Sources of error in mobile survey data collection

Alexander Wenz

A thesis submitted for the degree of

Doctor of Philosophy in Survey Methodology

Institute for Social and Economic Research

University of Essex

December 2017

Declarations

No part of this thesis has been submitted for another degree.

I am the sole author of Chapter 1 and Chapter 2. Chapter 3 is co-authored with Annette Jäckle, University of Essex, and Mick Couper, University of Michigan. I did the data management and analysis and wrote a first draft of the paper. Annette and Mick gave feedback on the analysis and edited the draft. We worked together in revising the paper.

Two of the chapters have been published as working papers.

An early version of Chapter 1 has been published as

> Wenz, A. (2017). Completing web surveys on mobile devices: does screen size affect data quality? *ISER Working Paper*, 2017-05.

An early version of Chapter 3 has been published as

> Wenz, A., Jäckle, A., & Couper, M. P. (2017). Willingness to use mobile technologies for data collection in a probability household panel. *Understanding Society Working Paper*, 2017-10.

Chapter 3 is currently under review at *Survey Research Methods*.

For Mama, Nana, and Papa

Acknowledgements

I am profoundly grateful to all people who have supported and encouraged me throughout all these years. I would like to thank my supervisors Peter Lynn and Tarek Al Baghal for their outstanding support, invaluable feedback and guidance. I am also grateful to my co-author Annette Jäckle. I learned a lot from her while collaborating on my third chapter and she became a very supportive mentor. I am thankful that I had the chance to work with colleagues from Kantar Public on various projects during my PhD. I especially would like to thank Joel Williams for very interesting discussions and for his support when working on the data for my first chapter.

The Institute for Social and Economic Research provided me with a vibrant, collaborative and inspiring research environment where I very much enjoyed working. I have greatly benefited from discussions that I had with Jon Burton and Olena Kaminska and with my fellow students in the survey methodology programme, Valerija Kolbas, Vilma Agalioti-Sgompou, Carlos Lagorio, Alexandru Cernat, Yamil Nares, Brendan Read, Jennifer Kelley, and Linh Nguyen.

I am extremely grateful that I had the opportunity to complete research visits at the University of Michigan and at the University of Mannheim. I would like to express my sincere thanks to Mick Couper for his time and support when he so generously hosted me in Ann Arbor and for all the inspiring discussions we had. I enjoyed working in the intellectually stimulating research environment at the Institute for Social Research and I am grateful that I could discuss aspects of my work with Fred Conrad, Andy Peytchev, Brady West, Felicitas Mittereder, Micha Fischer, Colleen McClain, and Yanna Yan. I would also like to thank Annelies Blom for generously hosting me in Mannheim and for many interesting discussions. My thanks also go to the other members of the German Internet Panel, Franziska Gebhard, Jessica Herzing, Carina Cornesse, Barbara Felderer, Ulrich Krieger, and Christian Bruch.

I have experienced survey methodology as a vibrant and very inspiring field to work in. My sincere thanks go to Peter Selb, Lars Kaczmirek and Nick Allum for introducing me to this field and for supporting my decision to apply for the PhD programme at the University of Essex.

I gratefully acknowledge the financial support from the Economic and Social Research Council and from Kantar Public that allowed me to focus on my research during these years. Thanks to an overseas institutional visit grant from the Economic and Social Research Council, I was able to complete the research visit at the University of Michigan. I also appreciate the funding that I received from ESSEXLab at the University of Essex to conduct the laboratory experiment for my second chapter as well as the travel funding from the American Association for Public Opinion Research that allowed me to present my third chapter at the AAPOR Conference 2017 in New Orleans.

I am very grateful for all the great opportunities that I was given during my PhD to grow both as a researcher and as a person. I would like to thank my friends in the UK, in Germany and all around the world who have accompanied and supported me along the way. Finally, I want to thank my family, Mama, Nana, and Papa. I cannot express with words how grateful I am for your emotional support and love throughout all these years. I dedicate this thesis to you.

Summary

The proliferation of mobile technologies in the general population offers new opportunities for survey research, but also introduces new sources of error to the data collection process. This thesis studies two potential sources of error in mobile survey data collection: measurement error and nonresponse.

Chapter 1 examines how the diagonal screen size of a mobile device affects measurement error. Using data from a non-mobile-optimised web survey, I compare data quality between screen size groups. Results suggest that data quality mainly differs between small smartphones with a screen size of below 4.0 inches and larger mobile devices. Respondents using small smartphones are more likely to break off during the survey, to provide shorter answers to open-ended questions, and to select fewer items in check-all-that-apply questions than respondents using devices with larger screens.

Due to the portability of mobile devices, mobile web respondents are more likely to be in distracting environments where other people are present. Chapter 2 explores how distractions during web survey completion influence measurement error. I conducted a laboratory experiment where participants were randomly assigned to devices (PC or tablet) and to one of three distraction conditions (presence of other people who have a loud conversation, presence of music, or no distraction). Although respondents felt more distracted in the two distraction conditions, I did not find significant effects of distraction on data quality.

Chapter 3 investigates correlates of nonresponse to data collection using mobile technologies. We asked members of a probability household panel about their willingness

to participate in various data collection tasks on their mobile device. We find that willingness varies considerably by the type of activity involved, to some extent by device, and by respondent: those who report higher security concerns and who use their device less intensively are less willing to participate in mobile data collection.

Contents

List of Figures

List of Tables

Introduction

We live in an exciting time for survey methodology. The field is confronted with three technology-related developments that radically change the way we do research and that "expand the range of tools available to us to understand society", as Mick Couper discussed in his keynote speech at the ESRA Conference 2013 in Ljubljana (Couper, 2013). First, in addition to traditional surveys, social scientists increasingly rely on new sources of data that are usually subsumed under the terms "Big Data" or "organic data" (Groves, 2011). These include social media data, sensor data, transactional data, and administrative data (Japec et al., 2015). Second, online opt-in panels and other surveys based on non-probability sampling methods increasingly gain acceptance in the scientific community as an alternative to probability sample surveys (Baker et al., 2013). Third, the emergence of mobile technologies including smartphones, tablets and other devices, such as smartwatches and activity trackers, provides new opportunities to survey researchers but also creates new methodological challenges (Link et al., 2014). This thesis focuses on the third of these three developments.

Over the last years, the availability of mobile devices has increased tremendously in the general population. While in 2011, 27 percent of households in the United Kingdom reported owning a smartphone and only 2 percent owning a tablet, this number has increased to 76 percent of households owning a smartphone in 2017 and 58 percent owning a tablet (Ofcom, 2017). Mobile devices not only allow administering web surveys in innovative ways but also capturing new forms of data that have distinct advantages over questionnaire-based methods of data collection. Mobile technologies can be used, for example, to conduct 'in-the-moment' surveys that are triggered at regular intervals and might reduce the respondent's need to recall information, or to collect objective measures by relying on integrated mobile device sensors, such as the in-built accelerometer (Couper, Antoun, & Mavletova, 2017; Link et al., 2014).

The widespread use of mobile technologies in the general population also creates new challenges for survey researchers. First, the large variety of mobile devices – which may vary in screen size, operating system, browser type and other technical features – make the design and testing of web surveys more challenging. As Vera Toepoel and Peter Lugtig (2015) pointed out, we should be aware that all of our web surveys are now "mixed-device surveys". Second, technology develops quickly and the survey profession needs to keep adapting to the technological advancements. When I started to write the first chapter of this thesis in 2014, research on mobile web surveys was at a relatively early stage and many large-scale social surveys were still hesitant towards the development of mobile-optimised questionnaires. Only three years later, large-scale social surveys, such as *Understanding Society* – The UK Household Longitudinal Study based at the University of Essex, not only adapt their web questionnaire for mobile web respondents but also experiment with new ways of data collection using mobile devices, as more research and best practices emerged on how to optimise survey instruments for mobile device users. Third, and this is presumably one of the main challenges in this area, mobile data collection introduces new sources of potential error to the data collection process that survey methodologists need to understand better.

The aim of this thesis is to evaluate two potential sources of error in mobile survey data collection: the first two chapters study two features of mobile data collection, the small screen size of mobile devices and the potentially more distracting environment of mobile device users, and the impact on measurement error, whereas the third chapter investigates nonresponse error by looking at correlates of nonresponse in mobile data collection. The next section outlines each of the three chapters.

Chapter 1 examines how the diagonal screen size of a mobile device affects measurement error. Using data from the web component of the Community Life Survey, a general population survey in England, I compare data quality between five different screen size groups: small smartphones (< 4.0 inches), large smartphones (≥ 4.0 inches), small tablets (< 8.0 inches), large tablets (≥ 8.0 inches), and PCs/laptops. Results suggest that data quality mainly differs between small smartphones and larger mobile devices. Respondents using small smartphones are more likely to break off during the survey, to provide shorter answers to open-ended questions, and to select fewer items in check-all-that-apply questions than respondents using larger devices. I do not find significant differences between screen size groups in completion times, response distributions, and straight-lining.

Due to the portability of mobile devices, mobile web respondents are more likely to be in distracting environments where other people are present compared to respondents with desktop PCs or laptops. In Chapter 2, I examine how distractions during web survey completion influence measurement error. I conducted a laboratory experiment where I randomly assigned participants to devices (PC or tablet) and to distraction conditions that are likely to occur in web survey settings (presence of other people who have a loud conversation, presence of music, or no distraction). I use eleven indicators to compare data quality between the experimental conditions, including non-differentiation in grid questions, length of responses to open-ended questions, and responses to an Instructional Manipulation Check. Although respondents feel more distracted if other people are present or if they listen to music, I do not find significant effects of distraction on data quality.

Chapter 3 investigates the stated willingness of the general population to participate in studies that involve mobile data collection, and factors that affect willingness. We asked members of the *Understanding Society* Innovation Panel, a probability household panel in Great Britain, about their willingness to participate in various data collection tasks on their mobile device. We find that stated willingness varies considerably depending on the type of activity involved: respondents are less willing to participate in tasks that require downloading and installing an app, or where data are collected passively. Stated willingness also varies between smartphones and tablets, and between types of respondents: those who report higher concerns about the security of data collected with mobile technologies and those who use their devices less intensively are less willing to participate in mobile data collection.

1. Completing web surveys on mobile devices: does screen size affect data quality?

Abstract

Using data from a non-mobile-optimised web survey in England, this paper compares the quality of survey data from mobile devices with different screen size. The findings suggest that data quality mainly differs between small smartphones with a diagonal screen size of below four inches and larger mobile devices. Users of small smartphones are significantly more likely to drop out of the survey, to provide shorter responses to open-ended questions, and to select fewer items in check-all-that-apply questions. There are no significant differences between screen size groups in completion times, response distributions, and straight-lining.

1.1. Introduction

Mobile technology has become an integral part of people's daily life. In 2017, 76 percent of households in the United Kingdom owned a smartphone and 58 percent owned a tablet (Ofcom, 2017). On average, British people spend 65 hours on their smartphone per month (around two hours per day), and women aged 16-24 even 89 hours per month (around three hours per day). Similar trends can be observed in the United States and in other Western countries (Anderson, 2015; Poushter, 2016).

This development of mobile technology also affects survey research. An increasing number of survey participants access web surveys on their mobile device, regardless of whether the survey designer intended mobile completion and optimised the questionnaire for mobile devices (de Bruijne & Wijnant, 2014; Lugtig & Toepoel, 2015; Peterson, 2012; Poggio, Bosnjak, & Weyandt, 2015; Revilla, Toninelli, Ochoa, & Loewe, 2016; Struminskaya, Weyandt, & Bosnjak, 2015). For example, in the LISS panel, a probability-based online panel in the Netherlands, the proportion of respondents who used a smartphone or a tablet for survey completion grew from 3 percent in March 2012 to 11 percent in September 2013 (de Bruijne & Wijnant, 2014). More recently, Struminskaya, Weyandt, and Bosnjak (2015) reported that in 2014, 18 percent of web respondents of the GESIS panel, a probability-based mixed-mode panel in Germany, completed the survey on a mobile device. When asked about their preferred device to participate in the survey, around 24 percent of panel members indicated either a tablet or a smartphone as their preferred device in 2015. Given that mobile device ownership continues to grow among the general population (e.g., Anderson, 2015), it can be expected that the proportion of mobile respondents in web surveys of the general population will further increase in the future.

The increasing use of mobile devices by survey respondents creates new challenges for survey researchers. One of the primary concerns is that certain characteristics of mobile devices, such as the smaller screen size or the touchscreen interface, make survey completion more burdensome, and that mobile respondents may hence provide survey data of lower quality compared to respondents who use desktop computers or laptops (Couper et al., 2017; Lugtig & Toepoel, 2015; Peytchev & Hill, 2010). Survey managers have various options to handle mobile devices: they can offer an optimised web questionnaire for mobile browsers, for example using a responsive or adaptive web design, implement the questionnaire within a mobile survey app, or administer the standard web questionnaire and discourage the use of mobile devices (Buskirk & Andrus, 2012; Callegaro, 2010; Callegaro, Lozar Manfreda, & Vehovar, 2015). While a mobile-optimised questionnaire or a mobile survey app may improve survey experience for mobile device users, the development and maintenance of such survey versions is costly and involves methodological challenges, for example how to best present grid questions or questions with horizontal rating scales on mobile devices (Callegaro et al., 2015). Therefore, until very recently, large-scale social surveys still adopted the more conservative approach of discouraging mobile survey completion. There was, however, little agreement as to which types of mobile devices should be discouraged from being used to maintain high data quality. Different thresholds were applied that were either based on diagonal screen size, for example in the *Understanding Society* Innovation Panel (Hanson, Matthews, & McGee, 2015), or based on screen resolution, such as in the 1958 National Child Development Study (TNS BMRB, 2014).

Previous research on data quality in non-mobile-optimised web surveys found larger quality differences between smartphones and PCs than between tablets and PCs (e.g., de Bruijne & Wijnant, 2013; Guidry, 2012; Lugtig & Toepoel, 2015), and it has been speculated that the small screen size of smartphones may be a major factor why mobile respondents provide data of lower quality compared to PC respondents. This assumption, however, has not been tested as existing studies have mainly compared data quality between smartphones, tablets and PCs but have not considered screen size differences within device classes.

Using data from a web survey which administers a non-optimised web questionnaire to mobile respondents, this paper provides novel evidence on how the diagonal screen size of a mobile device affects survey data quality. This study extends earlier research by focusing on screen size rather than categories of devices, which allows distinguishing between small and large smartphone as well as between small and large tablets. The results are intended to inform decisions that survey managers have to make when dealing with mobile respondents: which screen size appears to be problematic in terms of data quality? If a mobile-optimised version of the questionnaire has been developed, below which threshold should optimisation be triggered? Conversely, if a mobile-optimised questionnaire has not been developed yet, which types of devices should be discouraged from being used?

1.2. Background and Hypotheses

Previous research has shown that mobile device use negatively affects survey data quality. Mobile respondents, particularly smartphone users, have higher breakoff rates and longer completion times compared to PC respondents in surveys which have not been optimised for mobile devices (Bosnjak et al., 2013; Callegaro, 2010; Couper & Peterson, 2016; Guidry, 2012; Lugtig & Toepoel, 2015; Mavletova, 2013; Mavletova & Couper, 2013; McClain, Crawford, & Dugan, 2012; Peterson, 2012). Mobile users also tend to select responses at the left end of horizontal ratings scales (McClain et al., 2012). No differences were, however, found in the number of selected items in check-all-that-apply questions (Lugtig & Toepoel, 2015; Peterson, 2012). The findings are mixed regarding item-nonresponse rate (Bosnjak et al., 2013; Guidry, 2012; Lugtig & Toepoel, 2015; Mavletova, 2013; McClain et al., 2012), the length of answers to open-ended questions (Antoun, Couper, & Conrad, 2017; Bosnjak et al., 2013; Buskirk & Andrus, 2014; Lugtig & Toepoel, 2015; Mavletova, 2013; Peterson, 2012; Toepoel & Lugtig, 2014; Zahariev, Ferneyhough, & Ryan, 2009), primacy effects (Lugtig & Toepoel, 2015; Mavletova, 2013) and straight-lining in grid questions (Antoun et al., 2017; Guidry, 2012; Lugtig & Toepoel, 2015; McClain et al., 2012).

Usability issues on small-screen mobile devices

Why is the smaller screen size of mobile devices a potential source of measurement error in surveys? If a website is not optimised for mobile devices using an adaptive or responsive web design, but uses a liquid design that scales the width of objects relative to the screen width, website content is displayed proportional to screen size. Therefore, the survey page is displayed smaller on small screens, which may have a negative impact on the visibility and the visual design of the survey as well as on aspects of questionnaire navigation.

On small-screen devices the question text is smaller and more difficult to read and the response options and navigation buttons are more difficult to select compared to devices with larger screens (Callegaro, 2010). The screenshots in Appendix Figure 1.1from a range of devices illustrate this problem: whereas the question is quite large on a 9.7-inch tablet, the font size and the size of radio buttons decrease on a 7.0-inch tablet and are considerably smaller on a 4.5-inch smartphone. Respondents with small screens may need to zoom into the survey page to facilitate reading and selecting buttons (Appendix Figure 1.2). The disadvantage of zooming in is that it requires respondents to perform additional navigation steps before they are able to view and answer the question, which potentially makes survey completion more burdensome. Another potential problem is that once respondents have zoomed in, the survey page may exceed the small screen. Survey participants may need to scroll to see elements of the page that are not visible anymore, such as parts of the question text or response options (Appendix Figure 1.3). The usability problems that arise due to a small screen size may affect various aspects of response quality, including completion times, breakoffs and answer patterns.

Completion times

If respondents need to scroll and zoom in when completing a questionnaire on small-screen devices, the additional time required for these navigation activities may add to the overall survey completion time (Couper & Peterson, 2016; Wells, Bailey, & Link, 2014). Couper and Peterson (2016), for example, examined question-level response times of web surveys taken on PCs and mobile devices and suggest that the higher need for scrolling on mobile devices is the major factor why mobile web surveys take longer compared to surveys completed on PCs. Beyond issues with questionnaire navigation, the speed in which the respondent is able to read the question on small screens may be slower due to the smaller font size (Couper & Peterson, 2016; Wells et al., 2014).

Hypothesis 1. Respondents with small-screen devices have longer survey completion times than respondents using larger screens.

Survey breakoff

Respondents may find survey completion on small-screen mobile devices more burdensome than on larger devices as they need to use smaller buttons and a smaller keyboard to record their answers. They may decide to switch to larger devices or may drop out of the survey if they perceive survey completion as too burdensome. Extant research on survey breakoff in web surveys identified respondent burden experienced during survey participation as well as technical problems as one of the most important predictors of dropouts (Galesic, 2006; Peytchev, 2009).

Hypothesis 2. Respondents with small-screen devices are more likely to drop out of the survey than respondents using larger screens.

Response distribution

Once respondents have zoomed in to be able to read the question, the survey page may exceed small screens: it can be expected that some response options are not visible and require respondents to scroll vertically or horizontally. Respondents may pay more attention to visually prominent options and may process them more thoroughly than those that are initially not visible (Couper, Tourangeau, Conrad, & Crawford, 2004). This expectation is supported by existing research: McClain et al. (2012) found that mobile respondents are more likely to select options which are at the left end of horizontal scales.

Hypothesis 3. Respondents with small-screen devices are less likely to select response options at the bottom of vertical questions than respondents using larger screens.

Length of open responses

Answering open-ended questions may be particularly burdensome on small devices because the keys of the digital keyboard are smaller, which makes typing more difficult. To reduce their effort, users with small screens may try to minimise typing and give shorter answers to open-ended questions (Mavletova, 2013; Peytchev & Hill, 2010). Extant research provides partial support for this hypothesis. While several studies found shorter answers to open-ended questions among mobile respondents compared to PC respondents (Lugtig & Toepoel, 2015; Mavletova, 2013; Peterson, 2012; Wells et al., 2014), other studies found no significant differences by device (Bosnjak et al., 2013; Buskirk & Andrus, 2014; Toepoel & Lugtig, 2014; Zahariev et al., 2009), or even longer answers among mobile respondents (Antoun et al., 2017).

Hypothesis 4. Respondents with small-screen devices provide shorter answers to open-ended questions than respondents using larger screens.

Straight-lining and check-all-that-apply questions

In order to compensate for the additional effort required on small screens, respondents may be more likely to satisfice when answering survey questions on small-screen devices. Satisficing in the survey context means that respondents carry out the cognitive response process less thoroughly and may take cognitive shortcuts (Krosnick, 1991; Krosnick & Alwin, 1987). Thereby, they may provide an answer which seems reasonable but deviates from their true response, resulting in measurement error. Satisficing respondents may tend to select the same response option for all items in a grid question (straight-lining) instead of providing a more differentiated response (Krosnick, 1991), and may select fewer items in check-all-that-apply questions (Lugtig & Toepoel, 2015; Peterson, 2012). Extant research has mixed findings with regard to straight-lining (Antoun et al., 2017; Guidry, 2012; Lugtig & Toepoel, 2015; McClain et al., 2012) and no significant findings related to check-all-that-apply questions (Lugtig & Toepoel, 2015; Peterson, 2012). A potential explanation for these observations is that the studies compared smartphones and tablets without considering screen size differences within mobile devices.

Hypothesis 5. Respondents with small-screen devices are more likely to straight-line in grid questions than respondents using larger screens.

Hypothesis 6. Respondents with small-screen devices select fewer items in check-all-that-apply questions than respondents using larger screens.

1.3. Data

The analysis is based on data from the web survey component of the Community Life Survey 2013-2014 which were collected from October 2013 to April 2014 (Cabinet Office, 2014; Hamlyn, Fitzpatrick, & Williams, 2015). It is a repeated cross-sectional survey of adults living in England that asks about involvement and social engagement within the local community. A stratified random sample of addresses was drawn using the Postcode Address File held by the UK Post Office. Each sampled address received a letter which invited the household member aged 16+ with the closest birthday to complete the web survey. Username and password were enclosed in the letter. To increase response rates, two reminder letters were sent and a £10 e-voucher was offered upon completion of the survey. A household response rate of 27 percent was achieved for the web survey component. The questionnaire was programmed with a liquid design that adjusts to the width of the screen and was not optimised for mobile devices. In the invitation letter, respondents were discouraged from using a smartphone but survey access was not blocked for any device. Questions were presented using a paging design with one question per screen.

In total, N = 4,698 respondents took part in the web survey: 3,638 respondents (77.4 percent) completed the survey on a desktop PC or laptop whereas 1,060 respondents used a mobile device (22.6 percent). Among the mobile device users, 951 used a tablet and 109 used a smartphone. The analyses of completion time, survey breakoff and length of open responses were carried out on a subset of the dataset as these variables are only available in a reduced dataset covering one of the four fieldwork quarters of the survey. In this dataset, data from N = 1,195 respondents are available: 887 survey participants used a desktop PC or laptop, 260 participants a tablet and 48 a smartphone. Table 1.1 summarises the sample size available for each of the six data quality indicators used in the analysis[1].

[1] Data of later survey waves have been requested from the field agency to increase sample size. While survey data were available, it was not possible to get access to the paradata that include information on screen size and other device characteristics.

Table 1.1. Sample size in the Community Life web survey 2013-2014.

