this question blank.

The next five questions are about food and drinks.

7. What is your favorite fruit?

- oranges
- grapefruits
- apples
- pears
- bananas
- grapes
- strawberries
- blueberries
- tangerines
- plums
- none of the above

8. What is your favorite vegetable?

- green beans
- broccoli
- kale
- carrots
- spinach
- other, that is … Please note your favorite vegetable.

9. Which of the following nutrients is **most important** to you when selecting breakfast cereal?

- protein
- carbohydrates
- sugar
- fat
- fiber
- vitamin A
- vitamin C
- calcium
- iron
- vitamin E
- none of the above

10. Think about the past 30 days. On how many days did you have **5 or more** drinks of an alcoholic beverage on the same occasion?

11. Have you ever, in your entire life, driven a car (or other motor vehicle) when you were (at least a little) intoxicated?

- yes
- no

Next we'd like to know about your feelings toward some people and organizations in the news.

12 A – F. Please indicate how favorable or unfavorable you feel toward the following person or organization by entering a number between **0 and 100**. **0 = very unfavorable; 100 = very favorable.** How favorable or unfavorable would you rate…

Mark Rutte
Jan Peter Balkenende
Christian Democratic Appeal (CDA)
People's Party for Freedom and Democracy (VVD)
U.S. President Barack Obama
Former U.S. President George W. Bush

13. People from many different countries settle in the Netherlands. In your opinion are the Netherlands made a worse or better place to live by immigrants coming to live here? Please enter a number between **1 and 10**. 1 = much worse; 10 = much better.

The next few questions are about technology.

14. Overall, how satisfied are you with [*one half of participants*: the PC that you use most often / *other half:* the smartphone that you use most often? Please enter a number between **1 and 10**. **1** = **not at all satisfied; 10** = **extremely satisfied.**

15. What is your preferred device for going online?

- a mobile phone / smartphone
- a personal computer
- a tablet

16 A – E. For each of the following adjectives, please indicate how well it describes [*one half of participants*: the PC that you use most often / *other half*: the smartphone that you use most. Please enter a number between **1 and 10**. **1** = **not at all; 10** = **completely.**
Useful
Friendly
Powerful
Fun
Reliable

17. What is your current age? To answer, drag the black ball to your, approximate, age. The scale runs from 1 to 100 years. You can also tap anywhere on the bar.

18. What is your date of birth? Right-click on the circle to open the “date picker”.
Here you see three wheels, one for day, one for month and one for year.
Grasp the first roll with your finger and drag it to the correct day. Do the same for month and year.
Click to save. We ask you this question to compare your answer with our existing data.

In the next section, please do your best to answer the following puzzles.

19. A bat and a ball cost €1,10 in total. The bat costs €1,00 more than the ball. How much does the ball cost?

20. If it takes 5 machines 5 minutes to make 5 devices, how long would it take 100 machines to make 100 devices?

21. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

Finally, we would like to ask about your experience with the survey.

22. What type of device did you use to complete this survey?

- laptop computer
- desktop computer
- mobile phone / smartphone

23. Where did you complete this survey? *Check all that apply.*

- at home
- at work, not at home
- indoor, not at home or at work
- in transit
- outside, not at home or at work

24. Have you moved around while completing this survey?

25. Would you say you have been alone for the entire survey?

26. About how many other things have you done while completing this survey?
Please include activities like watching TV, eating, drinking, or checking email.

- 0

- 1
- 2
- 3
- 4
- 5 or more

27. Were you ever distracted by the things going on around you while completing this survey?

28. If you could choose, which device would you prefer to use to fill out your next questionnaire?

- laptop computer / desktop computer
- mobile phone / smartphone
- tablet

29 A – E. Finally; what did you think of this questionnaire?

Was it difficult to answer the questions?
Were the questions sufficiently clear?
Did the questionnaire get you thinking about things?
Was it an interesting subject?
Did you enjoy answering the questions?

- 1 certainly not
- 2
- 3
- 4
- 5 certainly yes

30. Do you have any remarks about this questionnaire?

Appendix B: Model diagnostics

- LMM predicting overall disclosure

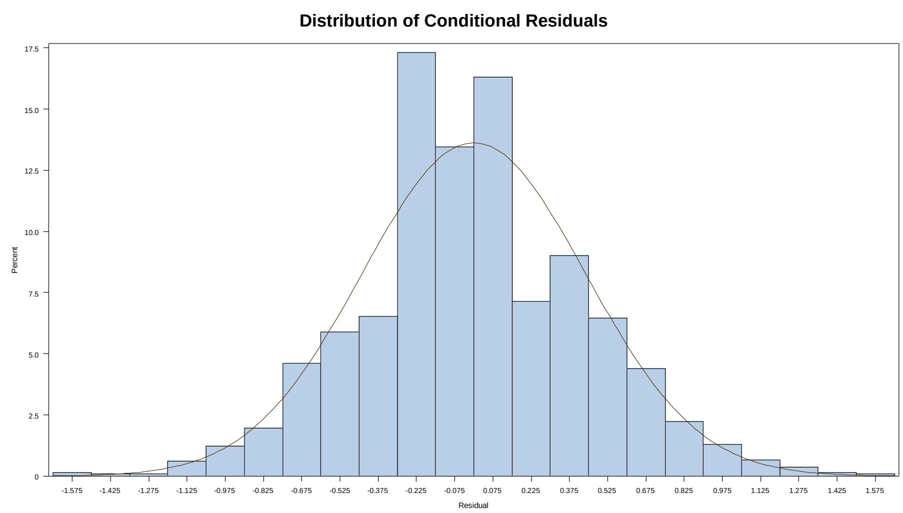


Figure B.1: Distribution of conditional residuals

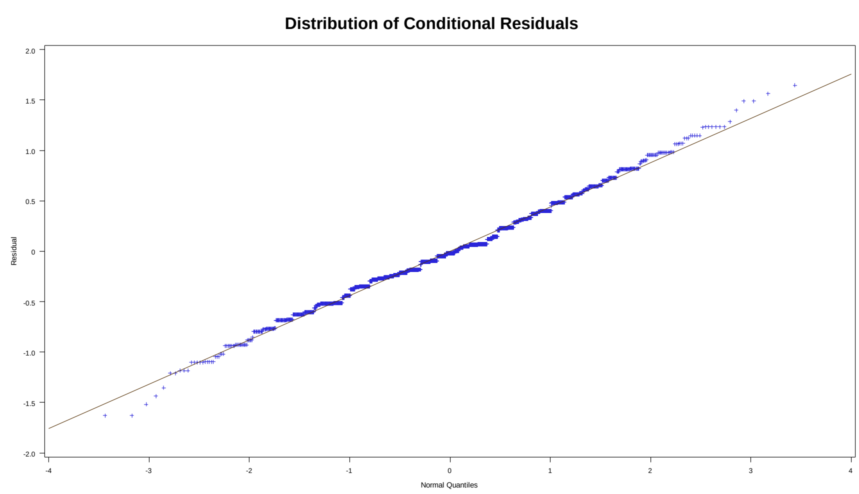


Figure B.2: Quantile-quantile plots of conditional residuals

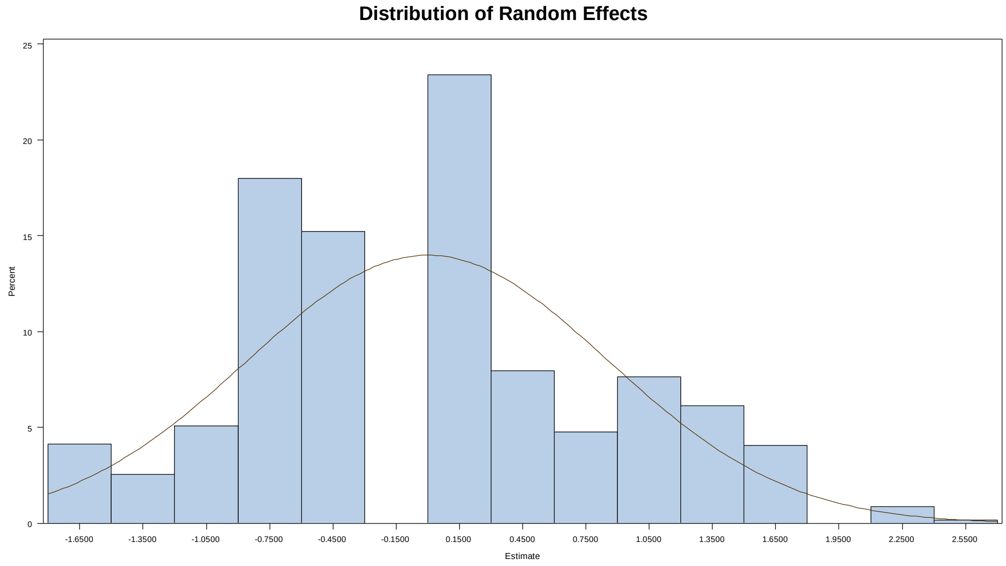


Figure B.3: Distribution of random effects

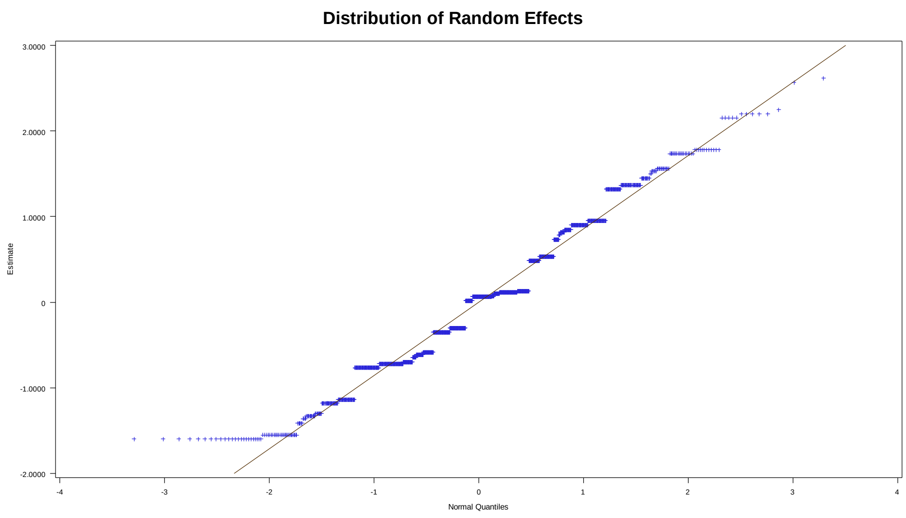


Figure B.4: Quantile-quantile plots of random effects

- GEE model predicting short open-ended answers

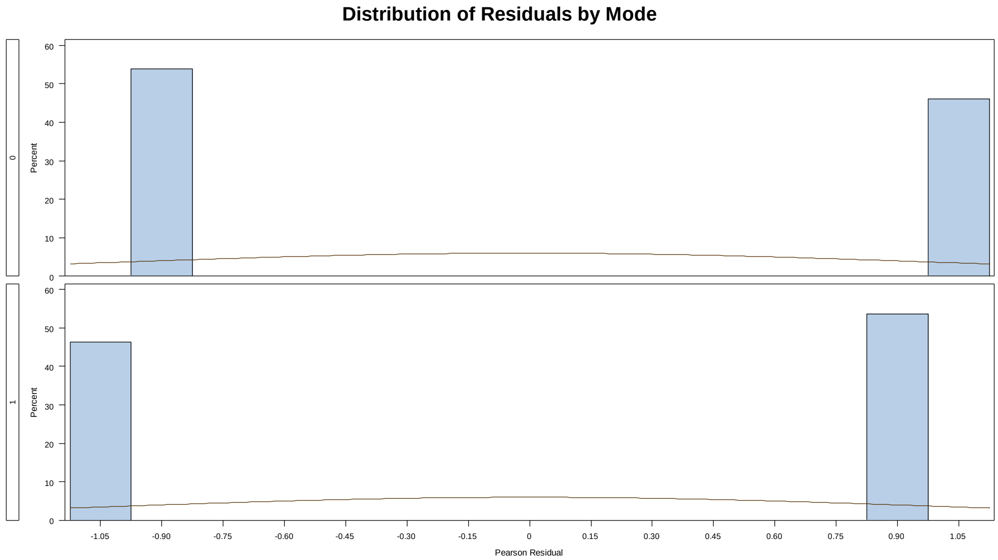


Figure B.5: Distribution of Pearson standardized residuals for observations in PC Web (top) and mobile Web (bottom).

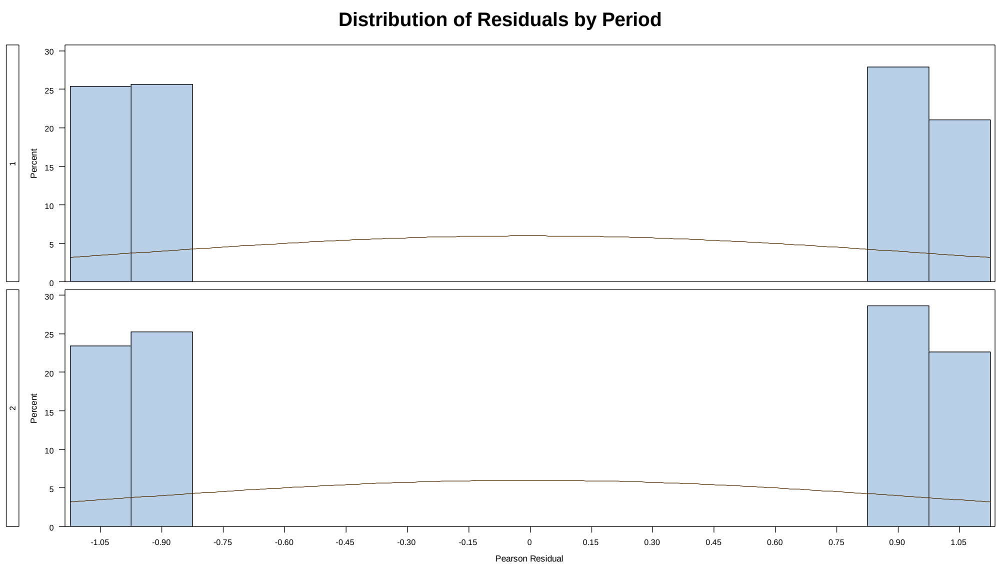


Figure B.6: Distribution of Pearson standardized residuals for observations in period 1 (top) and period 2 (bottom).