Sample	PC/Laptop	Tablet	Smartphone
Community Life web survey full sample	**3,625**	**951**	**109**
Response distributions	3,589	945	109
Straight-lining	3,610	950	109
Check-all-that-apply questions	3,625	951	109
Community Life web survey sub-sample	**887**	**260**	**48**
Completion time	776	237	35
Survey breakoff	887	260	48
Length of open responses	815	243	40

To capture screen size and other technical details of mobile devices, including device type (smartphone, tablet), manufacturer (e.g., *Samsung*) and model (e.g., *Galaxy S3*), the user agent string (UAS) of the respondent's web browser was recorded at the beginning of the survey and sent to *Device Atlas* (http://deviceatlas.com/), a web service which parses the string and extracts mobile-specific information. Using this method, the screen size of desktop PCs or laptops could not be identified; in the following analyses, PCs and laptops are therefore treated as a single group.

Screen size was classified according to a classification used in the human-computer interaction literature (cf. Firtman, 2010): smartphones with a screen size of 4.0 inches or larger were defined as large smartphones and tablets with a screen size of at least 8.0 inches were classified as large tablets (Table 1.2). If information on screen size was missing for a particular device, it was imputed based on the manufacturer and the model type of the device if these types of information were available.

Table 1.2. Screen size of mobile devices in the Community Life web survey 2013-2014.

Screen size (in inches)		Min	Max	Mean	SD	N
Community Life web survey full sample						
Smartphone	Small	2.42	3.92	3.41	0.30	71
	Large	4.00	6.30	4.80	0.53	38
Tablet	Small	6.98	7.00	6.99	0.01	105
	Large	8.00	10.50	9.71	0.18	846
Community Life web survey sub-sample						
Smartphone	Small	3.50	3.70	3.51	0.04	22
	Large	4.00	5.70	4.80	0.42	26
Tablet	Small	6.98	7.80	7.01	0.11	51
	Large	8.00	10.10	9.69	0.33	209
Example devices						
Smartphone	Small	iPhone 4S, Blackberry Curve 9320				
	Large	Samsung Galaxy S3, HTC One S				
Tablet	Small	Google Nexus 7, Amazon Kindle Fire HD				
	Large	iPad Air, Samsung Galaxy Note 10.1				

The six data quality indicators were operationalised as follows. The numbers in parentheses index the questions in Appendix Table 1.5 that were used to create the indicators.

Completion times. Completion times were calculated as the difference between timestamps of the first and last survey page. This type of response time measurement is error-prone because it does not account for respondents who interrupted the survey. To exclude outliers, completion times below 10 minutes and those above 150 minutes were removed from the analysis, which reflects the fieldwork agency's procedure for cleaning completion time data in the Community Life Survey. The analysis was replicated using alternative cut-off points, but this did not change the conclusion.

Survey breakoff. A breakoff measure was created based on a process variable in the dataset which indicates the last question that the respondent completed. If this question corresponds to the last question of the questionnaire, the respondent completed the entire survey (*breakoff* = 0), otherwise the respondent dropped out (*breakoff* = 1).

Response distributions. The analysis of response distribution is based on a check-all-that-apply question with a list of 18 response categories which asks whether the respondent has donated any money to the listed charities (Q17). A dichotomous variable was created which takes on the value of 1 if at least one response option in the lower half of the response list was selected, i.e. one of the nine lowest response options, and the value of 0 if only options in the upper half of the list were selected. Respondents who refused to provide an answer or answered with "don't know" were excluded.

Length of open responses. The response length analysis is based on three open-ended questions which ask respondents about different aspects of their current or previous employment (Q25; Q26; Q27). A length measure was created by adding up the number of characters provided to the three questions. As the question is not applicable to respondents who have never worked before, the analysis base drops (Table 1.1) and additional selection effects may be introduced.

Straight-lining. The only grid question available in the survey was used to measure straight-lining. The question has four items with eight response options and asks about the respondent's relationship with family members and friends (Q1). Survey respondents are defined as straight-liners if they give the same response to all items of the grid. One tablet user and 15 PC/laptop users refused to answer the question and were excluded.

Check-all-that-apply questions. The analysis of item selection in check-all-that-apply questions is based on all 25 multi-choice questions available in the survey (Q2-Q24; Q28; Q29). These questions ask about different aspects of community involvement. For each respondent, the average number of selected items was calculated across all questions that were applicable to them.

1.4. Methods

As respondents were not randomly allocated to devices of different screen size but self-selected into using a particular device, observed differences in data quality may be confounded with selection effects and may be driven by differences in the sample composition. In the mixed-mode literature, several approaches have been applied to separate selection effects from measurement effects. The majority of mixed-mode studies rely on the back-door method which aims to control for covariates related to the selection propensity of survey modes (Cernat, 2015), for example using regression modelling (Jäckle, Roberts, & Lynn, 2010), propensity score matching (Lugtig, Lensvelt-Mulders, Frerichs, & Greven, 2011) or weighting techniques (Hox, De Leeuw, & Zijlmans, 2015). For reasons of simplicity, the present study uses the regression approach to disentangle selection and measurement effects.

The analysis is carried out in two steps for each of the quality indicators. First, bivariate statistics of quality indicators are presented across the five screen size groups. Second, multivariate regressions are fitted to estimate the impact of screen size on data quality while controlling for selection effects.

To include the five screen size categories in the regression model, four dummy variables were created by setting the PC/laptop group as baseline category. A separate indicator for mobile device type, contrasting smartphones with tablets to disentangle device and screen size effect, was not included, assuming that screen size is the major difference between smartphones and tablets that potentially affects data quality (Lugtig & Toepoel, 2015).

Socio-demographic characteristics that are related to the propensity to use mobile devices for survey completion (and may also be correlated with quality indicators) are added to the model to control for selection effects. Previous research identified age, gender, education, working status, income and household composition as the main predictors of whether a respondent accesses surveys on a mobile device (de Bruijne & Wijnant, 2014; Peterson, 2012; Toepoel & Lugtig, 2014). In the multivariate analyses of completion times and survey breakoff, the models also include a control variable for motivation which was found to have a substantive effect on both indicators (Gummer & Roßmann, 2015; Peytchev, 2009). As the survey is about social engagement within the community, a question about whether the respondent is involved in any volunteering activities is used as proxy variable for motivation. There are no other questions related to survey motivation or familiarity with mobile devices. The model predicting completion times furthermore includes a count variable indicating how many items the respondent was asked in the survey as respondents may have answered a different set of questions due to routing.

Linear regressions are fitted to model continuous quality indicators (completion times, mean number of items selected in check-all-that-apply questions), logistic regressions for binary indicators (survey breakoff, response distributions, straight-lining), and a negative binomial regression for a count indicator (length of open responses).

The socio-demographic variables included in the multivariate models contain a considerable amount of missing data. In the full sample, around 27 percent of cases have missing values in at least one of the socio-demographic variables, particularly in the income variable. A complete-case analysis considering only respondents with non-missing values on all variables potentially leads to biased results. Therefore, missing values in the variables age, gender, education, employment status, household composition, income and volunteering were imputed using multiple imputation with n = 5 imputations. The imputation was conducted in SPSS using the fully conditional specification (FCS) algorithm.

1.5. Results

H1. Completion times

First, it was expected that respondents who use smaller screens take longer to complete the survey because they may need to scroll and zoom and may find it more difficult to read text with a small font size (H1). A bivariate analysis shows that mobile participants need on average 31-37 minutes to complete the survey while PC participants need 35 minutes (Table 1.3). Surprisingly, respondents using large tablets have on average the longest completion times whereas the completion times of the other four screen size groups are on a similar level of around 31-35 minutes. The difference in mean completion times between screen size groups is not statistically significant as determined by a one-way ANOVA, $F(4, 1043) = 1.606$, $p > 0.05$.

Table 1.3. Data quality indicators by screen size.

	Smartphone		**Tablet**		**PC**
	Small	Large	Small	Large	
Mean completion time (in minutes)	34.2	31.0	30.7	37.3	34.9
	(15)	(20)	(48)	(189)	(776)
Percent breaking off	31.8	15.4	7.8	4.3	9.4
	(22)	(26)	(51)	(209)	(887)
Percent selecting response option in	74.6	78.9	76.0	80.3	81.0
lower half of question	(71)	(38)	(104)	(841)	(3,589)
Mean length of open responses	49.4	85.5	74.8	93.3	109.3
(in characters)	(21)	(19)	(43)	(200)	(815)
Percent straight-lining	5.6	2.6	0.0	1.2	1.2
	(71)	(38)	(105)	(845)	(3,610)
Mean number of responses in	2.1	2.3	2.3	2.2	2.3
check-all-that-apply questions	(71)	(38)	(105)	(846)	(3,625)

Note. Sample size in parentheses.

In the second step, a multivariate linear regression is fitted to model completion times while controlling for selection effects (Table 1.4). Similar to the bivariate analysis, the model shows no significant differences in completion times between screen sizes.

Table 1.4. Data quality indicators, controlling for screen size and respondent characteristics.

	Completion times	Survey breakoff	Response distribution	Response length	Straight-lining	Check-that-apply questions
Intercept	-11.08	-2.61***	1.85***	4.19***	-2.79***	2.23***
Small SP	2.23	1.90***	-0.40	-0.82**	1.04	-0.21**
Large SP	0.57	0.87	-0.18	-0.25	0.26	-0.10
Small T	0.06	0.04	-0.34	-0.33*	-16.90	-0.10
Large T	2.49	-0.65	-0.06	-0.21**	-0.15	-0.11***
PC/Laptop	*-Baseline-*	*-Baseline-*	*-Baseline-*	*-Baseline-*	*-Baseline-*	*-Baseline-*
Age	0.20***	0.01	0.00	0.01**	-0.03**	0.00
Male	0.75	-0.12	-0.18*	-0.24**	-0.78*	-0.04**
A-levels	-0.75	-0.30	-0.01	0.34***	-0.14	0.19***
Employed	-2.13	-0.08	-0.14	0.12	0.07	0.04*
High income	-1.07	0.00	-0.13	0.00	-0.05	0.06**
Living alone in HH	2.03	-0.25	-0.23**	0.02	0.10	0.02
Volunteering	1.15	0.07	–	–	–	–
# Items completed	0.12***	–	–	–	–	–
N	1,048	1,195	4,643	1,098	4,669	4,685
Regression model	OLS	Logistic	Logistic	Negative binomial	Logistic	OLS

Note. SP = Smartphone. T = Tablet.
* $p < .05$, ** $p < .01$, *** $p < .001$. Results from multiple imputation.

As the effect of screen size on completion times is not statistically significant, no evidence is found for Hypothesis 1. Respondents who take the survey on a small-screen mobile device need on average the same amount of time for survey completion as respondents using larger screens.

The analysis, however, does not allow drawing conclusions about whether small-screen respondents have problems reading the small text, whether they actually scroll and zoom to a larger extent than users with large screens or whether these two factors add substantially to the overall survey completion time. Further research on the level of page-level response times is needed to better understand survey experience on small-screen devices, for example by collecting paradata that indicate whether the respondent has scrolled or zoomed on a particular survey page.

H2. Survey breakoff

Second, it was hypothesised that mobile respondents using small screens are more likely to break off the survey than respondents with larger screens (H2). The bivariate analysis supports the theoretical expectations (Table 1.3). Respondents with smaller screens have a higher propensity to drop out of the survey: almost one third (31.8 percent) of respondents using small smartphones and 15.4 percent of those using large smartphones failed to finish the survey. Among respondents with small tablets 7.8 percent dropped out and among those with large-screen tablets only 4.3 percent. Surprisingly, 9.4 percent of PC respondents dropped out of the survey, which lies in-between the figures for smartphone and tablet respondents. The Chi-square test of independence indicates that the relationship between screen size and breakoff rate is significant, $\chi^2 (4) = 21.219$, $p < 0.001$. To decompose the Chi-square test statistic and understand which screen size groups are significantly different from each other with regard to survey breakoff, standardised residuals can be considered. The standardised residuals for the small smartphone group and the large tablet group are significant at $p < 0.05$, which implies that the significant association between screen size and survey breakoff is mainly driven by the high percentage of breakoffs among small smartphone users and the low percentage among large tablet users.

Do these findings also hold true when controlling for selection effects? The results of the logistic regression indicate that respondents with small smartphones are significantly more likely to drop out of the survey compared to PC/laptop respondents, $p < 0.001$ (Table 1.4). The effects of the other screen size groups compared to PC/laptop respondents also point in the expected direction but are not statistically significant; large tablet respondents even have a lower propensity to break-off than PC/laptop respondents. Interestingly, breakoff probability falls monotonically with screen size among the mobile device groups. It is largest for the small smartphone group, smaller for the large smartphone group and small tablet group and smallest for the large tablet group.

The findings of the bivariate and multivariate analysis support Hypothesis 2. Respondents with smaller screens, particularly a small smartphone, have a higher propensity to drop out of the survey compared to PC/laptop respondents, presumably due to the higher burden that they might have experienced.

In addition to the overall breakoff rate, it was also examined whether respondents are more likely to drop out at particular question formats, such as grid or open-ended questions which may be more burdensome to complete on small-screen devices. The findings, however, suggest that dropouts do not cluster around specific survey items (analysis not shown).

H3. Response distribution

Third, it was expected that response options in the lower half of a long vertical list of options are likely to exceed small screens and that small-screen users are hence less likely to select an option in the lower part (H3).

A bivariate analysis suggests that the proportion of respondents who selected at least one item in the lower part of the question is similar across all screen size groups and ranges between 74.6 percent and 81.0 percent (Table 1.3). The maximum difference of 6.4 percentage points is between the small smartphone and the PC/laptop group. However, a Chi-square test of independence shows that overall, there is no statistically significant association between screen size and the distribution of responses, χ^2 (4) = 3.483, $p > 0.05$. Furthermore, the standardised residuals are not statistically significant for any of the groups.

A logistic regression predicting the likelihood to select an item in the lower half of the response list and controlling for selection effects confirms the bivariate findings (Table 1.4). None of the screen size effects are statistically significant at $p < 0.05$.

Bivariate and multivariate findings do not provide support for Hypothesis 3. Respondents of all screen size groups have a similar propensity to select items in the lower part of the response list. A possible explanation is that the question may not have exceeded the screen on most small-screen devices, for example as respondents may have used their device in vertical orientation where all response options were initially visible. An alternative explanation is that respondents had to scroll down in any case to press the “next” button and to proceed to the next survey screen, so that the lower response options became visible. However, more detailed information about which part of the survey page is actually visible would be required to further explore the effect of small screens on the distribution of responses.

H4. Length of open responses

Regarding the length of open responses, it was hypothesised that respondents with small-screen devices provide shorter answers to open-ended questions than those with large-screen devices as typing on small screens may be more burdensome (H4).

A bivariate analysis confirms the theoretical expectations (Table 1.3). Respondents with small smartphones provide the shortest responses to open-ended questions, with 49.4 characters on average, whereas PC/laptop users have the longest open answers, with an average length of 109.3 characters. The large smartphone, small tablet, and large tablet groups have an average response length which lies in-between as they provide responses of around 74.8 to 93.3 characters. A one-way ANOVA indicates that the difference in means across the five screen size groups is significant, F (4, 1093) = 2.567, $p < 0.05$.

To control for selection effects, a negative binomial regression is fitted to predict the number of characters in the three open-ended questions (Table 1.4). The model shows that small smartphone users provide significantly shorter answers than PC/laptop users ($p < 0.01$), but also small tablet users ($p < 0.05$) and large tablet users ($p < 0.01$) have a response length that is significantly shorter than the length of PC/laptop users. Surprisingly, the response length of large smartphone users is not significantly different from PC/laptop users when controlling for selection effects. Among the screen size groups with significant effects, the effect has the expected magnitude: it is largest for the small smartphone group and smaller for the small tablet group as well as the large tablet group.

The multivariate analysis provides evidence for Hypothesis 4. When controlling for selection effects, users with small mobile device screens, in particular small smartphone users, provide significantly shorter answers to open questions compared to those using PCs or laptops. The response length of small tablet users and large tablet users is also shorter than the length of PC/laptop users although the difference is not as large as between small smartphone and PC/laptop users.

H5. Straight-lining

Respondents using small-screen devices were expected to take cognitive shortcuts to reduce respondent burden due to usability problems on small screens. As a first indicator for survey satisficing, the occurrence of straight-lining response patterns in grid questions is examined (H5).

A bivariate analysis of straight-lining across the four screen size groups shows that there is a higher proportion of straight-lining respondents among the small smartphone group (5.6 percent) than among the large smartphone (2.6 percent), the large tablet group (1.2 percent) and the PC/laptop group (1.2 percent) (Table 1.3). Straight-lining is non-existent in the small tablet group. A Chi-square test of independence indicates that there is a statistically significant association between screen size and the occurrence of straight-lining, χ^2 (4) = 13.499, $p < 0.01$. Standardised residuals reveal that the significant association is mainly driven by the small smartphone group as only the standardised residuals for this group are significant ($p < 0.01$).

As the next step, a logistic regression is fitted to predict the likelihood to straight-line while controlling for socio-demographic characteristics that are related to the use of mobile devices in surveys (Table 1.4). The model indicates that none of the screen size effects are statistically significant at $p < 0.05$.

Although the bivariate analysis suggests that small smartphone users have the highest propensity to straight-line in grid questions, the significant effect disappears when controlling for selection effects in the multivariate model. Therefore, Hypothesis 5 cannot be supported. This finding suggests that small-screen respondents are not more prone to satisficing response behaviour in grid questions than respondents using larger screens.

H6. Check-all-that-apply questions

It was also expected that small-screen users are more likely to satisfice in check-all-that-apply questions, so that they select fewer items compared to survey participants using larger devices (H6).

A bivariate analysis shows that all screen size groups select on average 2.1 to 2.3 items per check-all-that-apply question (Table 1.3). The small smartphone group has the lowest mean of 2.1 whereas the other screen size groups have a mean of 2.2-2.3. A one-way ANOVA indicates that overall, the difference in means between the five screen size groups is significant, F (4, 4680) = 7.691, $p < 0.001$. A Tukey post-hoc test reveals that the difference between the small smartphone group and the PC/laptop group ($p < 0.05$) as well as between the large tablet group and the PC/laptop group ($p < 0.001$) are statistically significant.

A linear regression which predicts the mean number of selected items in check-all-that-apply questions confirms the bivariate findings (Table 1.4): the small smartphone group (p < 0.01) and the large tablet group (p < 0.001) select significantly fewer items in check-all-that-apply questions than the PC/laptop group. Surprisingly, there is no significant difference between the large smartphone group and the PC/laptop group as well as between the small tablet group and the PC/laptop group.

The bivariate and multivariate findings support Hypothesis 6. It seems that small-screen users, in particular small smartphone users, take more cognitive shortcuts in check-all-that-apply questions than users with PCs or laptops, presumably as they experience higher respondent burden when answering questions of this format. Interestingly, also large tablet users select fewer items than PC/laptop users although the difference is not as large as between small smartphone and PC/laptop users.

1.6. Discussion

The aim of the present study was to understand how screen dimensions of mobile devices affect the quality of web survey data. Using data from an online survey in England, it can be found that the use of small smartphones in surveys is detrimental to response quality if the questionnaire is not optimised for smartphone use. The results suggest that response quality mainly differs between small smartphones with a diagonal screen size of below four inches and larger mobile devices. Participants using small-screen smartphones are significantly more likely to drop out of the survey than survey participants who use larger devices. In addition, users of small smartphones provide the shortest answers to open-ended questions and select fewer items in check-all-that-apply questions than participants using PCs or laptops. However, contrary to what was expected, the study did not find any effect of screen size on completion times, on the response distribution of a question with a long response list and on straight-lining.

The present study provides evidence that surveys ought to provide a mobile-optimised questionnaire design to mobile respondents, in particular to respondents who use smartphones with a screen size of below four inches. A mobile-optimised design seems to be particularly important if the questionnaire contains a considerable amount of open-ended questions or check-all-that-apply questions as both question formats seem to be more burdensome on small-screen smartphones. The other main problem is the higher dropout rate of users with small smartphones, which may result in nonresponse error. The response quality of surveys taken on other mobile devices, however, seems to be relatively comparable, regardless of whether it is a smartphone with a screen size of at least four inches or a tablet of any size.

The analysis has two main limitations. First, the sample size of the available data is small, especially of the large smartphone and small smartphone group, which resulted in small statistical power and may have been a possible reason for some of the non-significant findings. The results presented here may therefore be a conservative estimate of the effects of screen size on data quality. Second, as this study is based on survey data where respondents were not randomly allocated to devices, observed differences in data quality may be confounded with selection effects. Although the fitted multivariate regression models control for variables related to mobile device use in surveys, measurement and selection effects may have not been fully disentangled with this approach.

The present study could be extended in several ways. Future research using larger samples could investigate whether there is an interaction effect of screen dimensions and respondent characteristics. It can be expected, for instance, that older or less tech-savvy respondents experience higher respondent burden when using mobile devices for survey completion than those who are younger or use mobile devices more frequently. Data quality of respondents who are more motivated and more interested in the survey topic may also be less affected by screen size. The small sample size of the available dataset does not allow to model interaction terms in the present study.

Furthermore, it would be interesting to analyse the response distribution of questions with horizontal response scales. In this study, it was only possible to analyse the answer distribution of questions in vertical format because the present survey did not administer horizontal questions other than the grid. However, small smartphone screens may particularly affect the response distribution of horizontal questions because smartphones are usually used in vertical orientation and it is likely that the right end of horizontal scales exceeds the screen.

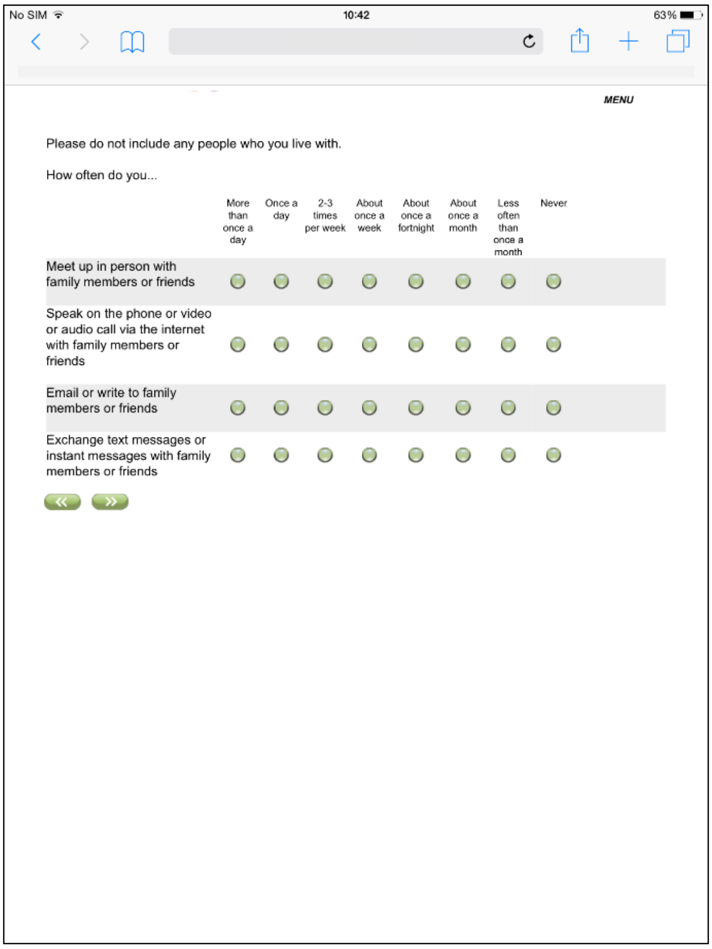
Paradata on device orientation, questionnaire navigation, such as scrolling and zooming, and screen resolution would be helpful to better understand how the questionnaire is actually displayed on a range of devices. In this study, it was assumed that questionnaire navigation may be one of the factors which makes survey completion on small-screen devices more burdensome. However, it could not be validated whether mobile users actually scroll and zoom to a larger extent on small screens than they would do on larger screens.