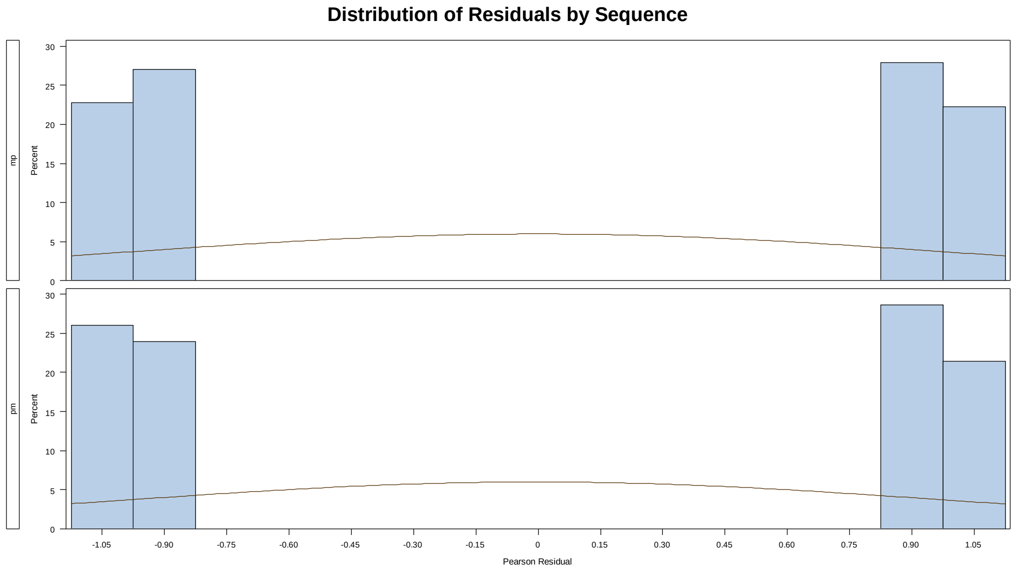


Figure B.7: Distribution of Pearson standardized residuals for observations in sequence B (top) and sequence A (bottom).

- GEE model predicting admission of ever driving while intoxicated

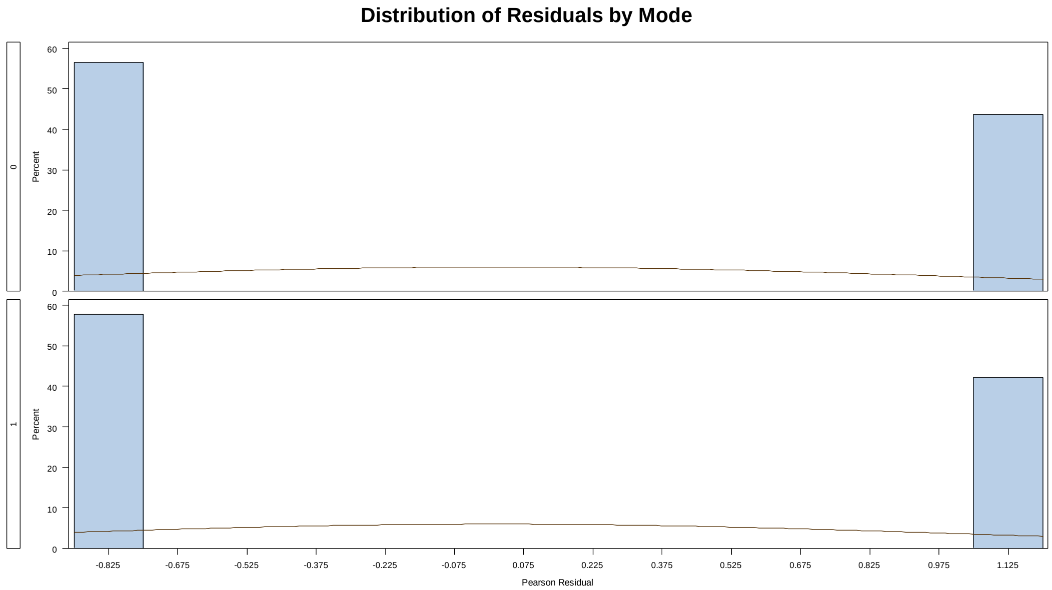


Figure B.8: Distribution of Pearson standardized residuals for observations in PC Web (top) and mobile Web (bottom).

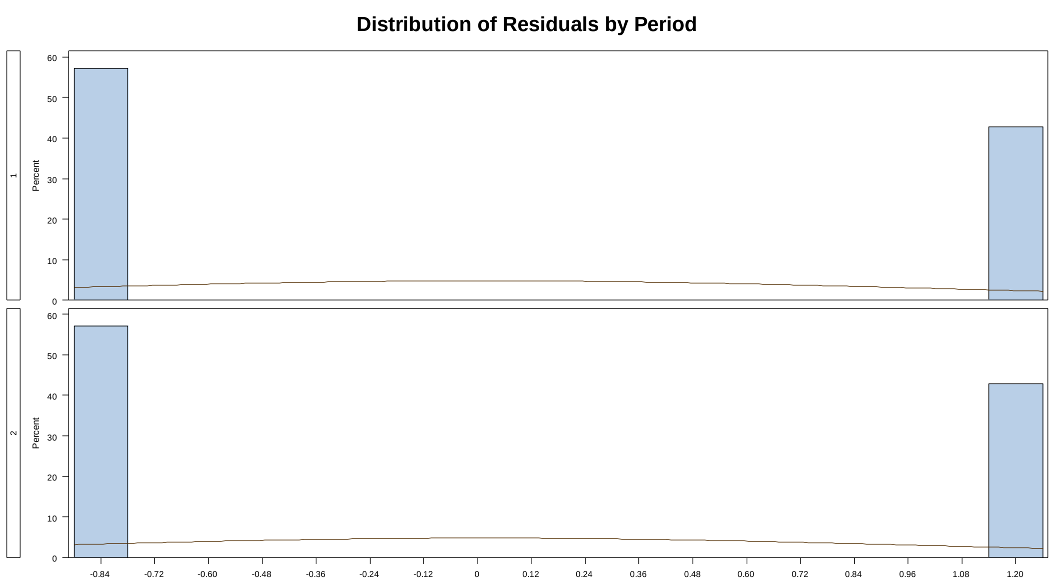


Figure B.9: Distribution of Pearson standardized residuals for observations in period 1 (top) and period 2 (bottom).