Finally, it would be interesting to explore how screen size affects the data quality of mobile-optimised compared to non-optimised questionnaires, ideally using an experimental approach. If the survey is optimised for small screens, small-screen smartphone users may not encounter the usability issues documented in this study and may be able to provide survey data of similar quality.

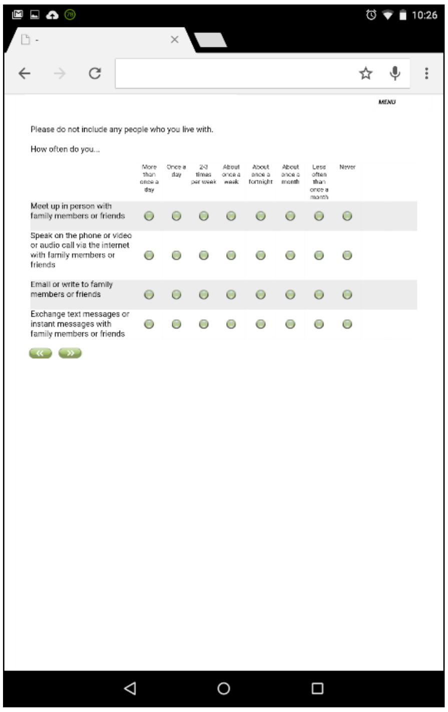
1.7. Appendix

Figure 1.1. Survey displayed on mobile devices with different screen size (scale 1:2).

a) 9.7-inch tablet *(iPad 4)*



b) 7.0-inch tablet (*Nexus 7*)



c) 4.5-inch smartphone (*Motorola Moto G*)

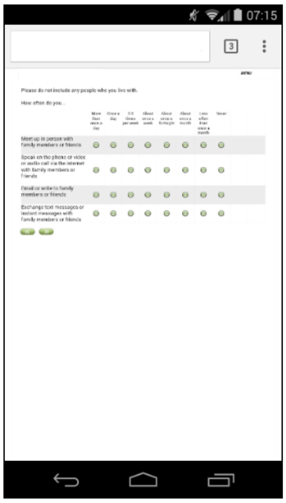
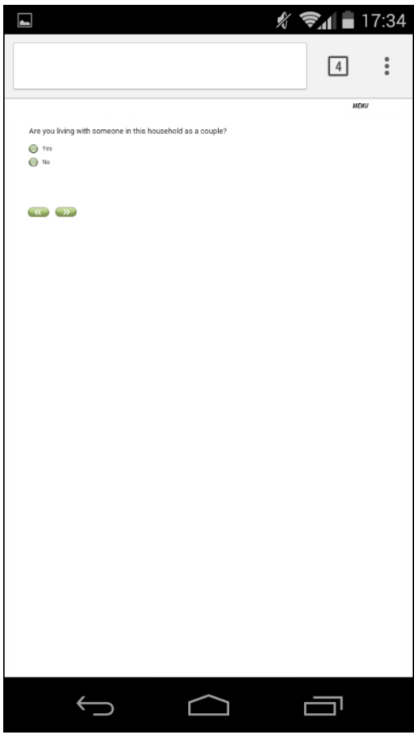


Figure 1.2. Zooming in on 4.5-inch smartphone (*Motorola Moto G*).

a) Not zooming in



b) Zooming in

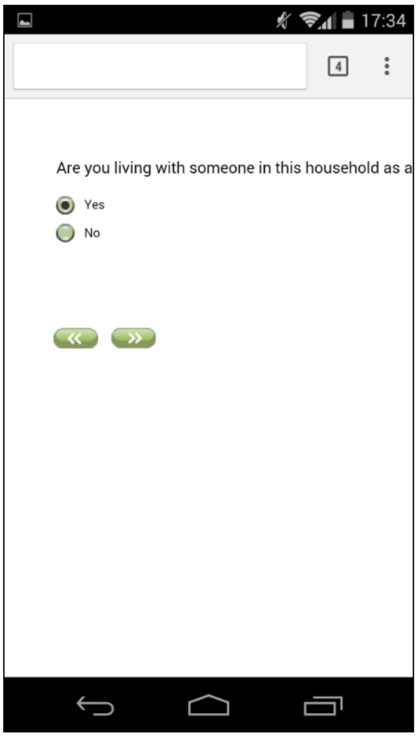
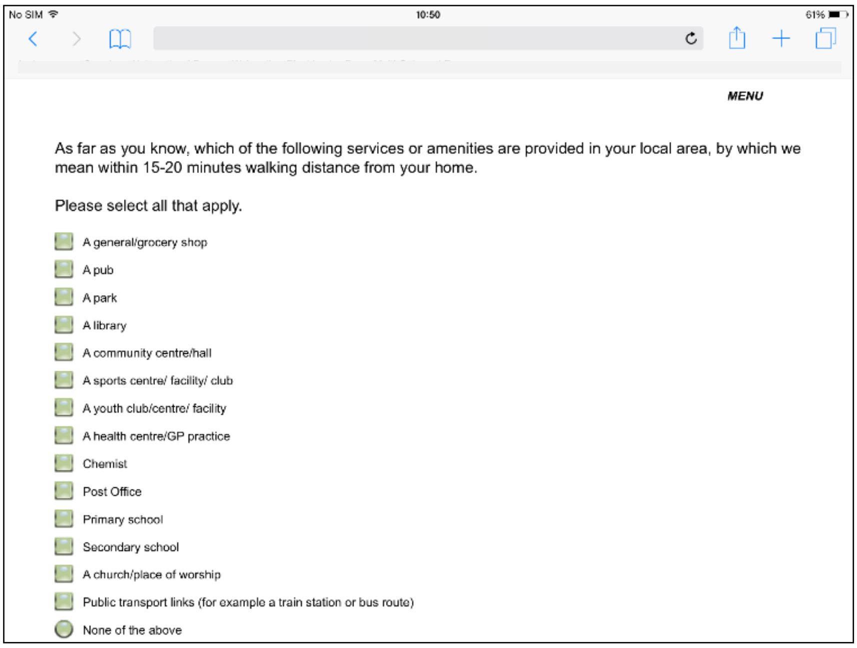


Figure 1.3. Question displayed in horizontal orientation on 9.7-inch tablet and 7.0-inch tablet.

a) 9.7-inch tablet *(iPad 4)*



b) 7.0-inch tablet *(Nexus 7)*

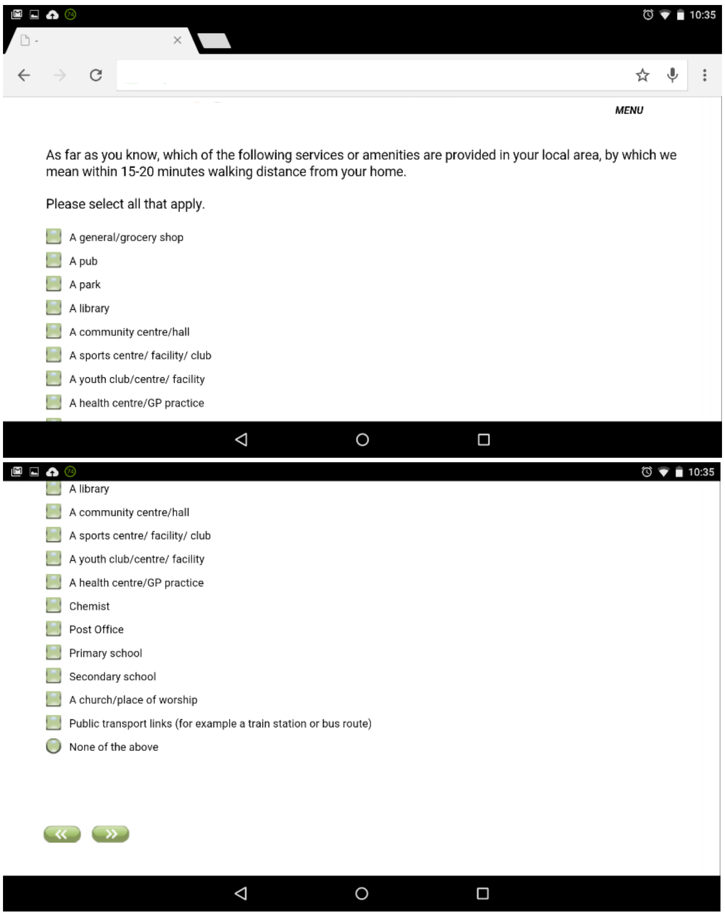


Table 1.5. Questionnaire.

	Question wording
Q1	Please do not include any people who you live with. How often do you … *More than once a day; Once a day; 2-3 times per week; About once a week; About once a fortnight; About once a month; Less often than once a month; Never* Meet up in person with family members or friends Speak on the phone or video or audio call via the internet with family members or friends Email or write to family members or friends Exchange text messages or instant messages with family members or frien*ds*
Q2	*[If respondent chats to neighbour once or twice a month, less than once a month, or never]* Why would you say you don't chat to your neighbours more often? Please select all that apply. *Prefer to keep myself to myself; Don't have time; Prefer to choose my friends/have enough friends already; Neighbours speak different language/have different culture; Don't trust/get on with my neighbours; Have no need to speak to neighbours; Don't feel I know my neighbours well enough; Nothing in common with my neighbours; I'm new to the area; Don't see neighbours very often; Have an illness/disability that prevents me from going out much; People just don't speak to each other round here; Other (specify)*
Q3	As far as you know, which of the following services or amenities are provided in your local area, by which we mean within 15-20 minutes walking distance from your home. Please select all that apply. *A general/grocery shop; A pub; A park; A library; A community centre/hall; A sports centre/facility/club; A youth club/centre/facility; A health centre/GP practice; Chemist; Post Office; Primary school; Secondary school; A church/place of worship; Public transport links (for example a train station or bus route); None of the above*
Q4	In the last 12 months, that is since the [date], have you done any of the following? Please select all that apply.

	Contacted a local official such as local councillor, MP, government official, mayor, or public official working for the local council or Greater London Assembly (Please do not include any contact for personal reasons, e.g. housing repairs or contact through work); Attended a public meeting or rally, taken part in a public demonstration or protest; Signed a paper petition or an online/e-petition; None of these
Q5	*[If age > 18]* In the last 12 months, that is since the [date], have you done any of the following? Please select all that apply. *Taken part in a consultation about local services or problems in your local area through completing a paper or online questionnaire; Taken part in a consultation about local services or problems in your local area through attending a public meeting; Taken part in a consultation about local services or problems in your local area through being involved in a face-to-face or online group; None of these*
Q6	In the last 12 months, that is since the [date] have you done any of the things listed below? Please include any activities you have already mentioned. Please do not include any activities related to your job. Please select all that apply. *Been a local councillor (for local authority, town or parish); Been a school governor; Been a volunteer Special Constable; Been a Magistrate; None of these*
Q7	And again in the last 12 months, that is since the [date] have you been a member of any of the following groups? Please include online groups and any activities you have already mentioned. Please do not include any activities related to your job. Please select all that apply. *A group making decisions on local health services; A decision making group set up to regenerate the local area; A decision making group set up to tackle local crime problems; A tenants' group decision making committee; A group making decisions on local education services; A group making decisions on local services for young people; Another group making decisions on services in the local community; None of these*
Q8	If you wanted to influence decisions in your local area **how** would you go about it? Please select all that apply.

	Contact the council/a council official; Contact my councillor; Contact my MP; Contact my assembly member (for London); Sign a paper petition; Sign an e-petition/online petition; Organise a paper petition; Organise an e-petition/online petition; Attend a council meeting; Attend a public meeting; Contact local media or journalists; Organise a group (e.g. campaign/action group); Something else (please specify)
Q9	Which, if any, of these might make it easier for **you** to influence decisions in your local area? Please select all that apply. *If I had more time; If the council got in touch with me and asked me; If I could give my opinion online/by email; If I know what issues were being considered; If it was easy to contact my local councillor; If I knew who my local councillor was; If I could get involved in a group (not online) making decisions about issues affecting my local area/neighbourhood; If I could get involved in an online group making decisions about issues affecting my local area/neighbourhood; Something else (specify)*
Q10	*[If respondent has taken part in, supported or helped any groups, clubs or organisations over the last 12 months]* In the last 12 months, that is since [date], have you given **unpaid** help to any of the **groups, clubs or organisations** you've just selected in any of the following ways? Please select all that apply. *Raising or handling money/taking part in sponsored events; Leading a group/member of a committee; Getting other people involved; Organising or helping to run an activity or event; Visiting people; Befriending or mentoring people; Giving advice/information/counselling; Secretarial, admin or clerical work; Providing transport/driving; Representing; Campaigning; Other practical help (e.g. helping out at school, shopping); Any other help; None of the above*
Q11	*[If respondent has taken part in, supported or helped any groups, clubs or organisations over the last 12 months and has given unpaid help]* How did you find out about opportunities to give unpaid help to these groups, clubs or organisations? Please select all that apply. *Through previously using services provided by the group; Frome someone else already involved in the group; From a friend not involved in the group/by word*

	of mouth; Place of worship; School, college, university; Doctor's surgery/Community centre/Library; Promotional events/volunteer fair; Local events; Local newspaper; National newspaper; TV or radio (local or national); Volunteer bureau or centre; Employer's volunteering scheme; www.do-it.org.uk; National Citizen Service; Other internet/organisation website; Other way
Q12	*[If respondent has taken part in, supported or helped any groups, clubs or organisations over the last 12 months and has given unpaid help]* Now thinking about the unpaid help you've given as part of a group, club or organisation in the last 12 months, have you mixed with any people who are different to you in terms of the following types of characteristics? Please select all that apply. *People of different age groups; People of different ethnic groups or religions; People with a different social or educational background; People who live in different neighbourhoods; None of these*
Q13	*[If respondent has taken part in, supported or helped any groups, clubs or organisations over the last 12 months and has given unpaid help]* People do unpaid work or give help to all kinds of groups for all kinds of reasons. Thinking about all the groups, clubs or organisations you have helped over the last 12 months, did you start helping them for any of the following reasons? Pick the reasons that were most important to you. You can choose up to five reasons. *I wanted to improve things/help people; I wanted to meet people/make friends; The cause was really important to me; My friends/family did it; It was connected with the needs of my family/friends; I felt there was a need in my community; I thought it would give me a chance to learn new skills; I thought it would give me a chance to use my existing skills; It helps me get on in my career; It's part of my religious belief to help people; It's part of my philosophy of life to help people; It gave me a chance to get a recognised qualification; I had spare time to do it; I felt there was no one else to do it; None of the above*
Q14	*[If respondent has not been involved with any groups, clubs or organisations in the last 12 months but has been involved in the last five years]*

	What would you say were the main reasons for stopping your involvement with giving unpaid help to any groups, clubs or organisations? Please check all that apply. *Not enough time – due to changing home/work circumstances; Not enough time – getting involved took up too much time; Group/club/organisation finished/closed; Moved away from the area; Due to health problems or old age; Group/club/organisation wasn't relevant to me anymore; Lost interest; It was a one-off activity or event; Felt I had done my bit/some else's turn to get involved; Got involved in another activity instead; Didn't get asked to do the things I'd like to; Felt the group/club/organisation was badly organised; Felt my efforts weren't always appreciated; It was too bureaucratic/too much concern about risk and liability; Activity linked to my school/college/university/job I have now left; Other reason (specify)*
Q15	*[If respondent has not taken part in, supported or helped any groups, clubs or organisations over the last 12 months but would like to spend time helping groups, clubs or organisations]* Listed below are some reasons people have given about why they don't give unpaid help to groups, clubs or organisations. Which, if any, of these are reasons why you don't give unpaid help to groups, clubs or organisations more regularly? Please select all that apply. *I have work commitments; I have to look after children/the home; I have to look after someone who is elderly or ill; I have to study; I do other things with my spare time; I'm not the right age; I don't know any groups that need help; I haven't heard about opportunities to give help/I couldn't find opportunities; I'm new to the area; I have never thought about it; I have an illness or disability that I feel prevents me from getting involved; It is not my responsibility; Other reason*
Q16	In the last 12 months, that is since [date], have you done any of these things, **unpaid**, for someone who was **not a relative**? Please select all that apply. *Keeping in touch with someone who has difficulty getting gout and about (visiting in person, telephoning or e-mailing); Doing shopping, collecting pension or paying bills; Cooking, cleaning, laundry, gardening or other routine household jobs; Decorating, or doing any kind of home or car repairs;*

	Babysitting or caring for children; Sitting with or providing personal care (e.g. washing, dressing) for someone who is sick or frail; Looking after a property or a pet for someone who is away; Giving advice; Writing letters or filling in forms; Representing someone (for example talking to a council department or to a doctor); Transporting or escorting someone (for example to a hospital or an outing); Anything else; No help given in last 12 months
Q17	**In the past 4 weeks**, have you given any money to charity in any of the following ways or through any other method? Please **exclude** donating goods or prizes. Please select all that apply. ***Donations.*** *Money to collecting tins (e.g. door-to-door, in the street, in a pub, at work, on a shop counter, etc.); Collection at church, mosque or other place of worship; Collections using a charity envelope/cheque in the post; Covenant or debit from salary, payroll giving; Donation – via direct debit, standing order; Giving to people begging on the street; Donation – in person or on phone (excluding online or via text message); Donation – online/via website; Donation – by text message; Donation – via an ATM/cash machine.* ***Purchases/fundraising.*** *Buying raffle tickets (NOT national or health lottery); Buying goods from a charity shop, catalogue or online; Making a purchase where the price includes a charitable donation/or where you can add a charitable donation to the purchase; Buying tickets or spending money at fundraising events (e.g. charity dinners, fetes, jumble sales).* ***Sponsorship.*** *Sponsorship (not online); Sponsorship (online).* ***Other.*** *Other method of giving (excluding donating goods or prizes) (specify); Did not give to charity.*
Q18	*[If respondent has donated money to charity]* To which, if any, of these types of cause have you given money in the **past 4 weeks**? Please select all that apply. *Schools, colleges, universities or other education; Children or young people (outside school); Sports/exercise; Religion/Place of Worship; The elderly; Overseas Aid/Disaster Relief; Medical Research; Hospital and Hospices; Physical/Mental Healthcare/Disabled people (including blind or deaf people); Social Welfare; The arts and museums; Hobbies/Recreation/Social clubs; Other (specify)*

Q19	The following list contains some things that people said would encourage them to give to charity. Would any of these things encourage you to start giving to charity or to increase the amount you currently give? Please select all that apply. *Having more information about the different charities or organisations that I could support; Knowing that my money is going to be spent locally; Receiving letter/email of thanks from the charity or organisation; Receiving information from the charity or organisation explaining what has been done with my donation; Being asked by the charity or organisation to increase my donation; Confidence that the charity or organisation uses the money efficiently; Being able to give money by tax efficient methods (e.g. Gift Aid, giving via self assessment); More generous tax relief (e.g. tax relief on the values of gifts or shares, land or buildings given to charities); Being asked by a friend or family member; If I had more money; If payroll giving became available to me; If the charity helped me or someone close to me; None of these*
Q20	Have you personally been involved in helping out with any of these types of activity in your local area in the last 12 months? Please only include unpaid involvement. Do not select any activities where you only signed a petition but took no further action. In the last 12 months I have been involved in … *Trying to set up a new service or amenity to help local residents; Trying to stop the closure of a local service or amenity; Trying to stop something happening in my local area; Running local services on a voluntary basis (e.g. childcare, youth services, parks and community centres); Organising a community event such as a street party; Another issue affecting my local area (specify); None of these*
Q21	*[If respondent has been involved in activities in local area]* In the last 12 months, in what ways have you been involved in all of these activities or issues? Please select all that apply. *I started up the activity (solely or jointly); I managed the activity (solely or jointly); I participated in a discussion on this issue/event (online or in person); I helped fundraise; I got more people involved; I contributed specialist skills; I donated money; I offered non-monetary donations or contributions; I campaigned; I helped raise awareness locally; I helped organise a petition; I signed a petition; I offered other practical support; Other (specify)*

Q22	*[If respondent has been involved in activities in local area]* How did you find out how to get involved with all of these activities or issues? Please select all that apply. *I was the person/one of the people who started the action; I was asked to get involved by someone I already knew; I was asked to get involved by someone I hadn't previously known; I saw a leaflet/poster/flyer; I read about it in the local newspaper; Via a local community/neighbourhood/residents group/Via an online forum or social network site; Other (specify)*
Q23	*[If respondent has been involved in activities in local area]* People get involved with activities and issues like this for all sorts of reasons. Thinking about all of the local issues or activities you have been involved in over the last 12 months, did you do this for any of the reasons listed below? Pick the reasons that were most important to you. Please select all that apply. *I wanted to serve my community/felt it was my responsibility; I wanted to improve local services/not happy with existing provisions; I wanted to resolve an issue; My political beliefs; An earlier positive experience of getting involved; I was asked to get involved; I wanted to have my say; I wanted to meet people/make friends; It was connected with the needs of my family/friends; I thought it would give me a chance to learn new skills/use my existing skills; I thought it would help my career; I had spare time to do it; Because I wanted an interest outside of work; Other (please specify)*
Q24	Which language do you speak most often at home? If you speak English and another language equally please select both of these codes. Otherwise please choose your main language. *English; Other language*
Q25	*[If respondent has ever had a paid job]* What [does/did] the firm/organisation you [work/worked] for mainly make or do (at the place where you [work/worked])? Please provide as much detail as possible. *[Textbox]*
Q26	*[If respondent has ever had a paid job]* What was your [main job in the week ending Sunday the [date]/your last main job]? Please enter your full job title.

	[Textbox]
Q27	*[If respondent has ever had a paid job]* What [do/did] you mainly do in your job? Please give as much detail as possible, and include any special qualifications and training needed to do the job. *[Textbox]*
Q28	*[If age < 70]* Do you have any qualifications from each of the following … Please select all that apply. *From school, college or university; Connected with work; From government schemes; No qualifications*
Q29	*[If respondent discussed any elements of the survey with other household members]* What did you discuss with other members of your household? Please select all that apply. *The survey in general (e.g. what it is about, whether to complete it); The broad content of the survey (e.g. what sort of issues it covers); How to respond to specific questions in the survey; Other (please type in details of what else you discussed)*

2. Do distractions during web survey completion affect data quality? Findings from a laboratory experiment

Abstract

This paper reports on results from a laboratory experiment that examines how distractions during web survey completion influence data quality. Participants were randomly assigned to experimental groups using a 2 (device type) x 3 (form of distraction) between-subject factorial design. They were asked to complete a web questionnaire on either a PC or a tablet and were allocated to one of three distraction conditions: a) the presence of other people in the room who have a loud conversation, b) the presence of music, or c) no distraction. The study examines the effect of distraction on various data quality measures, including item-nonresponse, straight-lining, extreme response styles, and response consistency.

2.1. Introduction

Web surveys are increasingly considered as cost-effective mode of data collection for large-scale social surveys. Existing face-to-face surveys have introduced mixed-mode approaches that include web (e.g., Jäckle, Lynn, & Burton, 2015) and a number of probability online and mixed-mode panels have recently been established in the United States and Europe that aim to cover the general population (Blom et al., 2015; Hays, Liu, & Kapteyn, 2015). Compared to interviewer-administered surveys, web surveys have considerable benefits related to costs, timeliness, and the possibility to implement rich visual information (Callegaro et al., 2015; Couper, 2008), but they are not free of disadvantages.

One of the problems of web survey data collection is that survey designers lose control over the environment in which respondents complete the survey. Web respondents are not ‘supervised’ by an interviewer but can decide when and where to fill in the questionnaire. As a consequence, they might experience various forms of distractions while completing the survey. These distractions might either come from the environment, over which the respondent has less control, or might be a form of multitasking that the respondent seeks deliberately (Ansolabehere & Schaffner, 2015; Sendelbah, Vehovar, Slavec, & Petrovcic, 2016; Zwarun & Hall, 2014). Distracted and multitasking respondents might not be able to fully concentrate on the survey task and to accurately process their responses, which potentially affects the quality of data they provide.

Distractions and multitasking seem to occur frequently among web respondents. Ansolabehere and Schaffner (2015), for example, examined the incidence of distractions across four political online surveys that were fielded in the United States and vary by survey length and sample design (one probability sample, three non-probability samples). For short surveys of around ten minutes, the authors found that 22-37 percent of respondents report that they were involved in other activities while completing the survey whereas for longer surveys of around 30 minutes, around half of respondents report that they were distracted. The study suggests that watching TV, having a conversation with another person, and making a phone call are the most common forms of distraction during web survey completion. Using a different approach, Sendelbah, Vehovar, Slavec, and Petrovcic (2016) studied the occurrence of multitasking in a student web survey in Slovenia by analysing web paradata: they recorded *focus-out events* which capture whether respondents opened another browser window during survey completion, and collected question-level response times. Combining these two measures, *focus-out events* and long response times, the authors showed that 62 percent of respondents multitasked at least once.

Although distractions during survey completion might occur across all device types, respondents taking web surveys on a mobile device, especially on a smartphone, might be more likely to get distracted or to multitask than respondents using a stationary PC or laptop. The portability of mobile devices allows users to complete the survey in environments where other people are present and where they might encounter a greater variety of distractions than at home, for example while using public transport or being in public spaces. Extant research confirms this expectation: although the majority of mobile web respondents still complete questionnaires at home or at work, they are significantly more likely than PC or laptop users to complete the survey away from home or work, in settings where other people are present (Antoun et al., 2017; Mavletova & Couper, 2013; Revilla, Toninelli, et al., 2016; Toepoel & Lugtig, 2014). Similarly, previous research suggests that respondents are more likely to multitask when using a mobile device rather than a PC for survey completion (Ansolabehere & Schaffner, 2015; Antoun et al., 2017).

Several studies have examined how distractions and multitasking affect response behaviour in web and telephone surveys. The first set of studies asked respondents to self-report if they were doing other activities during the interview, and compared the response quality of multitasking respondents with those who did not report any multitasking activity (Ansolabehere & Schaffner, 2015; Antoun et al., 2017; Lavrakas, Tompson, & Benford, 2010; Schober et al., 2015). Results suggest no significant effect of reported multitasking on the majority of quality indicators, including item-nonresponse, straight-lining in grid questions, rounding, and the length of responses to open-ended questions (Antoun et al., 2017; Lavrakas et al., 2010; Schober et al., 2015). The studies, however, found significantly longer survey completion times among multitasking respondents, and mixed evidence for the association between multitasking and the reliability of answers to attitudinal questions (Ansolabehere & Schaffner, 2015; Lavrakas et al., 2010). The second set of studies used data from mode experiments in which respondents were randomly allocated to participate in the survey via mobile phone or landline phone (Kennedy & Everett, 2011; Lynn & Kaminska, 2012). The authors compared the response quality between mobile phone and landline phone respondents, assuming that mobile phone respondents are more likely to be exposed to distractions or to do other activities while completing the survey. Although mobile phone respondents were shown to have a significantly longer interview length than those responding on their landline phone, no significant differences were found with regard to other quality indicators, including item-nonresponse, straight-lining in grid questions, rounding, the length of responses to open-ended questions, acquiescence, extreme response styles, and recency effects (Kennedy & Everett, 2011; Lynn & Kaminska, 2012). Finally, the study by Sendelbah et al. (2016), using the paradata-based approach described earlier, found no significant relationship between multitasking and straight-lining in grid questions, and a significant but weak effect of the number of *focus-out events* on item-nonresponse.

Although previous research suggests that distractions and multitasking do not have a significant effect on response behaviour and data quality, the previous studies have two major limitations. First, none of the studies experimentally manipulate the presence of distractions or multitasking activities and might be affected by selection bias if the research design does not sufficiently account for confounding factors. A possible explanation for the observed null findings might be that only those respondents choose to multitask who have high levels of cognitive capacity and whose performance is not compromised by multitasking. Second, with the exception of Sendelbah et al. (2016) who measure distractions using paradata, previous studies rely on self-reported measures of distractions that are susceptible to recall error and social desirability bias (Ansolabehere & Schaffner, 2015). Respondents might over- or underestimate the amount of time spent doing other activities, either because they cannot remember all instances of distractions or because they are reluctant to report distractions to the survey researcher.

In this paper, I report on results from a laboratory experiment that examines how distractions during web survey completion influence data quality. N = 261 participants were randomly assigned to experimental groups using a 2 (device type) x 3 (form of distraction) between-subject factorial design. They were asked to complete a web questionnaire on either a PC or a tablet, and were allocated to one of three distraction conditions: a) the presence of other people in the room who have a loud conversation, b) the presence of music, or c) no distraction. I examine the following research questions:

(1) Are respondents distracted by the presence of other people and by the presence of music?
(2) Do distracted respondents provide survey data of lower quality?
(3) Do different forms of distractions have differential effects on data quality?
(4) Does the effect of distractions on data quality vary by device type?

2.2. Background

Respondents might be involved in various forms of distractions during survey completion. Conceptually, we can make a distinction between distractions that originate from the environment over which the respondent has less control, for instance background noise, and multitasking activities that are initiated by the respondent, such as browsing other websites while completing a survey (Ansolabehere & Schaffner, 2015; Sendelbah et al., 2016; Zwarun & Hall, 2014). Multitasking activities can be further classified by the extent that they interfere with the primary activity of survey completion (Salvucci & Taatgen, 2011; Sendelbah et al., 2016): while some activities can be carried out in parallel to survey completion, such as listening to music (*concurrent* multitasking), other activities require to switch tasks between survey completion and the secondary activity, for example having a conversation with another person (*sequential* multitasking). In the presence of distractions and multitasking activities, respondents have to divide their attention between the survey task and the distraction or multitasking activity, either continuously or intermittently depending on the type of distraction. This division of attention might have implications for information processing, including the cognitive processing of a survey response (Tourangeau, Rips, & Rasinski, 2000). Distractions and multitasking might have similar implications because in both circumstances, respondents have to react to external stimuli (Kennedy, 2010).

Capacity theories of attention suggest that human attention is limited: individuals are able to process multiple tasks simultaneously depending on how much capacity the tasks demand and which type of cognitive resources they require (Kahneman, 1973; Kellogg, 2012). Task performance suffers if two tasks are similar and require the same set of cognitive resources, for example posting content to social media websites and completing a web survey, whereas two distinct tasks can be processed without loss in performance, such as eating and completing a web survey (Kellogg, 2012; Salvucci & Taatgen, 2011; Trafton & Monk, 2007).

If respondents get distracted and the distraction or multitasking activity draws on similar cognitive resources as the survey task, respondents might have insufficient cognitive capacity to accurately carry out the response process. Following the cognitive response process model (Tourangeau, 1984; Tourangeau et al., 2000), respondents who get distracted or multitask might perform the four stages of the response process less accurately (Kennedy, 2010; Lynn & Kaminska, 2012). Distracted respondents might not be able to pay sufficient attention to the question wording and might not fully comprehend a question. They might not have the cognitive capacity to retrieve all relevant information from memory or might not be able to accurately integrate the retrieved information to make a judgement. Finally, distracted respondents might fail to map their response to the available set of response options if they do not pay sufficient attention to the format and wording of the response options. As a result of more superficial response processing, these respondents might provide data of lower quality compared to those who are able to fully concentrate on the survey task. We might expect that distracted respondents show a similar response behaviour to respondents who use satisficing response strategies to cope with the cognitive demands of survey questions: in both circumstances, respondents might perform the four stages of the response process less thoroughly (Antoun, 2015; Krosnick, 1991; Krosnick & Alwin, 1987; Lynn & Kaminska, 2012). The presence of distractions or multitasking might also affect survey completion time: as cognitive capacity is limited, respondents might process multiple tasks sequentially rather than in parallel, by switching back and forth between the primary and the secondary task (Pashler, 1994). As a result of these additional switching processes, distracted or multitasking respondents might need more time for survey completion (Meyer & Kieras, 1997).

Various factors might moderate the extent to which distractions affect cognitive response processing and data quality. First, we can expect that the type of the distraction or multitasking activity has an influence. Distractions or activities that are more similar to the survey task itself might interfere with cognitive response processing to a larger extent than activities that are very different and draw on a different set of cognitive resources (Kellogg, 2012; Trafton & Monk, 2007). Second, the task difficulty of the survey might moderate the effect of distractions on data quality as more difficult tasks put greater demands on the respondent's cognitive capacity (Kahneman, 1973). Response processing of questions that contain rarely used words, vague or ambiguous terms or a complex syntax, or that ask respondents to perform complex retrieval strategies might be more affected by the presence of distractions than questions that are relatively easy to answer (Krosnick, 1991; Lenzner, Kaczmirek, & Lenzner, 2010). Finally, respondents might vary the extent to which they are able to cope with distractions (Krosnick, 1991). Those with higher levels of cognitive capacity might be less affected by the additional cognitive load of distractions (Kahneman, 1973; Kellogg, 2012). Respondents who get distracted but are highly motivated to complete the survey might also be less affected by the presence of distractions: despite being exposed to distractions, motivated respondents might try to focus on the survey task and to put more cognitive effort into survey responding, which might reduce the loss in cognitive performance that result from distractions (Krosnick, 1991). Similarly, the cognitive processing of people who are familiar with the task of completing surveys as well as those who are used to working in distracting environments might be less affected by distractions as they might have developed strategies to process multiple tasks efficiently (Sendelbah et al., 2016).

This study sets out to examine the impact of two forms of distraction that are likely to occur during web survey completion and that can reasonably be reconstructed in a laboratory setting: the presence of other people who have a loud conversation[2] and the presence of music. The two forms of distraction might affect data quality to a different extent: listening to music is only an auditory distraction and might have a smaller effect on data quality compared to the presence of other people in the room who are speaking loudly with each other, which might distract respondents both in auditory and visual ways. In the music condition, the type of music might play an important role in whether it is disruptive to cognitive performance. Previous research found that vocal music is significantly more disruptive to performance in recall tasks than instrumental music or silence (Belsham & Harman, 1977; Salamé & Baddeley, 1989), presumably as the information load of music might be greater if it contains vocals (Furnham, Trew, & Sneade, 1999; Kiger, 1989).

Given that respondents increasingly use their mobile device for survey completion, I was also interested to test whether the effect of distractions on data quality is consistent across device types. Extant research shows that there are no large differences in measurement error between respondents using a PC and those using a mobile device for survey completion (Couper et al., 2017), which suggests that cognitive response processing is similar across devices. I therefore do not expect that distractions have a differential effect on quality depending on device type.

[2] While this paper focuses on the potential distraction that people create when having a loud conversation, the presence of other people during survey completion might also affect responding to sensitive questions (e.g., Couper, Singer, & Tourangeau, 2003).

2.3. Data and Methods

Experimental design

Participants were randomly assigned to experimental groups using a 2 (device type) x 3 (form of distraction) between-subject factorial design. They were asked to complete a web questionnaire of around 20 minutes on either a desktop PC or a tablet[3] (iPad 4, 9.7-inch screen), and were allocated to one of three distraction conditions: a) the presence of other people in the room who have a loud conversation, b) the presence of music, or c) no distraction. In each experimental session, all participants were exposed to the same type of distraction and completed the survey on the same type of device. The survey was programmed in *Qualtrics*, optimised for both PC and tablet, and had a median length of 15 minutes.

The three distraction conditions were operationalised as follows. In the ***people*** condition, four confederates were sitting at a table in the middle of the room and were instructed to play a word-guessing game while the participants were filling out the online survey. The confederates were briefed about the purpose of the experiment and were instructed to have a loud conversation and to make noise. The partition walls in the laboratory were removed, so that participants were able to see the four confederates and the other participants in the room. The volume of the conversation might have varied across subjects depending on the location of their workspace, but I expect that differences were minimal due to the relatively small size of the laboratory. Participants allocated to the ***music*** condition listened to music

[3] It would have also been interesting to test the effect of distractions on data quality on smartphones. The laboratory where I conducted the experiment, however, was only able to provide tablets to participants; I would have had to rely on the participants' own device if I had wanted to test on smartphones. In order to control for device type by keeping the device constant across participants, I decided to only rely on the devices provided by the laboratory. Future research might investigate whether the findings of this paper also hold true for smartphones.

taken from a mid-morning radio programme from BBC Radio 1 while filling out the questionnaire. The music consisted of upbeat pop songs with vocals that were separated by a female radio presenter talking. Upbeat pop music with vocals has been shown to affect performance in recall tasks and reading comprehension (Furnham & Bradley, 1997). The participants listened to the music via headphones, and the volume was set to the same level for all participants. In the ***no distraction*** condition, participants completed the survey in the default laboratory setup and were not exposed to any form of distraction.

While laboratory experiments have limited external validity due to sample characteristics and the artificial nature of the laboratory setting (Jerit, Barabas, & Clifford, 2013), this study design has a number of advantages. The laboratory makes it possible to experimentally manipulate the type and magnitude of distractions, and to control for other factors that potentially affect measurement including device characteristics, such as screen size and the speed and quality of the Internet connection. By using a laboratory study, it can also be ensured that all respondents complete the survey in the mode which they were allocated to, which is often difficult to realise in the field (Millar & Dillman, 2012).

Measures of distraction

To check whether the distraction manipulation worked as intended and the respondents in the two distraction conditions felt more distracted, I included two debriefing questions about the perceived level of distraction at the end of the questionnaire. Participants were asked: “How distracted did you feel by the things going on around you while completing the survey?” (1 = Not distracted at all; 2; 3; 4; 5 = Extremely distracted), and “Overall, how much attention were you able to pay to the survey?” (1 = A lot of attention; 2; 3; 4; 5 = No attention at all).

Measures of data quality

I use ten indicators to compare data quality between the experimental treatment groups. This section describes how the indicators were operationalised; the number in parentheses index the questions in Appendix Table 2.5 that were used to create the quality indicators. Questions were adopted from the European Social Survey (Round 3 questionnaire modules on political attitudes and personal well-being) and from the LISS Panel (experiment: "Non-response and measurement in mobile web surveys"), among other sources.

Non-differentiation. Non-differentiation is a response pattern where respondents select the same or a similar response option for all items in a grid question, which might serve as a cognitive shortcut (Krosnick, 1991). To measure non-differentiation, I calculated the average standard deviation of responses to five grid questions with reverse-coded items (Q5a-h, Q6a-e, Q8a-j; Q13a-i; Q14a-i); the first grid contained eight items, the second grid five items, the third grid ten items, and the fourth and fifth grid nine items each. A lower standard deviation reflects a less differentiated response pattern and hence lower data quality. The standard deviation of responses is used as a measure of non-differentiation as it is more fine-grained than binary measures of whether respondents gave the same answer to all items in a grid question; the variance of responses as a measure of non-differentiation has been used in previous studies (Barge & Gehlbach, 2012; Kennedy & Everett, 2011; Lavrakas et al., 2010).

Agreeing. A greater tendency to agree in agree-disagree items might be a potential strategy to reduce cognitive effort (Krosnick, 1991; Krosnick, Narayan, & Smith, 1996; Lynn & Kaminska, 2012). I calculated the proportion of 'agree' responses (e.g., strongly agree,

agree) among 32 agree-disagree items (Q8a-j, Q13a-i, Q14a-i, Q20a-d). This indicator has previously been used by Lynn and Kaminska (2012).

Extreme and middle response. As a cognitive shortcut, respondents might select an extreme-point or middle-point answer rather than a more differentiated answer in scale questions (Kaminska, McCutcheon, & Billiet, 2010; Krosnick et al., 1996). I computed the proportion of extreme-point or middle-point answers (0, 5, 10) among eleven items with an eleven-point scale (Q7, Q21a-g, Q22, Q23, Q24); a higher proportion reflects lower data quality. The indicator has previously been used by Lynn and Kaminska (2012) and Kaminska, McCutcheon and Billiet (2010).

Don't know response. To reduce cognitive effort, respondents might choose the 'don't know' option rather than selecting a substantive answer category if they are explicitly provided with a 'don't know' option (Krosnick, 1991). For each respondent, I calculated the proportion of 'don't know' responses among 21 items with explicit 'don't know' option (Q5a-h, Q6a-e, Q7, Q20a-d, Q22, Q23, Q24); a higher proportion reflects lower data quality. This indicator has previously been used by Lynn and Kaminska (2012).

Inconsistent response. I included two logical question pairs to measure whether respondents provide an inconsistent response (Q16, Q17, Q18, Q19). This indicator has previously been used by Kaminska, McCutcheon, and Billiet (2010). Respondents were asked the following questions:

(1) "In your opinion, what is the ideal age for a [girl or woman/boy or man] to get married and live with [her/his] spouse?"

(2) “Before what age would you say a [woman/man] is generally too young to get married and live with [her/his] spouse?”

The response is inconsistent if the reported age in the second question (“too young”) is higher than the reported age in the first question (“ideal age”). I created a binary measure coded as 1 if the respondent provided an inconsistent response to at least one of the two logical question pairs, and 0 otherwise.

Length of open-ended response. Respondents might try to avoid typing responses to open-ended questions to reduce effort; shorter responses might therefore reflect larger measurement error (Lugtig & Toepoel, 2015). Response length as an indicator of data quality has been used in previous studies (Antoun et al., 2017; Kennedy & Everett, 2011; Lugtig & Toepoel, 2015; Mavletova, 2013). I asked respondents two open-ended questions (Q4, Q36):

(1) “Do you have any hobbies? If so, what are these? If you do not have any hobbies, please leave this question blank.”

(2) “Why have you chosen to study [subject] at the University of Essex?”

I added up the number of characters that respondents provided to the two questions.

Avoiding half-open ‘other’ response. I included two half-open questions to measure whether respondents select one of the closed-ended options, even if the response might be less plausible, rather than selecting the ‘other’ option that requires typing a response (Q2, Q3). This indicator of data quality has previously been used in the web survey literature (Antoun et al., 2017; Peytchev & Hill, 2010; Wells et al., 2014). Respondents were asked:

(1) “What is your favourite vegetable?”

(1) Green beans

(2) Broccoli

(3) Kale

(4) Carrots

(5) Spinach

(6) Other, that is [textbox]

(2) “What is the main reason you eat vegetables?”

(1) They are in my food already

(2) I like colours in my meals

(3) Other, please specify [textbox]

I use a binary measure coded as 1 if the respondent selects a closed-ended response option in both questions, and as 0 if they select the half-open ‘other’ option and provide an open-ended response to at least one of the questions.

Items selected in check-all-that-apply questions. For each respondent, I counted the number of items selected in one check-all-that-apply question with twelve items (Q1). The number of items selected in check-all-that-apply questions has previously been used by Lugtig and Toepoel (2015) as a measure of data quality; a larger number of selected items might indicate that respondents put more cognitive effort into survey responding. The question was adopted from Mavletova (2013) and asked respondents:

“Which of the following activities were you doing for the past 12 months to feel good and healthy? Please check all that apply.”

(1) I tried not to overeat

(2) I exercised

(3) I tried to relax, avoid stress

(4) I went to the gym or swimming pool

(5) I took a contrast shower

(6) I went to a health resort

(7) I tried to eat healthy food

(8) I took vitamins

(9) I saw a doctor

(10) I consumed alternative medicine products

(11) I bought food with low levels of cholesterol, fat, calories, or artificial food additives

(12) I walked outdoors

I also included two data quality indicators that measure more directly how much cognitive effort respondents made during survey completion and how much attention they paid to the questions: the cognitive reflection test and an instructional manipulation check.

Cognitive reflection test. I asked respondents to complete the cognitive reflection test developed by Frederick (2005). As a measure of survey data quality, it has first been used by Antoun, Couper, and Conrad (2017). The test consists of three items (Q27, Q28, Q29) which are designed to generate an intuitive but incorrect answer; to get the correct answer, respondents have to think deliberately. It therefore measures whether respondents process their responses deliberately or rely on cognitive shortcuts (Antoun et al., 2017). For example, the first item asks: “A bat and a ball cost £1.10. The bat costs £1.00 more than the ball. How much does the ball cost?”. The impulsive answer that might first come to one’s mind is “10 pence” but the correct response is “5 pence”. For each respondent, I counted the number of incorrect responses, ranging from 0 to 3; a higher score reflects lower data quality.

Instructional manipulation check. I also implemented an instructional manipulation check developed by Berinsky, Margolis, and Sances (2014) (Q15). The manipulation check measures if respondents read instructions diligently or just skim through the question (Berinsky et al., 2014; Jones, House, & Gao, 2015; Oppenheimer, Meyvis, & Davidenko, 2009). The first sentence of the instructional manipulation check tells respondents: “Before we proceed, we have a question about how you are feeling.” The first sentence is followed by a long paragraph with instructions. If respondents continue to read the instructions, they will notice the sentence “To show that you have read the instructions, please ignore the question below about how you are felling and instead check only the ‘none of the above’ option as your answer.” If they skip reading the instructions, they might proceed directly to the response options and might report how they are feeling. I use a binary indicator that takes on the value of 1 if the respondent selects a substantive answer about how they are feeling and fails the manipulation check, and the value of 0 if they select the ‘none of the above’ option and pass the manipulation check.

Survey duration. In addition to examining data quality, I was interested to test if the presence of distractions affects survey completion time. I computed survey duration in minutes by adding up the question-level response times, which measure the time between page load until the respondent clicks the ‘Next’ button, across all questions.

Participants

Students aged 18-25 from the University of Essex who signed up for the ESSEXLab participant database were invited to take part in the experiment. They were paid £6 for participating in the 30-minute session. The achieved sample included N = 261 participants who were distributed across the six cells of the experimental design (Table 2.1). Subjects were on average 21 years old; 66 percent were female; 59 percent were White, 21 percent Asian, and 15 percent Black. The large majority of participants (87 percent) were enrolled in undergraduate degrees, and 45 percent of subjects were native English speakers.

Table 2.1. Experimental design and sample size.