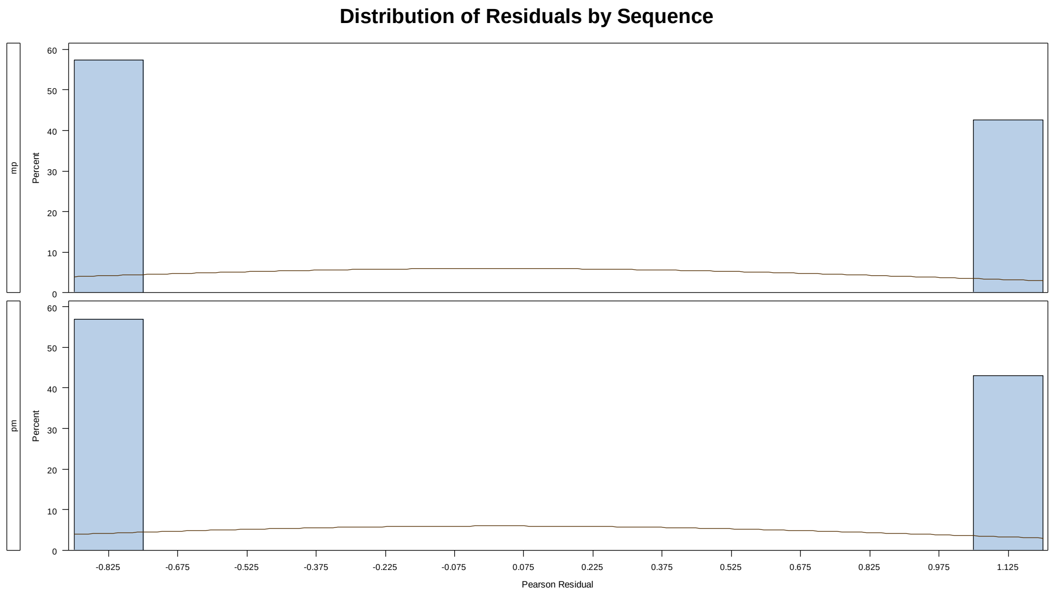


Figure B.10: Distribution of Pearson standardized residuals for observations in sequence B (top) and sequence A (bottom).

Appendix C: The relationship between willingness and participation

As shown in Figure C.1, there are at least three different theoretical frameworks that can be used to show the influence of predictors on willingness and participation. I highlight the limitations of the first two approaches and the advantages of the third approach.

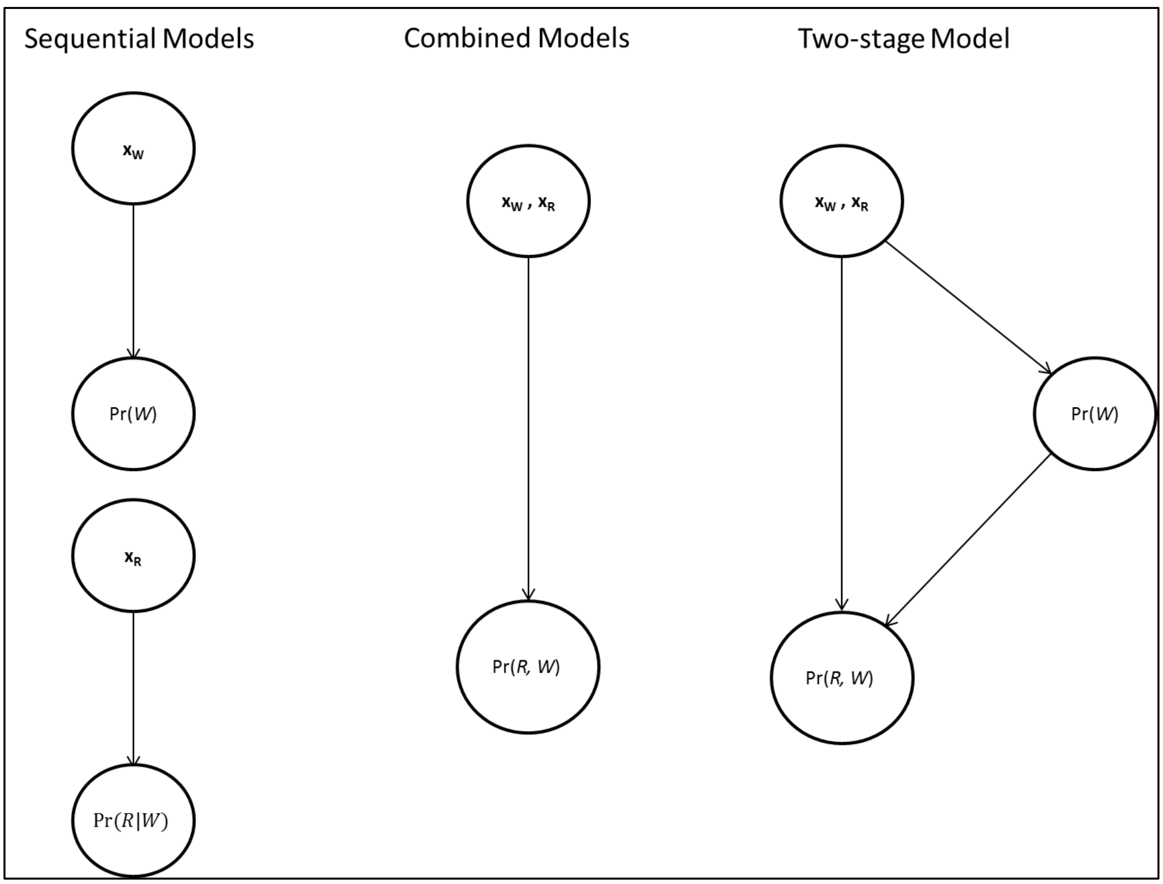


Figure C.1: Three theoretical models showing the influence of predictors on willingness and participation

In what I am calling the "sequential models" approach, a vector of covariates expected to be related to willingness denoted by $\mathbf{x_W}$ affect the likelihood of expressing willingness given by Pr(W), and a group of covariates expected to be related to response denoted by $\mathbf{x_R}$ influence the likelihood of participating given willingness $\Pr(R|W)$. This theoretical model suggests that willingness and response can be modeled separately. It is assumed that people first decide whether they are willing to participate or not which can be modeled with a simple logistic regression; it then focuses exclusively on those who are willing and models their response decision using another simple logistic regression. But this forces one to make the assumption that response is independent of the likelihood of expressing willingness, which is a strong assumption given that the intention to perform a behavior often predicts the behavior itself. Furthermore, in situations where doing a survey requires having first expressed willingness to do so, the model predicting response may not be very informative. For example, it might be the case that there are several differences between those who express willingness to participate and those who don't. But then only highly motivated individuals are selected for the survey so there is not much variation in response behavior left to explain. For example, in a PC Web survey, Couper et al. (2007) and Bosnjak et al. (2013b) report few differences between respondents and nonrespondents, contingent on willingness.