	People	Music	Control	*Total*
PC	46	40	42	*128*
Tablet	44	46	43	*133*
Total	*90*	*86*	*85*	*261*

As a result of differential no-shows to experimental conditions, some socio-demographic groups might be over- or under-represented in particular treatment groups. To check if there is any difference in the sample composition across treatment groups, I ran a multinomial logistic regression with treatment group as dependent variable and age, gender, ethnicity, level of study, and English proficiency as covariates. None of these covariates were significant, which suggests that treatment groups are balanced with regard to the set of socio-demographic characteristics that I examined.

2.4. Results

Before studying the impact on data quality, I want to examine whether participants felt distracted by other people in the room who had a loud conversation, or when listening to music. Responses to the two debriefing questions about perceived levels of distraction and attention suggest that participants indeed felt more distracted in the two distraction conditions than in the control condition (Table 2.2).

Whereas none of the participants in the *Tablet-Control* condition and only 4.8 percent of participants in the *PC-Control* condition said that they were very or extremely distracted by the things going on around them while completing the survey, the perceived level of distraction was higher in the two distraction conditions. For example, 27.5 percent of participants in the *PC-Music* condition and 36.4 percent in the *Tablet-People* condition reported being very or extremely distracted.

Similarly, 90.5 percent of participants in the *PC-Control* condition and 95.4 percent in the *Tablet-Control* condition said that they were able to pay a lot of attention to the survey while the self-reported level of attention was lower in the other conditions: for example, 67.4 percent of participants in the *Tablet-Music* condition and 67.4 percent in the *PC-People* condition said they could pay a lot of attention to the survey.

Table 2.2. Perceived level of distraction and attention by experimental condition (in percent).

	PC			**Tablet**		
	Control	Music	People	Control	Music	People
Extremely distracted (4-5)	4.76	27.50	26.09	0.00	23.91	36.36
A lot of attention (1-2)	90.48	70.00	67.39	95.35	67.39	70.45

N = 260.

A two-way ANOVA with distraction (control, music, people) and device (PC, tablet) as between-subject factors shows a significant main effect of distraction both for the distraction question, F (2, 254) = 38.62, $p < 0.001$, and the attention question, F (2, 254) = 10.65, $p < 0.001$. Tukey-Kramer post-hoc tests reveal that respondents in the *music* and *people* conditions felt significantly more distracted and were less able to pay attention to the survey compared to the control condition, while the average ratings of the *music* and *people* conditions were not significantly different from each other. The ANOVA shows no significant main effect of device type and no significant interaction effect of distraction and device type for either of the two debriefing questions.

These findings suggest that respondents in the two distraction conditions indeed felt more distracted, which gives us some confidence that the distraction treatments worked as intended. Interestingly, respondents did not feel differently in the *music* and *people* conditions: they felt more distracted, independent of whether the source of distraction were other people having a loud conversation or music. The perceived level of distraction was also not affected by the type of device that participants were using.

I next examine the effect of distractions on data quality. I first want to test whether distracted respondents, independent of the distraction type and the type of device they are using, provide data of lower quality. Table 2.3 shows the quality indicators in the control condition (with devices combined) compared to the distraction conditions (with types of distraction and devices combined).

Results suggest no statistically significant difference between distraction and control condition for any of the data quality indicators. For example, I find no significant difference in non-differentiation, extreme response styles, and in rates of selecting an ‘agree’ or ‘don’t know’ response, which replicates findings from previous observational studies (Antoun et al., 2017; Kennedy & Everett, 2011; Lavrakas et al., 2010; Lynn & Kaminska, 2012; Schober et al., 2015; Sendelbah et al., 2016). For some indicators, although not statistically significant, the effect points in the expected direction: distracted respondents type, on average, slightly shorter responses to open-ended questions than respondents in the control condition (142 vs. 158 characters). They are also more likely to select a closed-ended response option in 'half-open' questions than respondents who are not distracted (30 percent vs. 26 percent selected a closed-ended response). For the more direct measures of attention and cognitive effort, the difference is also as expected: distracted respondents provide more incorrect responses to the cognitive reflection test (2.35 vs. 2.11 number of incorrect items) and are more likely to fail the instructional manipulation check than respondents in the control condition (40 percent vs. 21 percent failed the manipulation check). Surprisingly, I do not find a significant difference in completion time between control and distraction conditions.

Table 2.3. Data quality indicators in control and distraction conditions, devices and distraction type combined.

	Control (n = 85)	**Distraction** (n = 176)	**Test statistic**
Mean standard deviation to grid questions	1.29 (0.03)	1.29 (0.03)	t (259) = 0.149, p = 0.882
Proportion 'agree' response	0.45 (0.01)	0.45 (0.01)	t (259) = -0.265, p = 0.791
Proportion extreme and middle response	0.24 (0.02)	0.26 (0.02)	t (259) = -0.651, p = 0.516
Proportion 'don't know' response	0.03 (0.005)	0.02 (0.003)	t (259) = 0.939, p = 0.349
Proportion inconsistent response	0.01 (0.01)	0.00 (--)	χ^2 (1) = 2.079, p = 0.149
Mean length open response (in characters)	157.99 (17.87)	141.51 (7.37)	t (259) = 1.010, p = 0.314
Proportion closed-ended response selected in both 'half-open' questions	0.26 (0.05)	0.30 (0.03)	χ^2 (1) = 0.379, p = 0.538

Mean number of items selected in check-all-that-apply question	5.68 (0.20)	5.23 (0.17)	t (259) = 1.606, p = 0.109
Mean number of incorrect responses to cognitive reflection test	2.11 (0.12)	2.35 (0.07)	t (259) = -1.913, p = 0.057
Proportion failed instructional manipulation check	0.31 (0.05)	0.40 (0.04)	χ^2 (1) = 2.335, p = 0.127
Mean survey duration (in minutes)	15.94 (0.47)	15.57 (0.32)	t (258) = 0.646, p = 0.519

Note. Standard errors in parentheses.

As the next step, I want to test whether the two different forms of distraction have a differential effect on data quality, and whether the effect varies by device type. Table 2.4 shows the quality indicators by experimental conditions. Similar to the pooled analysis, results suggest that data quality is on a similar level across the two types of distraction, and also across devices. To test for statistical significance, I ran linear regression models for continuous indicators and logistic regression models for binary indicators, with distraction (control, music, people), device type (PC, tablet) and the interaction of distraction and device type added as covariates. Results suggest no significant effect of the *music* treatment and the *people* treatment for any of the data quality indicators (analysis not shown). I also do not find any significant interactions effects of distraction and device type, which suggests that the findings are robust across devices.

Table 2.4. Data quality indicators by experimental condition.

	PC			Table		
	Control (n = 42)	Music (n = 40)	People (n = 46)	Control (n = 43)	Music (n = 46)	People (n = 44)
Mean standard deviation to grid questions	1.24 (0.04)	1.28 (0.05)	1.28 (0.04)	1.35 (0.04)	1.34 (0.03)	1.25 (0.05)
Proportion 'agree' response	0.44 (0.02)	0.46 (0.02)	0.43 (0.01)	0.46 (0.02)	0.45 (0.01)	0.47 (0.01)
Proportion extreme and middle response	0.24 (0.03)	0.25 (0.03)	0.25 (0.03)	0.24 (0.03)	0.22 (0.02)	0.32 (0.04)
Proportion 'don't know' response	0.02 (0.01)	0.02 (0.01)	0.02 (0.01)	0.03 (0.01)	0.02 (0.01)	0.03 (0.01)
Proportion inconsistent response	0.02 (0.02)	0.00 (--)	0.00 (--)	0.00 (--)	0.00 (--)	0.00 (--)
Mean length open response (in characters)	199.93 (33.42)	153.23 (15.81)	154.93 (17.85)	117.02 (10.82)	130.07 (11.87)	128.80 (12.66)
Proportion closed-ended response selected in both 'half-open' questions	0.29 (0.07)	0.38 (0.08)	0.20 (0.06)	0.23 (0.07)	0.33 (0.07)	0.30 (0.07)
Mean number of items selected in check-all-that-apply question	5.64 (0.33)	5.05 (0.40)	4.91 (0.31)	5.72 (0.25)	5.59 (0.32)	5.34 (0.34)
Mean number of incorrect responses to cognitive reflection test	2.17 (0.15)	2.33 (0.16)	2.24 (0.14)	2.05 (0.18)	2.37 (0.14)	2.48 (0.11)
Proportion failed instructional manipulation check	0.26 (0.07)	0.35 (0.08)	0.37 (0.07)	0.35 (0.07)	0.39 (0.07)	0.50 (0.08)
Mean survey duration (in minutes)	15.86 (0.68)	16.49 (0.76)	15.27 (0.59)	16.01 (0.67)	14.13 (0.45)	16.57 (0.73)

Note. Standard errors in parentheses.

I conducted a series of moderator analyses to examine the interaction of distractions and potentially moderating factors: language ability (whether English is the respondent's first language); personality, measured by a ten-item Big-Five personality scale (Gosling, Rentfrow, & Swann, 2003); motivation, with need for cognition used as a proxy (Cacioppo, Petty, & Morris, 1983), and individual susceptibility to distractions (Schepers, 2007). However, I do not find significant effects for any of the interactions (analysis not shown).

2.5. Discussion

To the best of my knowledge, this paper has been the first study to examine the impact of distractions on survey data quality using a laboratory experimental setup. The experiment aims to help survey researchers to get a better understanding of how the physical environment in which respondents complete a web survey might affect response behaviour and data quality. I studied two forms of distraction that are likely to occur in web survey settings, the presence of other people who have a loud conversation and the presence of music. Respondents in the two distraction conditions felt significantly more distracted and were less able to pay attention to the survey. However, I did not find any significant effect of distractions on data quality: respondents who get distracted by other people or by music provide survey data of similar quality as respondents who do not experience any distraction. The findings are encouraging for survey practitioners who administer web surveys: even if respondents choose to listen to music while completing a survey, or are in a noisy environment where other people are present, these forms of distraction do not seem to affect the quality of responses they provide, independent of whether they are using a PC or tablet for survey completion.

The study has a number of limitations. First, I only examined the effect of two forms of distraction that could reasonably be reconstructed in a laboratory setting, the presence of other people in the room and music. Other forms of distractions and multitasking activities that are very similar to the survey task itself, such as browsing websites or using social media, might have a larger effect on cognitive response processing. Most of these activities, however, are difficult to simulate in a laboratory setting because they rely on the respondent to initiate the additional task and cannot be externally manipulated. Second, the study is based on a homogeneous sample of university students aged 18-25 and might not be generalisable to the general population. Young respondents might have more cognitive capacity to cope with distractions, and might be used to work and study in the presence of distractions. For these reasons, we might expect to find a larger effect of distractions on survey data quality if the present study was replicated among the general population or among older respondents. Third, the statistical power for some of the analyses might have been low due to a small sample size, which might have contributed to some of the null findings. However, I was able to detect significant differences for the manipulation check.

The following avenues of further research might be worth pursuing. First, to reduce the problem of external validity due to the artificial nature of the laboratory, future research might use field experiments to replicate the present study in more natural settings. Respondents could be provided with the same devices, still controlling for device characteristics that potentially affect measurement, and asked to complete a questionnaire in different settings, for example in a quiet library compared to a busy street. Second, it might be interesting to use eye-tracking technology to get a better understanding of how distractions and multitasking activities affect the cognitive response process (e.g., Galesic, Tourangeau, Couper, & Conrad, 2008), by studying eye-movements rather than indirect measures of data quality derived from the questionnaire. Finally, it might be interesting to expand upon the existing observational studies that rely on self-reported measures of distraction by using mobile technologies: mobile devices allow capturing distractions passively, by relying on the integrated microphone or on passive data collection apps. Examples include recording noise or capturing whether the respondent switches to other apps while filling out the questionnaire, similar to the paradata approach by Sendelbah et al. (2016), which might also help to better understand the impact of distractions on data quality.

2.6. Appendix

Table 2.5. Questionnaire.

	Question wording
Q1	Which of the following activities were you doing for the past 12 months to feel good and healthy? Please check all that apply. *I tried not to overeat; I exercised; I tried to relax, avoid stress; I went to the gym or swimming pool; I took a contrast shower; I went to a health resort; I tried to eat healthy food; I took vitamins; I saw a doctor; I consumed alternative medicine products; I bought food with low levels of cholesterol, fat, calories, or artificial food additives; I walked outdoors*
Q2	What is your favourite vegetable? *Green beans; Broccoli; Kale; Carrots; Spinach; Other, that is [Textbox]*
Q3	What is the main reason you eat vegetables? *They are in my food already; I like colours in my meals; Other, please specify below [Textbox]*
Q4	Do you have any hobbies? If so, what are these? If you do not have any hobbies, please leave this question blank. *[Textbox]*
Q5	In this question, you can see a list of the ways you might have felt or behaved during the past week. Please indicate how much of the time during the past week *None or almost none of the time; Some of the time; Most of the time; All or almost all of the time; Don't know* … you felt depressed? … you felt that everything you did was an effort? … your sleep was restless? … you were happy? … you felt lonely? … you enjoyed life? … you felt sad? … you could not get going?
Q6	Please indicate to what extent *0 = Not at all; 1; 2; 3; 4; 5; 6 = A great deal; Don't know* … you get a chance to learn new things?

	… you feel that people in your local area help one another? … you feel that people treat you with respect? … you feel that people treat you unfairly? … you feel that you get the recognition you deserve for what you do?
Q7	Taking all things together, how happy would you say you are? *0 = Extremely unhappy; 1; 2; 3; 4; 5; 6; 7; 8; 9; 10 = Extremely happy; Don't know*
Q8	Here are a number of personality traits that may or may not apply to you. To what extent do you agree or disagree with each of the following statements? Please rate the extent to which the pair of traits applies to you, even if one characteristic applies more strongly than the other. I see myself as … *Disagree strongly; Disagree moderately: Disagree at little; Neither agree nor disagree; Agree a little; Agree moderately; Agree strongly* Extraverted, enthusiastic Critical, quarrelsome Dependable, self-disciplined Anxious, easily upset Open to new experiences, complex Reserved, quiet Sympathetic, warm Disorganised, careless Calm, emotionally stable Conventional, uncreative
Q9	How well can you concentrate if something in the background is distracting you? *1 = Extremely well; 2; 3; 4; 5 = Extremely badly*
Q10	How quiet must the environment be in order for you to study effectively? *1 = Extremely quiet; 2; 3; 4; 5 = Not quiet at all*
Q11	Please indicate to what extent the following statements apply to you. To what extent… *1 = A great deal; 2; 3; 4; 5 = Not at all* … will music in your immediate environment disturb your concentration? … will noise in your immediate environment disturb your concentration?

	… will chatter in your immediate environment disturb your concentration? … are you dependent on absolute silence if you want to concentrate? … will people moving around in your immediate environment distract your attention? … would it distract you if other people in the room spoke softly while you are studying? … would the noise of a TV in an adjacent room distract you from your studies?
Q12	How easy or difficult do you find it adapting to loud music whilst trying to solve a problem? *1 = Extremely easy; 2; 3; 4; 5; 6; 7 = Extremely difficult*
Q13	For each of the following statements please indicate to what extent they apply to you. Please note that there are no right or wrong answers. *Strongly disagree; Disagree; Neither agree nor disagree; Agree; Strongly agree* I would prefer complex to simple problems I like to have the responsibility of handling a situation that requires a lot of thinking Thinking is not my idea of fun I would rather do something that requires little thought than something that is sure to challenge my thinking abilities I try to anticipate and avoid situations where there is likely chance that I will have to think in depth about something I find satisfaction in deliberating hard and for long hours I only think as hard as I have to I prefer to think about small, daily projects to long-term ones I like tasks that require little thought once I've learned them
Q14	For each of the following statements please indicate to what extent they apply to you. Please note that there are no right or wrong answers. *Strongly disagree; Disagree; Neither agree nor disagree; Agree; Strongly agree* The idea of relying on thought to make my way to the top appeals to me I really enjoy a task that involves coming up with new solutions to problems

	Learning new ways to think doesn't excite me very much I prefer my life to be filled with puzzles that I must solve The notion of thinking abstractly appeals to me I would prefer a task that is intellectual, difficult and important to one that is somewhat important but does not require much thought I feel relief rather than satisfaction after completing a task that required a lot of mental effort It's enough for me that something gets the job done; I don't care how or why it works I usually end up deliberating issues even when they do not affect me personally
Q15	Before we proceed, we have a question about how you are feeling. Recent research on decision making shows that choices are affected by context. Differences in how people feel, their previous knowledge and experience, and their environment can affect choices. To help us understand how people make decisions, we are interested in information about you. Specifically, we are interested in whether you actually take the time to read the directions; if not, some results may not tell us very much about decision making in the real world. To show that you have read the instructions, please ignore the question below about how you are feeling and instead check only the 'none of the above' option as your answer. Thank you very much. Please check all words that describe how you are currently feeling. *Interested; Distressed; Excited; Upset; Strong; Guilty; Scared; Hostile; Enthusiastic; Proud; Irritable; Alert; Ashamed; Inspired; Nervous; Determined; Attentive; Jittery; Active; Afraid; None of the above*
Q16	In your opinion, what is the ideal age for a girl or woman to get married and live with her spouse? *[Textbox]*
Q17	In your opinion, what is the ideal age for a boy or man to get married and live with his spouse? *[Textbox]*

Q18	Before what age would you say a woman is generally too young to get married and live with her spouse? *[Textbox]*
Q19	Before what age would you say a man is generally too young to get married and live with his spouse? *[Textbox]*
Q20	The following statements are about men and women and their place in the family. How much do you agree or disagree with each of them? *Strongly agree; Agree; Neither agree nor disagree; Disagree; Strongly disagree; Don't know* A woman should be prepared to cut down on her paid work for the sake of her family. Men should take as much responsibility as women for the home and children. When jobs are scarce, men should have more right to do a job than women. When there are children in the home, parents should stay together even if they don't get along.
Q21	How much do you personally trust each of the following institutions? *0 = No trust at all; 1; 2; 3; 4; 5; 6; 7; 8; 9; 10 = Complete trust* The UK Parliament The legal system The police Politicians Political parties The European Parliament The United Nations
Q22	On the whole how satisfied are you with the present state of the economy in the UK? *0 = Extremely dissatisfied; 1; 2; 3; 4; 5; 6; 7; 8; 9; 10 = Extremely satisfied; Don't know*
Q23	Now thinking about the UK government, how satisfied are you with the way it is doing its job? *0 = Extremely dissatisfied; 1; 2; 3; 4; 5; 6; 7; 8; 9; 10 = Extremely satisfied; Don't know*

Q24	And what do you think overall about the state of health services in the UK nowadays? *0 = Extremely bad; 1; 2; 3; 4; 5; 6; 7; 8; 9; 10 = Extremely good; Don't know*
Q25	Do you think that US President-elect Donald Trump should be allowed to give speeches at British universities? *Yes; No*
Q26	Do you think that UK Prime Minister Theresa May should be allowed to give speeches at US universities? *Yes; No*
Q27	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? *[Textbox]*
Q28	If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? *[Textbox]*
Q29	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? *[Textbox]*
Q30	What is your gender? *Male; Female*
Q31	What is your age? *[Textbox]*
Q32	What is your ethnic group? *White; Asian or Asian British; Arab; Black or Black British; Mixed; Other, please specify [Textbox]*
Q33	What is your current level of study? *Undergraduate (BA, BSc, LLB, BEng); Postgraduate Taught (MA, MSc, MRes, LLM, MEng, MBA); Postgraduate Research (PhD, MPhil); Other, please specify [Textbox]*
Q34	What is your department at the University of Essex? *Biological Sciences; Computer Science and Electronic Engineering;*

	Economics; Edge Hotel School; Essex Business School; Government; Health and Human Sciences; History; Human Rights Centre; Interdisciplinary Studies in the Humanities; International Academy; Institute for Social and Economic Research; Language and Linguistics; Law; Literature, Film, and Theatre Studies; Mathematical Sciences; Philosophy and Art History; Psychoanalytic Studies; Psychology; Sociology
Q35	What course are you studying? *[Textbox]*
Q36	Why have you chosen to study *[Q35 response]* at the University of Essex? *[Textbox]*
Q37	Is English your first language? *Yes; No*
Q38	Overall, how much attention were you able to pay to the survey? *1 = A lot of attention; 2; 3; 4; 5 = No attention at all*
Q39	How distracted did you feel by the things going on around you while completing the survey? *1 = Not distracted at all; 2; 3; 4; 5 = Extremely distracted*
Q40	How private did you feel the survey was? *1 = Completely private; 2; 3; 4; 5 = Not private at all*
Q41	Finally, what did you think of this questionnaire? *1 = Certainly not; 2; 3; 4; 5 = Certainly yes* Was it difficult to answer the questions? Did you understand the questions? Did the questionnaire make you think? Was the topic interesting? Was the questionnaire too long? Did you enjoy answering the questions? Is the survey important for science?
Q42	If you were to complete this survey again, which device would you choose to use? *Laptop or desktop PC; Tablet; Smartphone*
Q43	If you would like to make any additional comments about the survey, please use this space. *[Textbox]*

3. Willingness to use mobile technologies for data collection in a probability household panel

Abstract

We asked members of the *Understanding Society* Innovation Panel about their willingness to participate in various data collection tasks on their mobile devices. We find that stated willingness varies considerably depending on the type of activity involved: respondents are less willing to participate in tasks that involve downloading and installing an app, or where data are collected passively. Stated willingness also varies between smartphones and tablets, and between types of respondents: respondents who report higher concerns about the security of data collected with mobile technologies and those who use their devices less intensively are less willing to participate in mobile data collection tasks.

3.1. Introduction

Mobile technologies, including smartphones and tablets, can be used in various ways for data collection. On the one hand, mobile devices allow administering survey questionnaires in innovative ways: respondents can be asked to answer questions sent via text messaging, or to complete questionnaires in a mobile web browser or in a survey app installed on a smartphone or tablet. These forms of survey administration allow near real-time data collection, for example as part of ecological momentary assessment in psychological studies (Moskowitz & Young, 2006), that make it possible to collect more detailed and more wide-ranging measures across multiple time points while reducing the need to recall information. On the other hand, mobile technologies enable researchers to collect new forms of data from survey respondents by relying on the additional measurement capabilities of mobile devices. GPS data can be collected from the respondent's mobile device to measure their location and travel patterns (e.g., Geurs, Veenstra, & Thomas, 2013), or to trigger surveys at pre-specified locations using geo-fencing (e.g., Ginnis, 2017). Accelerometer data can similarly be collected from the respondent's mobile device (e.g., Lathia, Sandstrom, Mascolo, & Rentfrow, 2017), as can data from external devices that are connected via Bluetooth, such as activity trackers (e.g., Scherpenzeel, 2017), smart scales (e.g., Kooreman & Scherpenzeel, 2014), or transdermal devices (e.g., Greenfield, Bond, & Kerr, 2014). Such data can be used to measure physical activity as well as other biological features, such as weight, body fat, and stress. Other possibilities of mobile data collection include asking respondents to take photos with the camera of their smartphone or tablet, for example to scan payslips or shopping receipts (e.g., Jäckle, Burton, Couper, & Lessof, 2017), or to track how respondents are using their mobile device (e.g., Revilla, Ochoa, & Loewe, 2016), for example which websites they are visiting. These new forms of data, some of which cannot feasibly be collected with survey questionnaires, can supplement or potentially even replace data collected using questionnaire-based methods.

Depending on the population of interest, however, not all subgroups will have access to mobile devices. In 2017, 76 percent of households in the United Kingdom reported owning a smartphone and 58 percent reported owning a tablet, but there are large differences by age and socio-economic status (Ofcom, 2017). Socio-demographic differences in coverage are similar in the United States and in other Western countries (Poushter, 2016). To reduce coverage bias in studies with mobile data collection, sample members without mobile device access or Internet access could be provided with a smartphone or tablet and a mobile Internet connection. This approach has already been implemented in two associated studies of the LISS Panel, a probability-based online panel in the Netherlands: the Smartphone

Time Use Study and the Mobile Mobility Study (Scherpenzeel, 2017). Among those who have access to mobile devices, further potential barriers are whether individuals would actually be able and willing to participate in studies involving mobile data collection.

A few studies have started to examine the stated willingness of respondents to perform additional data collection tasks on their mobile device as part of a survey, and to explore which factors are associated with willingness. Results suggest that the level of willingness varies by data collection task: stated willingness is higher for tasks where respondents have control over the transmitted content than for tasks where data are collected automatically, even if those tasks require more effort from the respondent (Revilla, Couper, & Ochoa, 2017; Revilla, Toninelli, et al., 2016). In addition, stated willingness varies with respondent characteristics. Respondents who use their device more intensively, measured by how often they download apps on their smartphone and the number of apps they regularly use, are more willing to participate in mobile data collection tasks (Keusch, Antoun, Couper, Kreuter, & Struminskaya, 2017; Pinter, 2015). In contrast, stated willingness is lower among people with higher privacy and security concerns and people with lower levels of trust that institutions will protect their data (Keusch et al., 2017; Revilla et al., 2017). Study characteristics also matter: stated willingness is higher for studies that are sponsored by a university rather than a government agency, studies that include incentives, and those that run over a shorter period of time overall (Keusch et al., 2017).

The literature examining stated willingness to participate in mobile data collection tasks has several limitations. First, all studies rely on data from opt-in online panels rather than probability samples of the general population. The sample members of these panels are self-selected and might be more cooperative than the general population. Second, existing research lacks a theoretical discussion of the underlying mechanisms of willingness. Third, while existing studies have examined the implications of respondent and study characteristics, no studies have examined the interactions of respondent and task characteristics in determining willingness.

In this paper, we examine the stated willingness of the general population to use mobile technologies for a range of data collection tasks, and what affects willingness. Studying hypothetical rather than actual willingness allows us to understand the determinants of willingness across a range of tasks among a general population sample. Although hypothetical measures of willingness might be influenced by context effects, as other subjective measures in surveys (Sudman, Bradburn, & Schwarz, 1996), these measures have been shown to reflect actual behaviour. Jäckle, Burton, Couper, and Lessof (2017) find that hypothetical willingness is predictive of participation in a mobile app study: respondents who indicated that they would be willing to download and install a survey app on their mobile device have a 4.4 percentage point higher predicted probability of using an app to provide data about their expenditure compared to respondents who reported that they are a little or not at all willing.

We propose a framework of how characteristics of the data collection task (that might constitute potential barriers to participation), respondent characteristics, and interactions between the two, can affect willingness to participate in mobile data collection. We use data on 1,660 survey respondents of the *Understanding Society* Innovation Panel, a nationally representative household panel study in Great Britain, who reported using a smartphone or tablet, to examine the following research questions:

(1) How does stated willingness to use mobile technologies vary across different data collection tasks?

(2) How does stated willingness to do different tasks vary between smartphone and tablet?

(3) Which respondent characteristics predict stated willingness to do different tasks?

(4) Which task characteristics predict stated willingness, and does the effect depend on respondent characteristics?

3.2. Task characteristics and respondent characteristics associated with willingness to participate in mobile data collection

Mobile data collection tasks have various characteristics that constitute potential barriers to participation and that might affect the respondent's willingness to take part. In Table 3.1, we outline five key characteristics for a range of data collection tasks.

A first characteristic is that most data collection tasks require respondents to download and install an app on their smartphone or tablet to be able to take part in the data collection process. For some tasks, respondents also need to activate features on their device (for example turning on Bluetooth) or give data capture permissions (for example allowing the app to capture GPS coordinates of the mobile device). Only a few tasks, including administering a web questionnaire in the mobile browser or administering a questionnaire by text messages, can solely rely on apps that are already installed on the respondent's device and that do not need any additional permissions by the respondent.

Second, the data collection activities differ in how actively they involve the respondent in the data collection process, which affects how much control respondents have over the content measured, and how much of their time the task takes. Some activities require respondents to actively complete measurements, such as answering questions in a survey app or taking photos. These activities give respondents full control over what information they provide to the researcher. Other activities, such as GPS location tracking, rely on passive measurement and do not involve respondents in the data collection process once they have downloaded and installed an app and given consent to data collection. For these activities, the only control respondents have over what is measured is that they can switch off the data collection process. Passive data collection activities allow the collection of continuous data: the GPS location of a mobile device, for example, can be tracked continuously over a certain period.

Table 3.1. Characteristics of mobile data collection tasks.

Mobile data collection task	(1) Requires downloading and installing an app	(2) Role of respondent	(3) Requires uploading mobile data	(4) Technical demands	(5) Potential privacy threat
Questionnaire	No	Active	Yes	Low	Content-dependent
Survey app	Yes	Active	Yes	Low	Content-dependent
Device usage tracking app	Yes	Passive	Yes	High	Yes
Text messages	No	Active	No	Low	Content-dependent
Camera	Yes	Active	Yes	High	Content-dependent
Accelerometer	Yes	Passive	Yes	High	Content-dependent
GPS	Yes	Passive	Yes	High	Yes
Bluetooth linkage to external device	Yes	Passive	Yes	High	Content-dependent

Third, all data collection tasks, except those that rely on text messaging for data transmission, require that data are uploaded as part of the data collection process, which might affect mobile data usage limits and associated costs. The amount of data to be uploaded varies between activities and also depends on how the activity is implemented. Uploading photos, for example, is likely to require more data than uploading responses from a mobile questionnaire; uploading GPS coordinates that are collected continuously is likely to require more data than uploading coordinates that are collected at certain intervals.

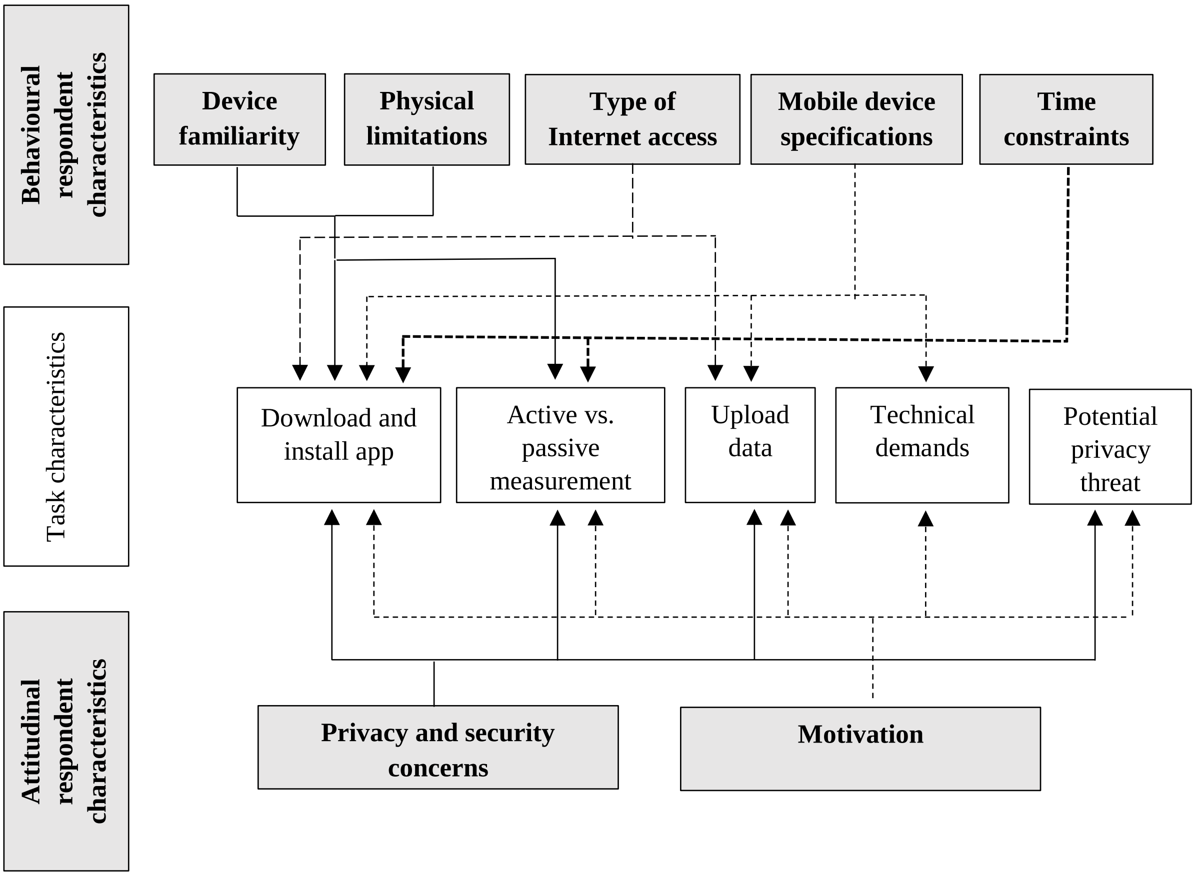
Fourth, mobile data collection tasks have different technical demands, including how much battery power and storage capacity they require. Tasks that collect data via sensors, such as GPS or accelerometer, as well as tasks that rely on apps that are continuously running in the background, such as an app that tracks how respondents use their mobile device, are likely to reduce battery life more than tasks that rely on apps that are only used intermittently, such as answering questions sent via text messaging. The required storage capacity also varies between tasks, for example taking photos for data collection requires more storage capacity, as photos need to be stored on the mobile device before they are sent to the researcher, whereas other tasks require no additional storage capacity, for example tasks that use the mobile browser that is already installed on the respondent's mobile device. In Table 3.1, we classify the technical demands of tasks in relative terms; we code tasks as highly demanding if they consume a lot of battery power, require a lot of storage capacity, or both. How each task is implemented, for example how frequently GPS coordinates are captured, can affect the technical demands.

Finally, the data collection activities differ in the extent to which they potentially intrude on the respondent's privacy. GPS data are of a more private nature as they could possibly be used to identify an individual. Similarly, data from an app that tracks the respondent's usage of their phone are of a more private nature. For other tasks, privacy concerns are likely to depend on the content of the data collected. For example, accelerometer data might be perceived as private by some people, in a similar way as self-reports on physical activity might be sensitive for some people.

As data collection tasks differ in what they require from respondents, willingness to use them is likely to vary between tasks, but also between types of respondents: some requirements might constitute barriers to participation for some people but not for others. Figure 3.1 represents the conceptual determinants of willingness: task characteristics, respondent characteristics, and interactions between the two. The relevant respondent characteristics include both behavioural and attitudinal characteristics.

Device familiarity. Respondents who feel more comfortable and confident with using their mobile device, who use their device more frequently, or who already use similar device features for their own purposes might be more willing to participate in mobile data collection tasks. Device familiarity might especially affect tasks that require respondents to download and install an app, and those that actively involve respondents in the data collection process. Previous research has shown that device familiarity is associated with increased smartphone use to complete web questionnaires (Couper et al., 2017), and a similar association can be expected between device familiarity and the willingness to use mobile technologies.

Figure 3.1. Task characteristics and respondent characteristics that can affect the willingness to participate in mobile data collection tasks.



Physical limitations. Respondents with physical limitations, in particular visual impairment and limited manual dexterity, may find it harder to use mobile devices (McGaughey, Zeltmann, & McMurtrey, 2013) and may therefore be less willing to participate in mobile data collection tasks. Physical limitations are also more likely to affect technologies that require respondents to download and install an app, and to be actively involved in the data collection process.

Type of Internet access. The way that respondents connect their mobile device to the Internet may be another determinant of how willing they are to participate in mobile data collection. Respondents who only use mobile Internet and have limited mobile data allowances or a pay-as-you-go plan may be less willing to participate in mobile data collection than those with unlimited data plans or WiFi access at home. The type of Internet access is particularly relevant for data collection tasks that require downloading an app and uploading a large amount of mobile data.

Mobile device specifications. The technical specifications of the mobile device that respondents use may also affect their willingness to participate in mobile data collection. Respondents may not have sufficient storage capacity on their device to download and install apps or to store data, they may use older mobile devices with shorter battery life and slower processing speed, they may not have an app store account, or they may use an operating system for which the data collection app has not been developed. Depending on the specification of their device, respondents may hence be less able and willing to participate in mobile data collection, in particular to complete tasks that require downloading an app, or that use a large amount of storage capacity and battery power.

Time constraints. Busy people, including respondents with long working and commuting hours, and those with young children and caring responsibilities, may be less willing to participate in data collection requests using mobile technologies. They may be particularly reluctant to complete tasks that require active involvement in the data collection process and repeated participation. People with time constraints were shown to have lower response propensities in surveys (Abraham, Maitland, & Bianchi, 2006; Groves & Couper, 1998), which suggests that a similar association can be expected between time constraints and willingness to participate in additional data collection requests that are beyond survey interviews.

Privacy and security concerns. Mobile technologies have the potential to collect personally identifying information automatically on a large scale, including photos, GPS coordinates and device use profiles. Respondents might consider these data collection activities intrusive to their privacy, and might be concerned about data security when providing sensitive information to researchers via mobile technologies (Chin, Felt, Sekar, & Wagner, 2012). Respondents who have greater concerns about privacy and data security might be less willing to participate in mobile data collection tasks, in particular to complete tasks that involve downloading an app, that are potentially intruding to privacy and tasks where respondents have little control over the transmitted content.

Motivation. Respondents who have a strong sense of loyalty or commitment to the study, or who are highly interested in the survey topic may be more willing to accept each of the potential barriers to participation in mobile data collection. Previous research on the willingness to comply with in-survey requests has, for example, found that respondents who were cooperative in previous survey interviews were also more likely to give consent to administrative data linkage (Sakshaug, Couper, Ofstedal, & Weir, 2012).

3.3. Data and Methods

Survey

We use data from wave 9 of the *Understanding Society* Innovation Panel, a nationally representative household panel study in Great Britain funded by the UK Economic and Social Research Council and led by the Institute for Social and Economic Research at the University of Essex (University of Essex. Institute for Social and Economic Research, 2017). The Innovation Panel is based on a stratified, clustered sample of households in England, Scotland, and Wales (Lynn, 2009). In addition to the original sample from wave

1, refreshment samples were drawn at waves 4 and 7. The interview is conducted annually among all household members aged 16 and older. Households where no household member participates in two consecutive years are no longer issued to the field. At wave 9, a random two-thirds of sample households were allocated to a sequential mixed-mode design, where non-respondents to the web survey were followed up by face-to-face interviewers. The other third of households were first approached by face-to-face interviewers. In the final phase of fieldwork non-respondents were given the option of completing the survey online or by telephone. Of 1,399 households issued at wave 9, 84.7 percent responded. In responding households, 85.4 percent of individuals completed a full interview (see Jäckle, Gaia, Al Baghal, Burton, & Lynn, 2017). Data for wave 9 were collected between May and September 2016. For details on the survey design and fieldwork see the documentation available at https://www.understandingsociety.ac.uk/documentation/innovation-panel.
The data are available from the UK Data Service at
https://discover.ukdataservice.ac.uk/catalogue/?sn=6849.

Measures of willingness to use mobile technologies

Respondents who indicated that they use the Internet for personal purposes were asked: "Which of the following devices do you use to connect to the Internet?" (yes, no)

(1) Desktop computer

(2) Laptop

(3) Smartphone

(4) Tablet

(5) Feature phone / non-touchscreen mobile phone

(6) E-book reader (e.g., Kindle)

(7) Smartwatch

(8) Other

Following the question about device use, we asked respondents who use a smartphone: "How willing would you be to carry out the following tasks on your smartphone for a survey?" (very willing, somewhat willing, a little willing, not at all willing)

(1) Complete an online questionnaire on your mobile phone
(2) Download a survey app to complete an online questionnaire
(3) Download an app which collects anonymous data about how you use your smartphone
(4) Answer a couple of questions sent via text messaging
(5) Use the camera of your smartphone to take photos or scan barcodes
(6) Allow built-in features of your smartphone to measure the frequency and speed at which you walk, run or cycle
(7) Share the GPS position of your smartphone
(8) Connect your smartphone via Bluetooth to other electronic devices (e.g., wearables such as Fitbit).

Similarly, respondents who reported using a tablet were asked about the subset of tasks for which tablets are typically used: "How willing would you be to carry out the following tasks on your tablet for a survey?" (very willing, somewhat willing, a little willing, not at all willing)

(1) Complete an online questionnaire on your tablet
(2) Download a survey app to complete an online questionnaire
(3) Download an app which collects anonymous data about how you use your tablet
(4) Use the camera of your tablet to take photos or scan barcodes
(5) Connect your tablet via Bluetooth to other electronic devices (e.g., wearables such as Fitbit).

If respondents reported using both devices, they were asked both sets of questions – first about their willingness to complete tasks on their smartphone, then about their tablet. As the questions were only asked of respondents who said that they have access to a smartphone, to a tablet, or both, our analyses of willingness are conditional on reported mobile device access.

In the face-to-face interview, the questions were implemented in the computer-assisted self-interviewing (CASI) section to reduce potential mode effects due to the mixed-mode design of the Innovation Panel. In this section, the interviewer passed the laptop to the respondents and asked them to complete the questions on their own.

Of the 2,174 respondents who gave a full interview, 48 respondents were excluded because they participated in the CAPI interview but refused or were not able to do the self-completion section; 31 respondents were excluded because they gave a CATI interview in the final non-response conversion stage and were not asked the self-completion section; a further 190 respondents were excluded because they do not use or have access to the Internet. This leaves 1,905 Innovation Panel respondents who were asked about mobile device access. Among those respondents, 87.1 percent reported having access to either a smartphone or a tablet and were hence asked about willingness (N = 1,660). The remaining 12.9 percent have no access to either mobile device or provided missing values to both questions on mobile device access and were excluded from the analytic sample (N = 245). The majority of respondents with mobile device access use both devices (59.0 percent) whereas 23.7 percent only use a smartphone and 16.5 percent only a tablet.

The data were weighted for all analyses to account for unequal selection probabilities and differential nonresponse. Standard errors were adjusted to account for the stratified, clustered sample design of the *Understanding Society* Innovation Panel. All analyses were conducted using the svy procedures in Stata.

Respondent-level predictors of willingness

This section describes how we operationalised the respondent-level predictors of our framework. Descriptive statistics for the predictors are shown in Table 3.2. The full wording of questions is documented in Appendix Table 3.9; numbers in parentheses index the corresponding questions in the Appendix.

Device familiarity. We use three measures of device familiarity which were asked separately for smartphone and tablet: frequency of use, intensity of use, and self-rated skill. We coded ***frequency of device use*** (Q4) as 1 if the device is used daily, and 0 otherwise. The categories were collapsed rather than included as an ordinal or continuous measure because the distribution is highly skewed. To measure ***intensity of use*** (Q5), we asked respondents which activities they carry out on their device. We include the number of activities carried out as a count variable, ranging from 0 to 12. Finally, we asked respondents to rate their ***skills using a mobile device*** (Q6). We include self-rated skill as a continuous variable, ranging from 1 = Beginner to 5 = Advanced.

Physical limitations. We include an indicator of whether the respondent has any ***physical limitations***: coded as 1 if the respondent has any visual impairment apart from wearing standard glasses or has limited manual dexterity, and coded as 0 otherwise. Note from Table

3.2 that this variable is highly skewed: among the sample of mobile device users, most respondents do not have any physical limitations.

Type of Internet access. To measure how respondents ***access the Internet*** (Q2), we use an indicator coded as 1 if the respondent has WiFi at home, and 0 if not. Again, note from Table 3.2 that most people have WiFi access from home. We also asked smartphone users about the ***type of data plan*** (Q3) they have. The variable is coded as 1 if the respondent has a pay-as-you-go plan, and 0 if the respondent has a fixed data plan with a monthly data allowance or uses WiFi only.

Time constraints. We derived an indicator for the respondent's ***time constraints***: coded as 1 if the respondent is employed or self-employed and works for more than 40 hours per week, or commutes to work for more than one hour one-way, or has young children under the age of five in the household or other caring responsibilities, and coded as 0 otherwise.

Security concerns. We asked respondents to rate their ***security concerns*** (Q8) when providing information using various mobile technologies: whether they are not at all concerned, a little concerned, somewhat concerned, very concerned, or extremely concerned. They were asked about the same set of technologies as in the willingness questions: smartphone users were asked about eight different technologies, tablet users about five technologies. Respondents with access to both smartphone and tablet were asked this question only once, about security concerns on smartphone and tablet at the same time. To measure the average level of security concerns across technologies, we use the mean of the individual security concern items, ranging from 1 (if the respondent is not at all concerned about any technologies) to 5 (if the respondent is extremely concerned about all technologies).

Table 3.2. Descriptive statistics of respondent characteristics.

		Smartphone users		**Tablet users**	
		%	**N**	**%**	**N**
Frequency of use	Every day	81.2		52.6	
	Less than every day	18.8	1,378	47.4	1,260
Number of activities	Mean	8.2		6.7	
	SD	3.2		3.4	
	Min; Max	0; 12	1,378	0; 12	1,258
Self-reported skill	Mean	3.7		3.6	
	SD	1.1		1.1	
	Min; Max	1; 5	1,378	1; 5	1,260
Physical limitations	Yes	4.5		5.3	
	No	95.5	1,376	94.7	1,259
WiFi access at home	Yes	97.5		98.6	
	No	2.5	1,379	1.4	1,261
Type of smartphone contract	Pay-as-you-go plan	11.0		--	
	Fixed data plan or WiFi only	89.0	1,377	--	
Time constraints	Yes	27.1		23.5	
	No	72.9	1,379	76.5	1,261
Security concerns	Mean	2.6		2.7	
	SD	1.0		1.1	
	Min; Max	1; 5	1,366	1; 5	1,250
Item-nonresponse	≥1 items missing	61.4		61.1	
	No items missing	38.6	1,379	38.9	1,261
Consent to data linkage	Yes	59.8		59.0	
	No	40.2	1,347	41.0	1,232
Mode of interview	Face-to-face	42.0		42.0	
	Web	58.0	1,379	58.0	1,261
Number of eligible waves	1-3	35.1		31.6	
	4-6	25.7		24.9	
	7-9	39.2	1,379	43.5	1,261
Proportion of full interviews	Mean	0.9		0.9	
	SD	0.2		0.2	
	Min; Max	0.1; 1	1,379	0.1; 1	1,261
Gender	Female	53.9		56.8	

	Male	46.1	1,379	43.2	1,261
Age	Mean	42.9		47.7	
	SD	15.7		16.7	
	Min; Max	16; 87	1,379	16; 91	1,261
Education	Higher degree	43.2		44.0	
	A-level	26.0		23.3	
	GCSE	23.8		24.5	
	No qualification	6.9	1,368	8.2	1,254
Labour force status	In work	68.3		61.9	
	Not in work	31.7	1,378	38.1	1,259
Individual monthly gross income in £	Mean	2,001.6		1,981.3	
	SD	1,828.3		1,711.4	
	Min; Max	0; 15,000	1,379	0; 15,000	1,261
Housing tenure	Has own house	75.0		80.6	
	Not own house	25.0	1,378	19.4	1,260

Motivation. We include several measures of respondent motivation and engagement with the study. The first indicator is whether the respondent has any ***item-nonresponse*** in the survey, coded as 1 if the respondent has at least one missing item among the questions prior to the questionnaire module on willingness, and 0 otherwise. The second indicator is whether the respondent gave ***consent to link*** their survey data with credit rating data held by the Financial Conduct Authority (FCA), coded as 1 if the respondent gives consent, and 0 if not. As the consent rate to data linkage is considerably lower in web than in face-to-face (Burton, 2016), we also control for the ***mode of data collection***, coded as 1 if web and 0 if face-to-face. The third indicator is the ***number of waves for which the respondent has been eligible***: whether the respondent has been a member of the *Understanding Society* Innovation Panel for 7-9 waves (original sample member or joined the panel in wave 2 or 3), for 4-6 waves (member of the wave 4 refreshment sample or joined the panel in wave 5 or 6), or for 1-3 waves (member of the wave 7 refreshment sample or joined the panel in wave 8 or 9). The final indicator is the ***proportion of waves in which the respondent was eligible and gave a full interview***, ranging from 0.11 to 1.