The second approach is shown as the "combined model" of association between covariates and response. In this model, a vector of covariates expected to be related to willingness and response ($\mathbf{x_W}$, $\mathbf{x_R}$) affect the likelihood of response given by Pr(R, W). The link between $\mathbf{x_W}$ and willingness is not specified. This suggests that a single model can be fit to predict response versus nonresponse and unwillingness; the latter two categories could be separated into two categories for a multinomial regression or combined into one reference category for a simple logistic regression. With either type of dependent variable, one cannot distinguish the influences on response from the influences on willingness, and so their effects on the two participation processes are confounded. For example, when trying to interpret a significant predictor, there is no information available in the model about whether that predictor influences only willingness, only response, or both processes.

The third framework is shown as the "two-stage model." Here, a vector of covariates expected to be related to willingness denoted by $\mathbf{x_W}$ affect the likelihood of expressing willingness given by $\Pr(W)$. These willingness propensities in turn affect the likelihood of response given by $\Pr(R, W)$. In addition, a group of covariates expected to be related to response denoted by $\mathbf{x_R}$ influence the likelihood of participating given by $\Pr(R, W)$. This has two advantages over the other approaches. One is that it takes into account the fact that this later process (response) may be related to the former one (willingness), and so the assumption that the two processes are independent does not have to be made. The other advantage is that it accounts for the fact that some predictors may have a unique effect on each process. For analysis, this approach suggests that willingness propensities should be estimated and used as an instrumental variable in a model predicting response.

Appendix D: Two-stage instrumental variable analysis

The simple approach of modeling survey response conditional on willingness is presented here as Model 1 in Table D.1 (earlier it was presented as Model 2 in Table 3.4). The alternative modeling strategy that accounts for the correlation between likelihood of expressing willingness and participation in the survey is presented as Model 2 in Table D.1. In this model, propensities of expressing willingness were estimated and included as a continuous predictor. The coefficients in this model represent the effect of response net of willingness.

Like in the simple model, smartphone use and extraversion are significant predictors of response in the two-stage version of the model. Agreeableness, measures of busyness, and age are no longer significant predictors of response, and education emerges as a significant predictor. As might be expected, the propensity of expressing willingness is also a significant predictor of response which follows the "two-stage model" approach presented in Figure C.1 in Appendix C.

Table D.1: Logistic regression models to predict participation in the survey using willingness as an instrumental variable

	Model 1: Survey response (given willingness)		Model 2: Survey response (controlling for willingness)	
	Est	*(SE)*	*Est*	*(SE)*
	N=1311 (unweighted)		N=4318 (unweighted)	
Intercept	-0.210	(1.141)	-3.229	(0.714)
Sociodemographic				
age	0.014*	(0.007)	-0.006	(0.009)
male	0.272	(0.166)	0.177	(0.098)
education	0.095	(0.056)	0.080*	(0.035)
inputted household income	-0.029	(0.068)	0.007	(0.046)
urbanization level	-0.071	(0.057)	-0.027	(0.033)
Civic engagement				

civic engagement index	0.012	(0.013)	0.017	(0.011)
Social integration				
social isolation	0.009	(0.033)	0.005	(0.021)
social trust	0.050	(0.035)	0.013	(0.021)
married	-0.047	(0.165)	-0.001	(0.104)
renting	-0.137	(0.163)	-0.066	(0.106)
Busyness				
satisfaction w/ leisure time	0.028	(0.037)	0.019	(0.026)
number of children in household	-0.164**	(0.059)	-0.053	(0.041)
employed	-0.460*	(0.221)	-0.141	(0.138)
Personality				
need for cognition	-0.009	(0.006)	0.000	(0.004)
openness to experience	0.019	(0.018)	0.005	(0.011)
extraversion	-0.026*	(0.011)	-0.018*	(0.007)
agreeableness	0.032*	(0.015)	0.012	(0.010)
conscientiousness	0.014	(0.013)	-0.004	(0.011)
emotional stability	0.005	(0.011)	0.008	(0.007)
Attitudes about surveys				
survey enjoyment	0.025	(0.024)	0.054	(0.030)
survey value	-0.069	(0.042)	-0.014	(0.026)
survey burden	-0.005	(0.023)	-0.009	(0.015)
Internet use				
smartphone use	0.359*	(0.160)	0.375*	(0.168)
tablet use	-0.202	(0.146)	-0.045	(0.089)
social media use	-0.181	(0.178)	0.166	(0.143)
computer Internet use (weekly hours)	0.005	(0.005)	0.004	(0.004)
e-mail use (weekly hours)	-0.010	(0.008)	-0.009	(0.006)
Experimental variable				
wave (=1)	0.329*	(0.138)		
Instrumental variable				
willingness propensity			**2.455***	**(1.024)**

*p<0.05, **p < 0.01, ***p < 0.001
Model 2: R^2 = .16, Max-rescaled R^2 = .21
Model 2: Weighted n = 5147

References

Abraham, K. G., Maitland, A., & Bianchi, S. M. (2006). Nonresponse in the American Time Use Survey: Who is missing from the data and how much does it matter?. *Public Opinion Quarterly*, *70*, 676-703.

The American Association for Public Opinion Research (2011) Standard definitions: final dispositions of case codes and outcome rates for surveys. 7th edition. Available: http://www.aapor.org/AM/Template.cfm?Section=Standard_Definitions2&Template=/CM/ContentDisplay.cfm&ContentID=3156.

The American Association for Public Opinion Research (2014). *Mobile Technologies for Conducting, Augmenting and Potentially Replacing Surveys: Report of the AAPOR Task Force on Emerging Technologies in Public Opinion Research*. Deerfield, IL: American Association for Public Opinion Research.

Aquilino, W. S. (1993). Effects of spouse presence during the interview on survey responses concerning marriage. *Public Opinion Quarterly*, *57*, 358-376.

Aquilino, W. S., Wright, D. L., & Supple, A. J. (2000). Response effects due to bystander presence in CASI and paper-and-pencil surveys of drug use and alcohol use. *Substance Use & Misuse*, *35*, 845-867.

Baker-Prewitt, J. (2014), Do smartphones really produce lower scores? Understanding device effects on survey ratings. Paper presented at CASRO Digital Research Conference, March 11-12, San Antonio, Texas, USA.

Baker-Prewitt, J. (2013), Mobile research risk: What happens to data quality when respondents use a mobile device for a survey designed for a PC. Paper presented at the CASRO Online Research Conference, March 6-8, San Francisco, USA.

Bi, X., Li, Y., & Zhai, S. (2013, April). Fitts law: modeling finger touch with fitts' law. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1363-1372). ACM.

Biemer, P. P., & Lyberg, L. E. (2003). *Introduction to survey quality*. John Wiley & Sons.

Bosnjak, M., Poggio, T., Becker, K. R., Funke, F., Wachenfeld, A., & Fischer, B. (2013a). Online survey participation via mobile devices. Paper presented at AAPOR, 2013, Boston, MA. Retrieved Jan 8, 2015, from http://www.psyconsult.de/bosnjak/pages/publications/conference-contributions.php

Bosnjak, M., Haas, I., Galesic, M., Kaczmirek, L., Bandilla, W., & Couper, M. P. (2013b). Sample composition discrepancies in different stages of a probability-based online panel. *Field Methods*, 1525822X12472951.

Bosnjak, M., Metzger, G., & Gräf, L. (2010). Understanding the willingness to participate in mobile surveys: exploring the role of utilitarian, affective, hedonic, social, self-expressive, and trust-related factors. *Social Science Computer Review*, *28*, 350-370.

Brüggen, E., Wetzels, M., de Ruyter, K., & Schillewaert, N. (2011). Individual differences in motivation to participate in online panels. The effect on response rate and response quality perceptions. *International Journal of Market Research*, *53*, 369-398.

Buskirk, T.D., & Andrus, C. (2012). Smart surveys for smart phones: Exploring various approaches for conducting online mobile surveys via smartphones. *Survey Practice*, 5.

Buskirk, T. D., & Andrus, C. (2014). Making mobile browser surveys smarter: results from a randomized experiment comparing online surveys completed via computer or smartphone. *Field Methods*, 1525822X14526146.

Buskirk, T.D. (2013). Smarter smartphone surveys 201: Data collection methods and survey design considerations. AAPOR Webinar. September 25, 2013.

Butt, S., & Phillips, J. G. (2008). Personality and self reported mobile phone use. *Computers in Human Behavior*, *24*, 346-360.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*, 116.

Callegaro, M. (2010). Do you know which device your respondent has used to take your online survey? Using paradata to collect information on device type. *Survey Practice*, 3.

Card, S. K., English, W. K., & Burr, B. J. (1978). Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a CRT. *Ergonomics, 21*, 601-61

Chang, L., & Krosnick, J. A. (2009). National surveys via RDD telephone interviewing versus the internet comparing sample representativeness and response quality. *Public Opinion Quarterly, 73*, 641-678.

Cialdini, R. B. (1988). *Influence: Science and practice*. Glenview, IL: Scott, Foresman.

Conrad, F. G., Brown, N.R. & Cashman, E. R. (1998). Strategies for estimating behavioural frequency in survey interviews. *Memory, 6*, 339–366.

Conrad, F., M. Schober, C. Zhang, H. Yan, L. Vickers, M. Johnston, A. Hupp, L. Hemingway, S. Fail (2013). Mode choice on an iPhone increases survey data quality. Paper presented at the Annual Conference of the American Association for Public Opinion Research, Boston, MA.

Cook, C., Heath, F., Thompson, R. L., & Thompson, B. (2001). Score reliability in Web or internet-based surveys: unnumbered graphic rating scales versus Likert-type scales. *Educational and Psychological Measurement, 61*, 697-706.

Costa, P. T., & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological assessment*, *4,* 5.

Couper, M. P. (2000). Review: Web surveys: A review of issues and approaches. *Public opinion quarterly*, 464-494.

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

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

Couper, M. Antoun, C. & Mavletova, A. (2015). Mobile Web surveys: A total survey error perspective. In forthcoming Total Survey Error edited volume.

Couper M. P., & Peterson, G. (2015). Exploring why mobile Web surveys take longer. Paper presented at General Online Research Conference. March 18-20, 2015, Germany.

Couper, M. P., Kapteyn, A., Schonlau, M., & Winter, J. (2007). Noncoverage and nonresponse in an Internet survey. *Social Science Research*, *36*, 131-148.

Couper, M. P., Singer, E., & Kulka, R. A. (1998). Participation in the 1990 decennial census: Politics, privacy, pressures. *American Politics Research*, *26,* 59-80.

Couper, M. P., Tourangeau, R., Conrad, F. G., & Singer, E. (2006). Evaluating the effectiveness of visual analog scales a web experiment. *Social Science Computer Review, 24,* 227-245.

Couper, M. P., Traugott, M. W., & Lamias, M. J. (2001). Web survey design and administration. *Public Opinion Quarterly, 65*, 230-253.

Cui, Y., & Roto, V. (2008). How people use the Web on mobile devices. *Proceedings of the 17th International Conference on World Wide Web*, New York, ACM: pp. 905-914.

Davis Jr, F. D. (1986). *A technology acceptance model for empirically testing new end-user information systems: Theory and results* (Doctoral dissertation, Massachusetts Institute of Technology).

De Bruijne, M., & Wijnant, A. (2013a). Comparing survey results obtained via mobile devices and computers: An experiment with a mobile Web survey on a heterogeneous group of mobile devices versus a computer-assisted Web survey. *Social Science Computer Review*, *31*, 482-504.

De Bruijne, M., & Wijnant, A. (2013b), Can mobile Web surveys be taken on computers? A Discussion on a Multi-Device Survey Design. *Survey Practice, 6* (4), http://surveypractice.org/index.php/SurveyPractice/article/view/238/

De Bruijne, M., & Wijnant, A. (2014).Mobile response in Web panels. *Social Science Computer Review*, DOI: 0894439314525918.

De Leeuw, E. D. (2005). To mix or not to mix data collection modes in surveys. *Journal of Official Statistics, 21*, 233-255.

De Nazell, A., Seto, E., Donaire-Gonzalez, D. et al. (2013). Improving estimates of air pollution exposure through ubiquitous sensing technologies. *Environmental Pollution, 176*, 92-99.

Dillman, D.A. (2000). *Mail and Internet surveys: The Tailored Design Method*. New York: John Wiley.

Duggan, M., & Smith, A. 2013. Cell Internet Use 2013. *Pew Internet and American Life Project.*

Fan, W., & Yan, Z. (2010). Factors affecting response rates of the web survey: A systematic review. *Computers in human behavior*, *26*, 132-139.

Fecso, R., & Pafford, B. (1988). Response errors in establishment surveys with an example from an agribusiness survey. In *Proceedings of the Section on Survey Research Methods, American Statistical Association* (pp. 315-320).

Fernee, H., & Sonck, N. (2013). Is everyone able to use a smartphone in survey research?. *Survey Practice,* 6.

Forlines, C., Wigdor, D., Shen, C., & Balakrishnan, R. (2007, April). Direct-touch vs. mouse input for tabletop displays. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 647-656). ACM.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic perspectives*, 25-42.

Fuchs, M., & Busse, B. (2009). The coverage bias of mobile web surveys across European countries. *International Journal of Internet Science*, 4, 21-33.

Gibbs, N. (2012). Your life is fully mobile. Time Magazine. Retrieved from http://techland.time.com/2012/08/16/your-life-is-fully-mobile/

Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, *70*, 646-675.

Groves, R. M. (2004). *Survey errors and survey costs*. John Wiley & Sons.

Groves, R. M., Cialdini, R. B., & Couper, M. P. (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly*, *56*, 475-495.

Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household interview surveys*. John Wiley & Sons.

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

Groves, R. M., Presser, S., & Dipko, S. (2004). The role of topic interest in survey participation decisions. *Public Opinion Quarterly*, *68*, 2-31.

Groves, R. M., Singer, E., & Corning, A. (2000). Leverage-saliency theory of survey participation: Description and an illustration. *Public Opinion Quarterly*, *64*, 299-308.