Socio-demographics. Finally, we control for a set of socio-demographic characteristics, including gender, age, education, labour force status, income, and housing tenure, to help identify the genuine effects of respondent characteristics and attitudes. ***Gender*** was coded as 1 if female and 0 if male. We include a variable for ***age*** and one for ***age-squared*** as age was found to have a curvilinear relationship with willingness. ***Education*** was coded in four categories: whether the respondent has a professional or a university degree, has A-levels (equivalent to 13 years of schooling in the UK), has GCSE (equivalent to 11 years of schooling in the UK), or has no qualifications. ***Labour force status*** was coded as 1 if the respondent is in work (employed or self-employed), and 0 if not in work. To measure ***income***, we use a derived indicator of the respondent's monthly gross income that is provided with the data set, including earnings from employment and self-employment as well as unearned income from benefits, pensions and other sources. Income was top-coded to a maximum value of £15,000. In the model, we take the natural logarithm as the distribution of income is highly skewed. ***Housing tenure***, used as a measure of wealth, was coded as 1 if the respondent lives in their own house (with a mortgage or owned outright), and 0 otherwise.

Task-level predictors of willingness

To examine the association between task characteristics and the willingness to participate in mobile data collection, we coded the characteristics of each of the eight types of mobile data collection tasks according to Table 3.1: whether the data collection task requires respondents to ***download and install an app*** (coded as 1 if yes 0 if no); whether respondents have an ***active role*** in the data collection process (coded as 1 if respondents are actively and 0 if they are passively involved); whether the task has relatively ***high technical demands*** (coded as 1 for high technical demands and 0 for low demands); and to what extent the data collection ***intrudes on the respondent's privacy*** (coded as 1 if the activity represents a privacy threat and 0 if the privacy threat is content-dependent). We do not include an indicator of whether the data collection task involves uploading mobile data because it would only represent one activity: completing a survey by text messages.

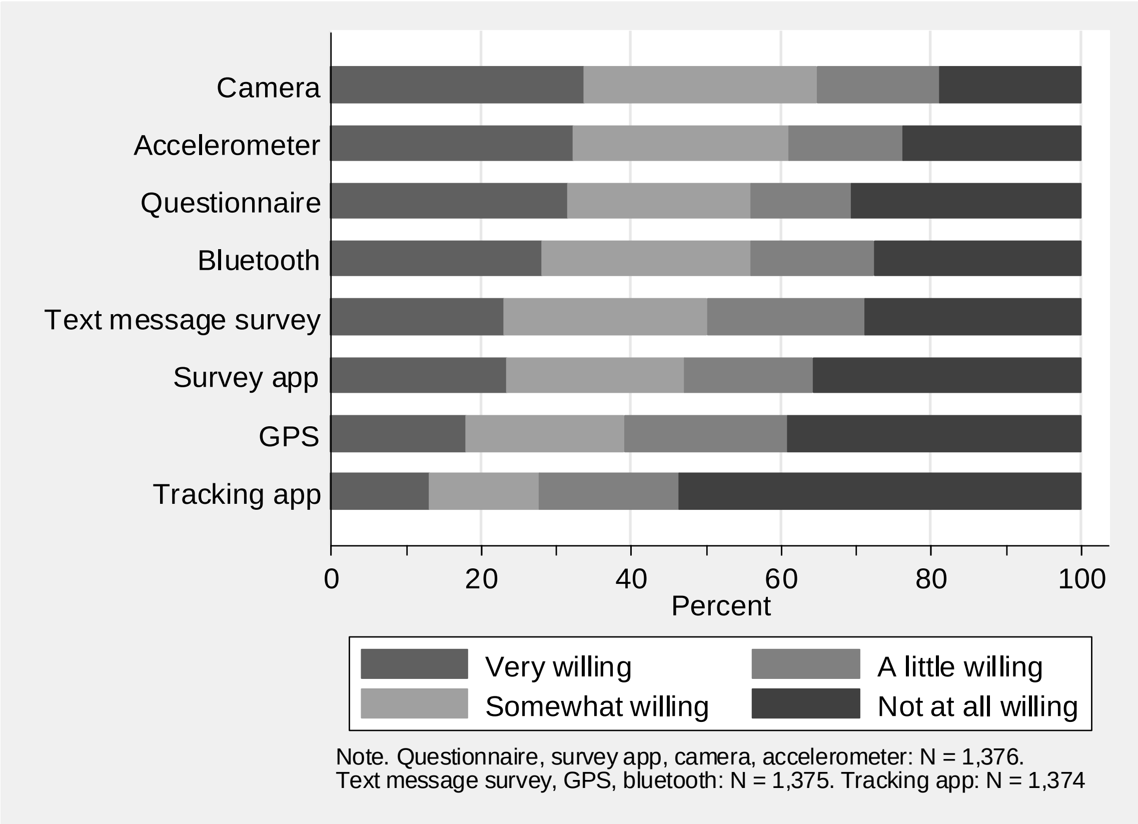
3.4. Results

RQ1. How does stated willingness to use mobile technologies vary across different data collection tasks?

Stated willingness to use mobile technologies on a smartphone for data collection varies considerably by data collection task (Figure 3.2, Table 3.5 in the Appendix). On the one hand, the majority of smartphone users would be (very or somewhat) willing to use the camera of their smartphone to take photos or to scan barcodes for a survey (65 percent). A similar proportion of respondents would be willing to allow the accelerometer built into their smartphone to measure their physical movement (61 percent). On the other hand, a much smaller proportion of smartphone users would be willing to share the GPS position of their phone (39 percent) and only 28 percent would be willing to download and use a tracking app that collects anonymous data about how they use their phone. More than half of respondents would be not at all willing to do this task.

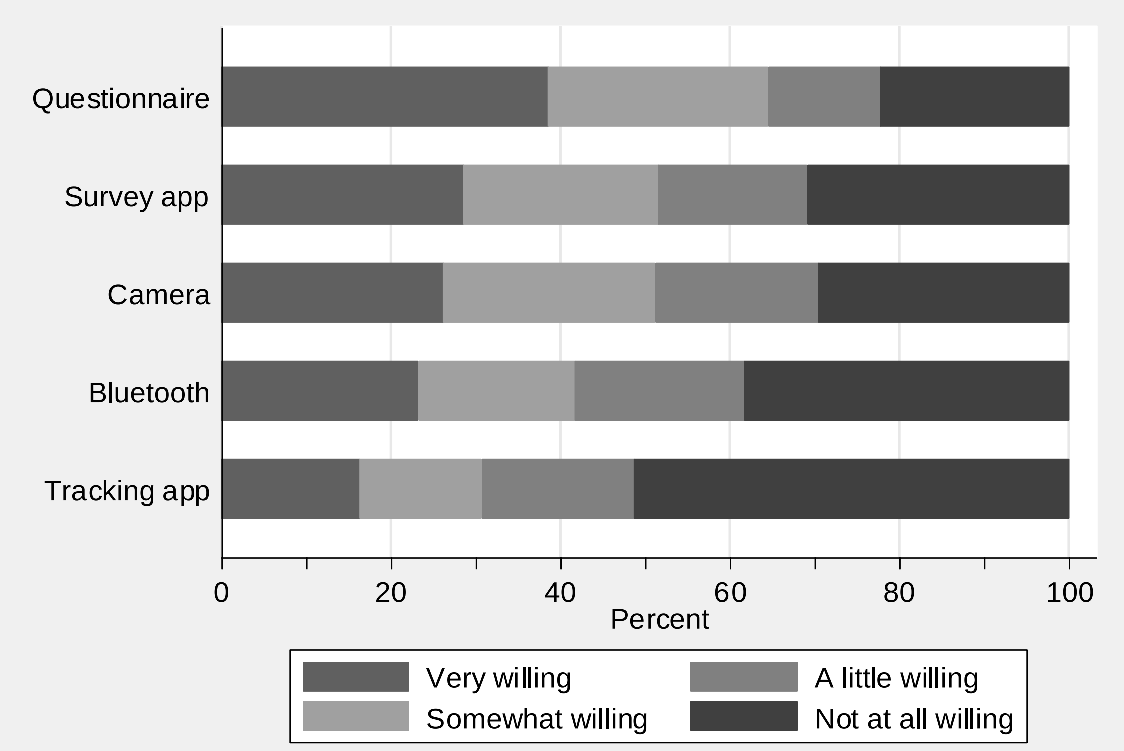
These findings suggest that not all smartphone users would be willing to use all kinds of technologies on their phone for data collection, and that they make a clear distinction between different tasks, depending on what type of technology the tasks involve.

Figure 3.2. Stated willingness to complete data collection tasks on a smartphone.



When asking tablet users about their stated willingness to participate in mobile data collection, we find that willingness varies across data collection tasks in a similar way, but there are some notable differences compared to smartphone users (Figure 3.3, Table 3.6 in the Appendix). A smaller percentage of tablet users would be willing to use the camera of their tablet to take photos or scan barcodes for a survey (51 percent), presumably as they are less used to taking photos on their tablet. A larger percentage, however, would be willing to complete an online questionnaire on their tablet (64 percent), presumably because it is easier to complete surveys on devices with a larger screen size.

Figure 3.3. Stated willingness to complete data collection tasks on a tablet.



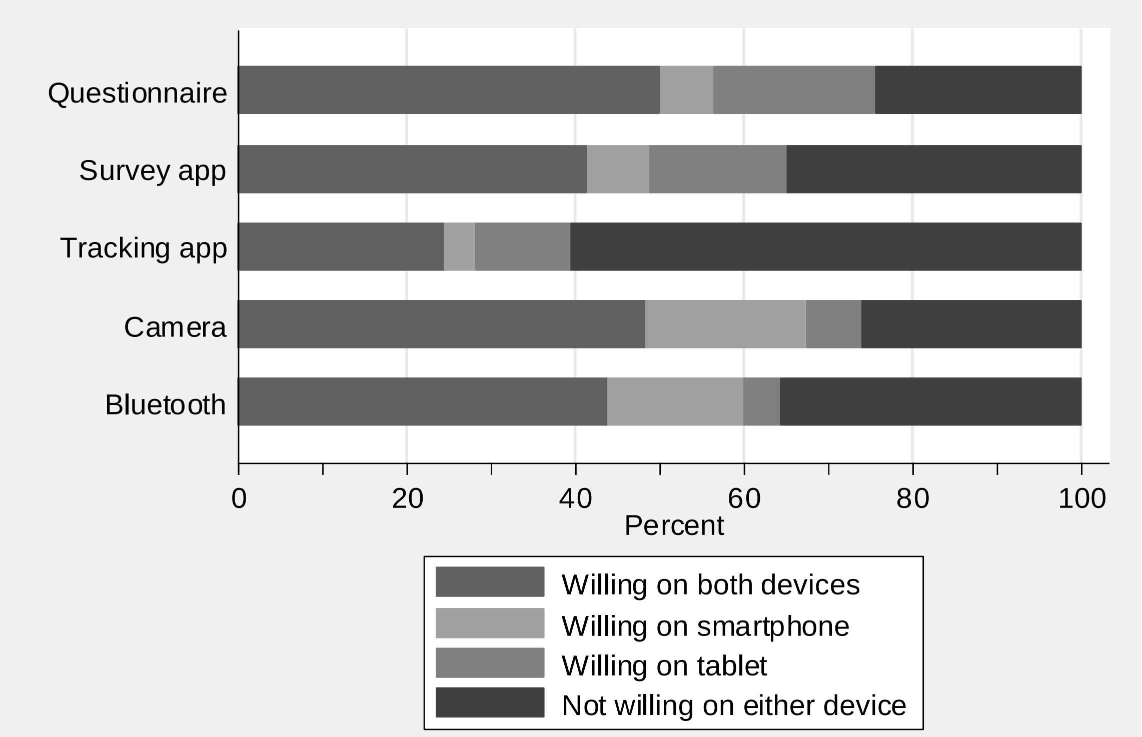
Note. Questionnaire, survey app, camera, bluetooth: N = 1,258. Tracking app: N = 1,257.

Comparing the stated willingness of smartphone users and tablet users gives a first indication that respondents also make a distinction between devices: they are more willing to complete certain tasks on a smartphone than on a tablet or vice versa. This first set of analyses, however, is based on two different albeit overlapping populations: those who use a smartphone compared to those who use a tablet. In the next section, we examine the stated preferences of the 980 respondents who have access to both devices to better understand how willingness differs between smartphones and tablets.

RQ2. How does stated willingness to do different tasks vary between smartphone and tablet?

To simplify the analysis, we dichotomised the four-point willingness scale: we coded very willing and somewhat willing as *willing*, and a little willing and not at all willing as *not* willing. We then compared if respondents are willing to complete data collection tasks on both devices, only on one device, or on neither device. As shown in Figure 3.4 (and in Table 3.7 in the Appendix), we find that a large majority of respondents have consistent levels of willingness: they are equally willing or equally unwilling to complete data collection tasks on a smartphone or on a tablet. The level of consistency varies slightly by data collection task. Respondents are most consistent in their willingness to use a tracking app that collects anonymous data about how they use their mobile device (85 percent are equally willing or equally unwilling), and least consistent in their willingness to complete a questionnaire in the mobile browser (still 75 percent are equally willing or equally unwilling).

Figure 3.4. Consistency of stated willingness among respondents with access to smartphone and tablet.



Note. Questionnaire, survey app, camera: N = 977. Bluetooth: N = 976.
Tracking app: N = 975.

To test the relationship between willingness to complete a given task on a smartphone and willingness to complete the task on a tablet, we computed Kendall's tau-b correlation coefficients, that measure the association between two ordinal variables. We find a moderate to strong positive correlation for all tasks, ranging from $\tau_b = 0.49$ for completing an online questionnaire to $\tau_b = 0.65$ for connecting to other devices via Bluetooth, which confirms the interpretation of Figure 3.4, that willingness is moderately consistent between devices.

Among respondents who expressed different levels of willingness across devices, the preference is task-related: the majority would be more willing to use their tablet to complete an online questionnaire, to use a survey app, or to use a tracking app that collects anonymous data about how they use their device, but would be more willing to use their smartphone to take photos or to connect to other devices via Bluetooth. These differences in preference may reflect how respondents use the devices. Respondents may use the camera of their smartphone more often than the camera of their tablet. For survey-related tasks including completing an online questionnaire and using a survey app, respondents seem to prefer devices with a larger screen size.

These findings suggest that stated willingness is consistent for the majority of respondents, but some respondents make a distinction between different devices. We therefore cannot assume that all respondents who have multiple devices would be equally willing to do the same type of task on all devices.

RQ3. Which respondent characteristics predict stated willingness to do different tasks?

Table 3.8 in the Appendix shows the bivariate relationship of respondent characteristics and stated willingness to complete different data collection tasks. To facilitate later analyses, the willingness scale was dichotomised into *willing* (combining very willing and somewhat willing) and *not willing* (combining a little willing and not at all willing). We find a significant association in the expected direction for most characteristics, including device familiarity, physical limitations, and security concerns. Two of the indicators of motivation, however, suggest a significant relationship with willingness that is opposite to what we expected: respondents who were sampled longer ago and are still in the panel appear to be less willing to participate in mobile data collection than panel members who were sampled more recently, although the effect is statistically significant for only three of the tasks. Contrary to our expectation, respondents who completed all previous interviews in which they were eligible seem to be less willing to participate in mobile data collection than those who did not complete all previous interviews, but the effect is statistically significant for only two of the tasks. Type of Internet access as well as time constraints do not have a significant bivariate relationship with willingness for any of the data collection tasks.

To further understand which respondent characteristics are associated with stated willingness to complete different data collection tasks, we ran regression models for each of the individual tasks, using different specifications. First, we fitted a series of ordered logistic regression models using the ordinal willingness scale as dependent variable, separately for smartphone and tablet. Second, we fitted a series of binary logistic regression models using the dichotomised willingness scale as dependent variable. Table 3.3 shows the results of the binary logistic regression models for willingness to complete data collection tasks on a smartphone. The binary logistic regression models for tablet and the ordered logistic regression models for smartphone and those for tablet all yield very similar results, so we do not present them in this paper.

We show the average marginal effects that denote the increase in the predicted probability of being willing for a one-unit change in the explanatory variable. The average marginal effect of frequency of smartphone use in the first model, for example, shows that respondents who use their smartphone every day have a 6.8 percentage point higher predicted probability to be willing to take photos on their smartphone for a survey compared to those who use their device less frequently, although the effect is not statistically significant. To recall the different levels of willingness across data collection tasks, we also show the proportion of smartphone users who reported that they are very or somewhat willing to complete the individual tasks in the first row of the table (shaded). As we replicate the models for the eight different smartphone data collection tasks, we adjusted the p-values of the average marginal effects estimated from the logistic regressions using the Holm-Bonferroni method to account for multiple testing (Holm, 1979).

Table 3.3. Logistic regression models predicting willingness to complete data collection tasks on a smartphone. Average marginal effects.

	Camera	Accelero-meter	Questionnaire	Bluetooth	Text messages	Survey app	GPS	Tracking app
Proportion Willing (n = 1,379)	0.648	0.609	0.559	0.559	0.501	0.470	0.391	0.276
Device familiarity								
Use smartphone every day	0.068	-0.036	0.091	-0.043	0.080	0.037	-0.027	0.014
	(0.041)	(0.029)	(0.046)	(0.040)	(0.060)	(0.051)	(0.053)	(0.055)
Number of activities on	0.015	0.035***	0.036***	0.028**	0.018	0.036***	0.027***	0.030**
smartphone	(0.007)	(0.005)	(0.005)	(0.006)	(0.009)	(0.006)	(0.005)	(0.006)
Self-rated skill	0.038	0.033	0.043	0.053	0.011	0.054	0.029	0.018
	(0.018)	(0.016)	(0.018)	(0.018)	(0.020)	(0.021)	(0.018)	(0.017)
Physical limitations	-0.046	-0.122	-0.016	0.061	-0.078	-0.052	0.079	-0.004
	(0.056)	(0.063)	(0.077)	(0.084)	(0.075)	(0.081)	(0.088)	(0.073)
Internet access								
WiFi access	-0.139	-0.054	-0.162	-0.083	-0.173	-0.142	-0.172	-0.097
	(0.093)	(0.078)	(0.095)	(0.092)	(0.116)	(0.100)	(0.098)	(0.086)
Pay-as-you-go plan	0.058	0.044	0.017	-0.002	0.031	-0.032	0.052	0.055
	(0.042)	(0.045)	(0.048)	(0.047)	(0.042)	(0.058)	(0.049)	(0.052)
Time constraints	-0.062	-0.019	-0.013	-0.048	0.025	0.002	-0.049	-0.015
	(0.034)	(0.035)	(0.037)	(0.034)	(0.039)	(0.037)	(0.038)	(0.034)
Security concerns	-0.142***	-0.138***	-0.114***	-0.160***	-0.171***	-0.124***	-0.197***	-0.172***
	(0.013)	(0.014)	(0.015)	(0.014)	(0.015)	(0.013)	(0.015)	(0.015)
Motivation								
Item-nonresponse	-0.035	-0.039	0.015	-0.063	-0.036	-0.017	-0.033	0.001
	(0.032)	(0.033)	(0.028)	(0.027)	(0.031)	(0.031)	(0.029)	(0.032)
Consent to data linkage	0.049	0.069	0.033	0.020	0.017	0.048	-0.013	0.064

	(0.029)	(0.032)	(0.029)	(0.027)	(0.036)	(0.033)	(0.028)	(0.030)
Mode of data collection:	0.063	0.067	0.041	-0.023	-0.003	0.057	-0.013	0.016
Web	(0.034)	(0.031)	(0.028)	(0.034)	(0.027)	(0.031)	(0.028)	(0.033)
Number of eligible waves								
1-3	0.025	0.055	0.082	0.010	0.026	0.070	0.027	0.034
	(0.037)	(0.031)	(0.039)	(0.041)	(0.042)	(0.039)	(0.033)	(0.032)
4-6	0.056	0.039	0.032	0.005	-0.008	0.024	0.056	0.035
	(0.041)	(0.037)	(0.046)	(0.044)	(0.047)	(0.044)	(0.043)	(0.040)
7-9	*-Baseline-*	*-Baseline-*	*-Baseline-*	*-Baseline-*	*-Baseline-*	*-Baseline-*	*-Baseline-*	*-Baseline-*
Proportion of full interviews	-0.006	-0.124	-0.054	-0.079	-0.003	0.035	-0.334	-0.216
	(0.093)	(0.100)	(0.093)	(0.097)	(0.095)	(0.105)	(0.092)	(0.084)
Socio-demographics								
Female	0.026	0.033	0.068	-0.055	0.063	0.033	-0.039	0.019
	(0.036)	(0.025)	(0.027)	(0.030)	(0.028)	(0.033)	(0.029)	(0.026)
Age	0.013	0.006	0.002	0.004	0.016	0.014	0.011	0.020
	(0.007)	(0.007)	(0.008)	(0.008)	(0.007)	(0.008)	(0.006)	(0.006)
Age-squared	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Education								
Higher degree	0.066	0.078	0.081	0.078	0.095	0.042	0.064	0.143
	(0.053)	(0.056)	(0.070)	(0.055)	(0.070)	(0.060)	(0.068)	(0.042)
A-level	0.102	0.094	0.149	0.059	0.126	0.094	0.021	0.204**
	(0.056)	(0.052)	(0.056)	(0.060)	(0.066)	(0.061)	(0.067)	(0.043)
GCSE	0.068	0.060	0.091	0.054	0.111	0.064	0.021	0.166
	(0.056)	(0.052)	(0.070)	(0.058)	(0.077)	(0.068)	(0.065)	(0.045)

No qualification	*-Baseline-*	*-Baseline-*	*-Baseline-*	*-Baseline-*	*-Baseline-*	*-Baseline-*	*-Baseline-*	*-Baseline-*
In work	-0.032	-0.007	-0.033	-0.049	-0.110	-0.066	-0.071	-0.126
	(0.048)	(0.038)	(0.045)	(0.044)	(0.040)	(0.050)	(0.043)	(0.038)
Income (ln)	0.003	-0.003	0.005	0.012	-0.006	-0.007	-0.006	-0.003
	(0.010)	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.008)	(0.008)
Own house	0.018	0.022	-0.006	0.014	0.024	-0.002	-0.047	-0.095
	(0.037)	(0.039)	(0.036)	(0.033)	(0.037)	(0.045)	(0.040)	(0.039)
N	1,317	1,317	1,317	1,316	1,316	1,317	1,316	1,315

Note. * p<0.05, ** p<0.01, *** p<0.001. P-values were adjusted using the Holm-Bonferroni method. Standard errors in parentheses.