Guidry, K. R. (2012). Respondents using a mobile device on a Web-based survey. Retrieved from http://cpr.indiana.edu/uploads/Response%20Quality%20and%20Demographic%20Characteristics%20of%20Respondents.pdf

Gupta, A., & Lee J. (2013). How does online survey mode affect answers to customer feedback loyalty surveys? Presented at Joint Statistical Meetings, August 3-8, Quebec, Canada.

Hainmueller, J., & Hiscox, M. J. (2007). Educated preferences: Explaining attitudes toward immigration in Europe. *International Organization*, *61*, 399-442.

Hansen, M. H., Hurwitz, W. N., & Bershad, M. A. (1961). Measurement errors in censuses and surveys. *Bulletin of the International Statistical Institute*, *38*, 359-374.

Heerwegh, D., & Loosveldt, G. (2008). Face-to-face versus web surveying in a high-internet-coverage population differences in response quality. *Public Opinion Quarterly*, *72*, 836-846.

Holbrook, A. L., Anand, S., Johnson, T. P., Cho, Y. I., Shavitt, S., Chávez, N., & Weiner, S. (2014). Response heaping in interviewer-administered surveys is it really a form of satisficing?. *Public Opinion Quarterly, 78*, 591-633.

Hosmer, D. W., & Lemeshow, S. (2013). *Applied Logistic Regression*. New York: Wiley

Hu, S., & Dayton, J. (2014) Using Smartphones to collect information about health behaviors: A formative and process evaluation. Presented at 18th Annual Federal Computer Assisted Survey Information Collection, March 17, 2014, Washington, D.C.

Hyman, I. E., Boss, S. M., Wise, B. M., McKenzie, K. E., & Caggiano, J. M. (2010). Did you see the unicycling clown? Inattentional blindness while walking and talking on a cell phone. *Applied Cognitive Psychology*, *24*, 597-607.

Johnson, D. E. (2010). Crossover experiments. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*, 620-625.

Jones, P., Drury, R., & McBeath, J. (2011). Using GPS-enabled mobile computing to augment qualitative interviewing: Two case studies. *Field Methods*, *23*, 173-187.

Kelly, F., & Stevens, S. (2013). 'Bite-sized chunks' – mobile and CAWI parallel proposal. Paper presented at Online Research Conference, March 7-8, San-Francisco, USA.

Kennedy, Courtney K. (2010). *Nonresponse and measurement error in mobile phone surveys.* University of Michigan. http://pocarisweat.umdl.umich.edu/handle/2027.42/75977.

Klausch, T., Schouten, B., & Hox, J. (2014). The use of within-subject experiments for estimating measurement effects in mixed-mode surveys. Statistics Netherlands Discussion Paper.

Keusch, F., Batinic, B., & Mayerhofer, W. (2014). Motives for joining nonprobability online panels and their association with survey participation behavior. In Callegaro, M., Baker, R. P., Bethlehem, J., Göritz, A. S., Krosnick, J. A., & Lavrakas, P. J. (Eds.). (2014). *Online panel research: a data quality perspective (pp 171-191)*. John Wiley & Sons.

Keusch, F. (2015). Why do people participate in Web surveys? *Management Review Quarterly*, *65*, 183-216.

Kinesis (2013). Online survey statistics from the mobile future. Kinesis Whitepaper, http://www.kinesissurvey.com/wp-content/uploads/2013/10/UPDATED-with-Q3-2013-Data-Mobile-whitepaper.pdf.

Klausch, T., Schouten, B., & Hox, J. (2014). The use of within-subject experiments for estimating measurement effects in mixed-mode surveys. Statistics Netherlands Discussion Paper.

Kreuter, F., Presser, S., & Tourangeau. R. (2008). Social desirability bias in CATI, IVR, and Web surveys the effects of mode and question sensitivity. *Public Opinion Quarterly*, *72*, 847–865.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*, 213–236.

Krosnick, J. A., & Alwin, D. F. (1987). An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly, 51*, 201–219.

Lattery, K., Park Bartolone, G., & Saunders, T. (2013). Optimizing surveys for smartphones: Maximizing response rates while minimizing bias. Paper presented at the CASRO Online Research Conference, March 6-8, San Francisco, USA.

Lavrakas, P. J., Tompson, T. N., Benford, R., & Fleury, C. (2010). Investigating Data Quality in Cell Phone Surveying. In proceedings of American Association for Public Opinion Research conference, Chicago, Illinois (pp. 13-16).

Lee, S. Y. (2014). Examining the factors that influence early adopters' smartphone adoption: The case of college students. *Telematics and Informatics, 31*, 308-318.

Lee, S. (2006). An evaluation of nonresponse and coverage errors in a prerecruited probability web panel survey. *Social Science Computer Review*, *24*, 460-475.

Lee, I., Kim, J., & Kim, J. (2005). Use contexts for the mobile internet: a longitudinal study monitoring actual use of mobile internet services. *International Journal of Human-Computer Interaction, 18*, 269-292.

Link, M. (2013). Measuring compliance in mobile longitudinal repeated-measures design study. *Survey Practice,* 6.

Link, M. (2014). Emerging technologies in public opinion research - the current landscape. AAPOR Webinar. October 22, 2014.

Lipsman, A., & Aquino, C. (2013). Mobile future in focus. *comScore Report.* http://www.comscore.com/Insights/Presentations-and-Whitepapers/2013/2013-Mobile-Future-in-Focus

Lugtig, P., & Toepoel, V. (2015). The use of PCs, smartphones, and tablets in a probability-based panel survey effects on survey measurement error. *Social Science Computer Review*, 0894439315574248.

Lynn, P., & Kaminska, O. (2012). The impact of mobile phones on survey measurement error. *Public Opinion Quarterly*. doi:10.1093/poq/nfs046

Macer, T. (2012). Developments and the impact of smart technology. *International Journal of Market Research*, 54, 567-570.

Manski, C. F., & Molinari, F. (2010). Rounding probabilistic expectations in surveys. *Journal of Business & Economic Statistics, 28*, 219-231.

Mavletova, A. (2013). Data quality in PC and mobile web surveys. *Social Science Computer Review, 31*, 725-743.

Mavletova, A., & Couper, M. P. (2013). Sensitive topics in PC Web and mobile Web surveys: Is there a difference?. *Survey Research Methods, 7*, 191-205.

Mavletova, A., & Couper, M.P. (2015). Device use in Web surveys: The effect of differential incentives. Paper presented at the General Online Research Conference, March 18-20, Cologne, Germany.

McClain, C. A., Crawford, S.D. & Dugan, J.P. (2012). Use of mobile devices to access computer-optimized Web instruments: Implications for respondent behavior and data quality. Presented at the annual conference of the American Association for Public Opinion Research, Orlando, FL.

McGeeney, K. (2015). Tips for creating Web surveys for completion on a mobile device. Pew Research Center. Retrieved from http://www.pewresearch.org/2015/06/11/tips-for-creating-web-surveys-for-completion-on-a-mobile-device/

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika, 12*, 153-157.

Medallia Webinar (2012). Mobile feedback best practices. Retrieved from http://www.medallia.com/resource/mobile-feedback-best-practices/

Metzler, A., & Fuchs, M. (2014), Coverage error in mobile Web surveys across European countries. Paper presented at the Internet Survey Methodology Workshop, December 1-3, Bozen-Bolzano, Italy.

Nagelkerke, N. J. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, *78*, 691-692.

Naimi, T. S., Brewer, R. D., Mokdad, A., Denny, C., Serdula, M. K., & Marks, J. S. (2003). Binge drinking among US adults. *Jama*, *289*, 70-75.

Neter, J., Kutner, M.H., Nachtsheim, C.J., & Wasserman, W. (1996). *Applied linear statistical models*. WCB McGraw-Hill, MA.

Norman, D. A. (1981). Categorization of action slips. *Psychological review, 88,* 1.

Olson, K., & Wagner, J. (2013). A field experiment using GPS devices to measure interviewer travel behavior. Paper presented at the AAPOR Annual Conference, May 16-19, Boston, USA.

Pan, W. (2001). Akaike's information criterion in generalized estimating equations. *Biometrics, 57*, 120-125.

Park, Y., & Chen, J. V. (2007). Acceptance and adoption of the innovative use of smartphone. *Industrial Management & Data Systems*, *107*, 1349-1365.

Peterson, G. (2012). Unintended mobile respondents. Paper presented at the CASRO Technology Conference, May 31, New York, USA.

Peterson, G., J. Mechling, J. LaFrance, J. Swinehart, & G. Ham (2013). Solving the unintentional mobile challenge. Paper presented at the CASRO Online Research Conference, March 6–8, San Francisco, USA, http://c.ymcdn.com/sites/www.casro.org/resource/collection/0A81BA94-3332-4135-97F6-6BE6F6CEF475/Paper_-_Gregg_Peterson_-_Market_Strategies_International.pdf.

Peytchev, A., & Hill, C. A. (2010). Experiments in mobile web survey design similarities to other modes and unique considerations. *Social Science Computer Review, 28*, 319-335.

Peytchev, A., Baxter, R. K., & Carley-Baxter, L. R. (2009). Not all survey effort is equal reduction of nonresponse bias and nonresponse error. *Public Opinion Quarterly*, nfp037.

Phonescoop. (2012). Definition of a Smartphone. Retrieved from http://www.phonescoop.com/glossary/term.php?gid=131.

Reason, J. (1990). *Human error.* Cambridge university press.

Reips, U. D., & Funke, F. (2008). Interval-level measurement with visual analogue scales in Internet-based research: VAS Generator. *Behavior Research Methods 40,* 699-704.

Revilla, M., Toninelli, D., Ochoa, C., & Loewe, G. (2014). Do online access panels really need to allow and adapt surveys to mobile devices?. RECSM-Research and Expertise Centre for Survey Methodology, 41.

Roe, D., Zhang, Y., & Keating, M. (2014). Decisions, observations, and considerations for developing a mobile survey app and panel. In Murphy, J., Hill, C. A., & Dean, E., (Eds.) (2014). *Social Media, Sociality, and Survey Research (pp 149-178).* John Wiley & Sons, Inc.

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

Sakshaug, J. W., Yan, T., & Tourangeau, R. (2010). Nonresponse error, measurement error, and mode of data collection: tradeoffs in a multi-mode survey of sensitive and non-sensitive items. *Public Opinion Quarterly*, *74*, 907-933.

Saucier, G., & Goldberg, L. R. (2002). Assessing the Big Five: Applications of 10 psychometric criteria to the development of marker scales. *Big Five Assessment*, 29-58.

Scherpenzeel, A. (2011). Data collection in a probability-based internet panel: how the LISS panel was built and how it can be used. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique, 109*, 56-61.

Schober, M.F., Conrad, F.G., Antoun, C., Carroll, D., Ehlen, P., Fail, S., Hupp, A.L., Johnston, M., Kellner, C., Nichols, K., Percifield, L., Vickers, L., Yan, H., & Zhang, C. (2012). Disclosure and quality of answers in text and voice interviews on iPhones. Presented at the annual conference of the American Association for Public Opinion Research, Orlando, FL.

Schouten, B., van den Brakel, J., Buelens, B., van der Laan, J., & Klausch, T. (2013). Disentangling mode-specific selection and measurement bias in social surveys. *Social Science Research*, *42*, 1555-1570.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63,* 129.

Simon, H. A. (1957). *Models of man; social and rational*. New York: Wiley.

Smith, A. (2012). Nearly half of American adults are smartphone owners. *Pew Internet & American Life Project*.

Smith, A. (2015). US Smartphone Use in 2015. *Pew Internet & American Life Project*.

Smyth, J. D., Dillman, D. A., Christian, L. M., & Mcbride, M. (2009). Open-ended questions in Web surveys. Can increasing the size of answer boxes and providing extra verbal instructions improve response quality?. *Public Opinion Quarterly*, nfp029.

Stapleton, C. (2013). The smart (phone) way to collect survey data. *Survey Practice, 6*.

Statistics Netherlands. (2013). Smartphones used most for mobile internet access. *Web magazine,* 09 July 2013. Retrieved January 8, 2015, from http://www.cbs.nl/

Toepoel, V., & Couper, M. P. (2011). Can verbal instructions counteract visual context effects in web surveys?. *Public Opinion Quarterly*, *75*, 1-18.

Toepoel, V. & Lugtig, P. (2014). What happens if you offer a mobile option to your Web panel? Evidence from a probability-based panel of Internet users. *Social Science Computer Review*, DOI: 0894439313510482.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, *39*, 1275-1289.

Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions the impact of data collection mode, question format, and question context. *Public Opinion Quarterly*, *60*, 275-304.

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

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. New York: Cambridge University Press.

Bosnjak, M., & Tuten, T. L. (2001). Classifying response behaviors in web-based surveys. *Journal of Computer-Mediated Communication*, 6.

Walton, L., Buskirk, T., & Wells, T. (2013). Smarter online panels for smartphone users: exploring factors associated with mobile panel participation. Paper presented at the AAPOR Annual Conference, May 16-19, Boston, MA.

Wells, T., Bailey, J. T., & Link, M. W. 2014. Comparison of smartphone and online computer survey administration. *Social Science Computer Review, 32*, 238-255.

West, B.T., Welch, K.B., & Gałecki, A.T. (2007). *Linear mixed models: A practical guide using statistical software*. Boca Raton: Chapman & Hall/CRC

Wikipedia. (2013). Smartphone. Retrieved from http://en.wikipedia.org/wiki/Smartphone.

Wood, E., Willoughby, T., Rushing, A., Bechtel, L., & Gilbert, J. (2005). Use of computer input devices by older adults. *Journal of Applied Gerontology, 24*, 419-438.

Zahariev, M., Ferneyhough, C., & Ryan, C. (2009). Best practices in mobile research. Paper presented at ESOMAR Online Research. October 26-28, 2009. Chicago, USA.

Zwarun, L., & Hall, A. (2014). What's going on? Age, distraction, and multitasking during online survey taking. *Computers in Human Behavior, 41*, 236-244.