N = 58 respondents had missing values in at least one of the predictor variables and were dropped from the analysis using listwise deletion.

Intensity of smartphone use, one of our indicators of device familiarity, is predictive of willingness for six of the data collection tasks. Respondents who use their smartphone more intensively, measured by the number of activities they carry out on their phone, are significantly more willing to allow the accelerometer to measure their physical activity, to complete a web survey in a mobile browser or in a survey app, to connect their smartphone to other devices via Bluetooth, to share the GPS position of their smartphone, and to use an app that tracks how they use their device. The effect has a similar magnitude across tasks: for every additional activity that respondents do on their smartphone, they have a 1.5 to 3.6 percentage point higher predicted probability of being willing to engage in mobile data collection. The other two indicators of device familiarity, ***frequency of smartphone use*** and ***self-rated skill using a smartphone***, however, do not have a significant effect on willingness in the multivariate models, despite having a significant bivariate relationship with willingness. When controlling for other characteristics, respondents who use their smartphone every day and who consider themselves proficient smartphone users are no more willing to participate in mobile data collection than those who use their smartphone less frequently and have lower self-rated skills.

The level of ***security concerns*** about mobile technologies is a second factor which is predictive of willingness to participate in mobile data collection. The more concerned respondents are about the security of providing information via mobile technologies, the less willing they are to complete each of the possible data collection tasks. The magnitude of the effect varies depending on the type of technology involved: it is larger for activities that are potentially threating to the respondent's privacy. Respondents with greater security concerns have a 19.7 percentage point lower predicted probability to be willing to share the

GPS location of their phone, but only an 11.4 percentage point lower predicted probability to be willing to complete an online questionnaire in a mobile browser.

In the multivariate models, we do not find a significant effect of ***physical limitations*** on willingness for any of the data collection tasks, presumably because we control for age. Respondents with physical limitations do not report lower levels of willingness compared to those without these limitations. We also do not find a significant effect of type of Internet access on willingness for any of the data collection tasks, similarly to the bivariate analysis: respondents without ***WiFi access*** at home and those with a ***pay-as-you-go plan*** are as willing to participate in mobile data collection as respondents with WiFi access or a fixed data plan. ***Time constraints*** are also not associated with willingness for any of the data collection tasks: respondents who have long working or commuting hours, children under the age of five or other caring responsibilities are not less willing to participate in mobile data collection compared to those without these time constraints.

Our indicators of respondent motivation and engagement with the study are also not predictive of willingness in the multivariate model. Respondents with ***item-nonresponse*** in the Innovation Panel questionnaire and those who ***gave consent to data linkage*** have similar levels of willingness to participate in mobile data collection as respondents without item-nonresponse and those who did not give consent. Similarly, time in panel, measured by the ***number of eligible waves***, is not predictive of willingness for any of the tasks. Respondents who were sampled longer ago and are still in the panel are as willing to participate in mobile data collection as panel members who were sampled more recently, when controlling for other respondent characteristics. We also do not find a significant association in the multivariate model between willingness and ***proportion of full interviews*** among eligible waves: panel members who have been cooperative in past waves are equally willing to complete additional data collection requests as members who have been less cooperative. We also tested the interaction between the number of eligible waves and the proportion of full interviews: respondents who were eligible for a larger number of waves and completed all interviews might have more experience with the survey compared to respondents who were eligible for fewer waves and completed all interviews. The interaction effect, however, is not significant (analysis not shown).

RQ4. Which task characteristics predict stated willingness, and does the effect depend on respondent characteristics?

To examine which task characteristics are associated with the varying levels of willingness that we observe across data collection tasks, we fitted multilevel logistic regression models predicting willingness to use mobile technologies on a smartphone. We used the dichotomised willingness as dependent variable to match the analysis for Research Question 3 and included random intercepts for each respondent. Given the small number of data collection tasks that we examined, we have limited variation in characteristics across tasks. We also ran models using the individual tasks as predictors of willingness. As will be shown in this section, however, the analysis of task characteristics reveals determinants of willingness that cannot be identified just by comparing the tasks.

Table 3.4 shows the average marginal effects of two multilevel logistic regression models: in the first model, we only include task characteristics as covariates; in the second model, we include task characteristics and respondent characteristics. On average across all data collection tasks, we find that 48.1 percent of respondents would be willing to participate in mobile data collection (n = 10,539).

Table 3.4. Multilevel logistic regression models predicting willingness to complete data collection tasks on a smartphone. Average marginal effects.

	Model 1		Model 2	
	AME	SE	AME	SE
Task characteristics				
App download required	-0.071***	0.014	-0.058***	0.012
Active role of respondent	0.075***	0.014	0.063***	0.012
High technical demands	0.211***	0.016	0.175***	0.014
Potential privacy threat	-0.314***	0.011	-0.258***	0.010
Device familiarity				
Use smartphone every day			0.014	0.020
Number of activities on smartphone			0.030***	0.003
Self-rated skill			0.036***	0.008
Physical limitations			-0.034	0.032
Internet access				
WiFi access			-0.156***	0.043
Fixed data plan			0.025	0.021
Time constraints			-0.013	0.015
Security concerns			-0.163***	0.006
Motivation				
Item-nonresponse			-0.024	0.013
Consent to data linkage			0.028*	0.013
Mode of data collection: Web			0.026	0.013
Number of eligible waves				
1-3			0.036*	0.015
4-6			0.029	0.016
7-9			*-Baseline-*	*-Baseline-*
Proportion of full interviews			-0.101*	0.040
Socio-demographics				
Female			0.024	0.013
Age			0.010***	0.003
Age-squared			-0.00012***	0.00003
Education				
Higher degree			0.086**	0.027
A-level			0.117***	0.028
GCSE			0.092***	0.028
No qualification				
In work			-0.064***	0.018
Income (ln)			-0.002	0.004
Own house			-0.013	0.016

Random-effects parameters		
Respondent variance	2.149	1.652
ICC	0.395	0.334

Note. * p<0.05, ** p<0.01, *** p<0.001.
Responses = 10,531 and respondents = 1,317. ICC = intra-class correlation.
N = 58 respondents had missing values in at least one of the predictor variables and were dropped from the analysis using listwise deletion.

In the first model, we find that all four task characteristics are significant predictors of willingness to participate in mobile data collection. Respondents have a 7.1 percentage point lower predicted probability of willingness to participate in tasks that ***require downloading and installing an app*** on their smartphone compared to tasks without this requirement. This result supports our expectation that downloading and installing an app is a potential barrier to participation. Data collection tasks that ***actively involve respondents in the data collection process*** have higher levels of willingness than passive tasks: respondents have a 7.5 percentage point higher predicted probability to report that they are willing to participate in active tasks compared to passive tasks, presumably because they have more control over the content of the data if they are actively involved in the data collection process. Surprisingly, respondents are more willing to complete tasks that have relatively high ***technical demands***, such as those requiring a lot of battery power or storage capacity, compared to tasks with relatively low technical demands: they have a 21.1 percentage point higher predicted probability of willingness to complete more technically demanding tasks than those with relatively low demands. This effect might be driven by other aspects of the tasks: albeit technically demanding, the tasks might be frequently used by respondents (e.g., the smartphone camera), and might have higher acceptance levels than tasks that have low technical demands but are rarely used by respondents. Finally, we find that tasks that are ***potentially threatening to the respondent's privacy*** have lower levels of willingness, which confirms our expectation that a potential privacy threat might represent a possible barrier to participation. Respondents have a 31.4 percentage point lower predicted probability of willingness to complete tasks that potentially threaten their privacy compared to tasks where the potential privacy threat is content-dependent. When we control for respondent characteristics in the second model, we find that the effect of each of the task characteristics remains significant, although the magnitude of the predicted probabilities decreases slightly.

Regarding respondent characteristics, the multilevel model confirms some findings of the task-specific models shown in Table 3.3: characteristics that have a significant effect on willingness in the task-specific models, including ***intensity of smartphone use*** and ***security concerns***, also have a significant effect in the multilevel model. There are, however, some differences. The multilevel model suggests that respondents with high ***self-rated skill*** using a smartphone are significantly more willing to participate in mobile data collection and those with ***WiFi access*** at home are significantly less willing to participate; neither of these variables significantly affects willingness in the task-specific models. Three of the motivation indicators, ***consent to data linkage***, ***number of eligible waves***, and ***proportion of full interviews***, also have a significant effect on willingness in the multilevel model, although not being predictive of willingness in the task specific models.

In addition to examining the main effect of task characteristics on willingness, we empirically tested the interactions of task characteristics and respondent characteristics that we proposed in our framework. Among all interaction effects that we specified in Figure 3.1, we only find significant interaction effects between frequency of smartphone use and task characteristics as well as between prior survey cooperativeness and task characteristics (analysis not shown).

For respondents who do not ***use their smartphone every day***, the requirement to ***download and install an app*** does not significantly affect their willingness to participate in mobile data collection (main effect: AME = +0.1 percentage points, p = 0.980). Respondents who use their smartphone every day, however, are significantly less willing to participate in mobile data collection compared to less frequent smartphone users if the task requires downloading and installing an app (interaction effect: AME = -6.9 percentage points, p = 0.002). Infrequent smartphone users have similar levels of willingness for both ***active and passive tasks*** (main effect: AME = +0.5 percentage points, p = 0.808), whereas respondents who use their smartphone every day are more willing to participate in mobile data collection than infrequent users if the task actively involves them in data collection (interaction effect: AME = +7.1 percentage points, p < 0.001).

We also find significant interaction effects between prior survey cooperativeness and three of the task characteristics. First, respondents who have been relatively uncooperative in previous survey waves, measured by a low ***proportion of waves*** in which they gave a full interview, are less willing to participate in ***active*** than in ***passive tasks*** (main effect: AME = -13.4 percentage points, p = 0.002). Those who have previously been more cooperative, however, are more willing to complete tasks where they are actively involved in data collection than less cooperative respondents (interaction effect: AME = +21.0 percentage points, p < 0.001). Second, we find that relatively uncooperative panel members are more willing to complete tasks with relatively high ***technical demands*** compared to tasks with lower demands (main effect: AME = +32.7 percentage points, p < 0.001). Those who have been cooperative, however, have lower levels of willingness for tasks that are technically demanding compared to uncooperative respondents (interaction effect: AME =-16.3 percentage points, $p < 0.001$). Third, the results suggest that relatively uncooperative panel members are willing to participate in mobile data collection independent of whether the task is ***intruding on their privacy*** (main effect: AME = -2.4 percentage points, $p = 0.619$). Cooperative respondents, however, are less willing to complete data collection tasks that are potentially threatening to their privacy compared to uncooperative respondents (interaction effect: AME = -25.1 percentage points, $p < 0.001$).

3.5. Discussion

In this paper, we examine the stated willingness of the general population to participate in mobile data collection tasks, using data from a nationally representative household panel study in Great Britain. We provide novel evidence on how stated willingness varies between eight different mobile data collection tasks and on how willingness varies between different mobile devices (smartphones and tablets). We also provide novel evidence on the relative importance of respondent characteristics, task characteristics, and their interactions, by proposing and testing a theoretical framework of the determinants of willingness to participate in different mobile data collection tasks.

We find that the level of stated willingness varies by data collection task and, to a lesser extent, by device. Respondents seem to make a clear distinction between different tasks: fewer people would be willing to share the GPS position of their mobile device than to take a photo for a survey or to complete a questionnaire in a mobile browser. More than half of respondents would not be at all willing to download an app which collects anonymous data about how they use their mobile device. These findings are consistent with previous results based on online access (volunteer) panels in other countries (Revilla et al., 2017; Revilla, Toninelli, et al., 2016). The majority of people who use both a smartphone and a tablet have consistent preferences: they are equally willing or equally unwilling to use either of their devices for data collection. For some respondents, the device type, however, makes a difference: a tablet would be the preferred device for completing an online questionnaire in a mobile browser or survey app, whereas a smartphone would be the preferred device for taking photos or for connecting to other devices via Bluetooth.

We also find that stated willingness varies with respondent characteristics: those who use their mobile device more intensively and have lower levels of security concerns are more willing to use mobile technologies for data collection. These findings are consistent with previous findings from access panels (Keusch et al., 2017; Pinter, 2015; Revilla et al., 2017). Other respondent characteristics that we examined do not significantly affect willingness.

The difference in stated willingness between different data collection tasks is related to the characteristics of the tasks: respondents are more willing to participate in tasks where they actively complete the measurements than in tasks where data are collected passively. This finding is consistent with previous results from an access panel in Spain, Portugal and Latin America (Revilla et al., 2017; Revilla, Toninelli, et al., 2016). In addition, we find that respondents are less willing to participate in tasks that require downloading an app and in tasks that measure highly private information. Somewhat surprisingly, respondents are more willing to participate in tasks that place higher technical demands (such as battery usage) on their devices; however, this may be an effect of the specific tasks we studied.

Finally, we find some evidence that the effect of task characteristics on stated willingness depends on respondent characteristics: for respondents who use their device every day, the requirement to download an app reduces willingness, while the requirement to actively complete the measurement increases willingness. For respondents who use their devices less frequently neither task characteristic affects stated willingness. This could be because frequent users are likely to have a larger number of apps and files stored on their device, and therefore less available storage space than infrequent users. Conversely, they are likely to be more confident in actively completing tasks using their device, and might find active completion less burdensome than infrequent users.

These findings suggest that willingness to participate in mobile data collection depends on the type of data that researchers want to collect as well as on characteristics of the population of interest that they want to study. Researchers who aim to implement mobile data collection in surveys might adjust the data collection request to the potential barriers of participation that the specific tasks entail. When asking respondents, for example, to complete data collection tasks that require downloading and installing an app on their mobile device, researchers might provide additional instructions or screenshots to respondents on how to access the app store and to download and install apps on their device. For data collection activities that are potentially intruding on the respondent's privacy, including sharing GPS coordinates, researchers might leverage data confidentiality and other data security aspects of the study as part of the data collection request.

In order to maximise participation rates in studies with mobile data collection, researchers might also consider tailoring data collection requests to respondents based on information available from a screening questionnaire. Respondents who have access to a mobile device but are not sufficiently familiar with using the device or use the device less intensively could be offered one-time support by an interviewer who helps them to install and use a data collection app, or could be provided with assistance during data collection, for example by setting up a support hotline. Respondents who report high levels of security concerns could receive invitation letters that contain more information about procedures to ensure data confidentiality. Those with lower levels of motivation and engagement with the study could receive motivational statements in the invitation letter which state the importance of the respondent's participation for the study or could be provided with higher levels of incentives, particularly in studies that ask respondents to share data from their accelerometer, to connect their mobile device to other devices via Bluetooth, or to use an app that tracks how they use their mobile device.

A limitation of our study is that we focused on a relatively small set of feasible mobile data collection tasks. While we classified the characteristics of these tasks a priori, we did not investigate the full set of potential tasks: we would need 32 (= 2^5) tasks to fully test our theoretical model with five task characteristics. We would be hard pressed to find realistic mobile data collection tasks to fit each of these cells. The aim of this paper, however, is to give researchers an idea which task characteristics to consider when examining willingness on a particular data collection task.

While this paper focuses on willingness to participate in mobile data collection generally, a potential avenue for further research is to examine compliance over time in repeated data collection tasks, and the factors that are associated with compliance. Respondents might be willing to engage in mobile data collection for one-off tasks but might drop out of tasks that are continuous or require repeated participation. In studies that track the GPS location of a smartphone, for example, respondents might decide to turn off the GPS function of their mobile device once they realise that GPS consumes a considerable amount of battery power. More research is also needed to further understand some of the findings of this paper. Further research could explore, for example, why frequent smartphone users appear less willing to participate in mobile data collection if the task requires downloading and installing an app, or why cooperative panel members appear less willing to complete some of the data collection tasks.

As survey researchers and others continue to find ways of exploiting the powerful mobile devices that many people carry around with them all day, we need to be mindful of what tasks people might be willing to do, and who might be willing to do what tasks. This paper begins to lay out the issues and provides initial empirical evidence on these important sources of variation in willingness to perform additional data collection tasks using these devices.

3.6. Appendix

RQ1. How does stated willingness to use mobile technologies vary across different data collection tasks?

Table 3.5. Stated willingness to complete data collection tasks on a smartphone (in percent).

	Very willing	Somewhat willing	A little willing	Not at all willing	Missing	Total
Camera	33.7	31.1	16.3	18.7	0.2	100.0
Accelerometer	32.2	28.7	15.2	23.7	0.2	100.0
Questionnaire	31.5	24.4	13.4	30.5	0.2	100.0
Bluetooth	28.0	27.8	16.3	27.5	0.3	100.0
Text message survey	23.0	27.1	21.1	28.6	0.2	100.0
Survey app	23.4	23.6	17.1	35.6	0.2	100.0
GPS	18.0	21.1	21.8	39.0	0.2	100.0
Tracking app	13.2	14.5	18.7	53.3	0.3	100.0

N = 1,379.

Table 3.6. Stated willingness to complete data collection tasks on a tablet (in percent).

	Very willing	Somewhat willing	A little willing	Not at all willing	Missing	Total
Questionnaire	38.5	25.9	13.2	22.0	0.5	100.0
Survey app	28.6	22.8	17.5	30.6	0.5	100.0
Camera	26.1	24.9	19.1	29.4	0.5	100.0
Bluetooth	23.2	18.3	19.9	38.0	0.5	100.0
Tracking app	16.3	14.3	18.0	50.9	0.5	100.0

N = 1,261.

RQ2. How does stated willingness to do different tasks vary between smartphone and tablet?

Table 3.7. Consistency of stated willingness among respondents with access to smartphone and tablet (in percent).

	Willing on both devices	Willing on smartphone	Willing on tablet	Not willing on either device	Missing	Total
Questionnaire	49.9	6.4	19.1	24.3	0.3	100.0
Survey app	41.3	7.5	16.2	34.8	0.3	100.0
Tracking app	24.3	3.9	11.1	60.3	0.4	100.0
Camera	48.3	18.9	6.7	25.9	0.3	100.0
Bluetooth	43.6	16.1	4.3	35.5	0.4	100.0

N = 980.

RQ3. Which respondent characteristics predict stated willingness to do different tasks?

Table 3.8. Bivariate relationship of respondent characteristics and willingness to participate in mobile data collection.

		% Willing	**Camera**	**Accelero-meter**	**Questionnaire**	**Bluetooth**	**Text messages**	**Survey app**	**GPS**	**Tracking app**
Device familiarity										
	Use smartphone every day	No	48.3	42.0	25.3	37.9	33.5	22.1	28.3	12.4
		Yes	68.4	65.0	62.4	59.8	53.7	52.3	41.4	30.8
		Prob>F	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.004*	*0.000*
	Number of activities on smartphone	0-2	31.1	18.6	4.6	5.7	21.5	1.0	9.7	1.2
		3-4	51.5	34.4	16.1	32.9	35.3	15.3	28.7	7.0
		5-6	52.1	40.6	31.7	32.2	39.2	24.7	26.7	10.6
		7-8	64.6	53.5	54.1	60.3	48.6	42.9	36.0	20.6
		9-10	70.8	70.3	68.0	58.1	53.6	56.7	39.5	34.8
		11-12	74.2	78.8	76.0	74.5	60.7	65.9	51.6	41.7
		Prob>F	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
	Self-reported skill	1 Beginner	17.0	13.3	7.6	6.4	6.0	8.4	8.2	0.0
		2	48.6	32.9	18.3	27.7	40.0	16.9	21.1	6.8
		3	62.2	53.0	42.9	46.9	50.3	31.5	36.1	20.0
		4	67.9	64.0	60.6	58.0	50.1	50.3	39.2	28.0
		5 Advanced	72.4	75.1	74.4	72.1	57.1	65.9	48.5	40.5
		Prob>F	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
Physical limitations		No	65.8	62.7	57.2	56.7	51.1	48.4	39.5	28.3

	Yes	44.5	26.5	31.5	40.7	31.6	20.6	32.3	13.4
	Prob>F	*0.021*	*0.000*	*0.006*	*0.091*	*0.019*	*0.000*	*0.355*	*0.030*
Internet access									
WiFi access	No	71.0	59.4	62.4	58.0	64.3	54.0	56.6	42.8
	Yes	64.7	61.1	55.8	55.9	49.8	46.9	38.6	27.2
	Prob>F	*0.570*	*0.868*	*0.595*	*0.849*	*0.244*	*0.550*	*0.098*	*0.117*
Pay-as-you-go plan	No	65.2	61.9	57.0	56.9	50.7	48.4	38.8	27.3
	Yes	61.8	55.0	49.2	49.2	46.3	37.9	42.1	31.2
	Prob>F	*0.428*	*0.213*	*0.150*	*0.151*	*0.281*	*0.091*	*0.545*	*0.400*
Time constraints	No	65.3	59.8	54.2	55.2	48.9	45.1	39.1	26.2
	Yes	63.8	64.0	60.6	58.0	53.5	52.1	39.2	31.3
	Prob>F	*0.701*	*0.326*	*0.122*	*0.445*	*0.226*	*0.085*	*0.986*	*0.217*
Security concerns	Not at all concerned	86.0	85.3	78.5	84.1	75.9	72.9	71.8	64.8
	A little …	78.6	75.5	70.7	70.0	63.4	58.7	51.5	35.3
	Somewhat …	55.2	51.2	46.0	45.6	40.9	38.8	23.8	14.1
	Very…	37.2	32.3	30.5	28.2	22.7	20.6	16.2	5.4
	Extremely…	31.4	18.7	11.7	9.1	7.9	7.4	4.3	2.2
	Prob>F	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
Motivation									
Item-nonresponse	No	70.2	66.7	57.5	62.9	55.7	51.1	44.2	30.7
	Yes	61.5	57.4	55.1	51.6	46.7	44.6	35.9	25.8
	Prob>F	*0.013*	*0.009*	*0.555*	*0.001*	*0.006*	*0.073*	*0.009*	*0.164*
Consent to data linkage	No	61.1	54.4	52.0	49.8	47.2	42.4	36.5	22.8
	Yes	67.8	65.5	58.8	59.9	52.7	50.6	40.9	31.1
	Prob>F	*0.031*	*0.003*	*0.056*	*0.002*	*0.151*	*0.026*	*0.170*	*0.010*

Number of eligible waves	1-3	65.0	65.2	63.2	59.0	52.7	53.3	40.2	30.0
	4-6	70.4	62.3	56.4	56.9	50.5	47.7	41.8	31.1
	7-9	60.2	55.0	47.0	51.6	47.0	39.2	35.6	22.1
	Prob>F	*0.110*	*0.044*	*0.004*	*0.234*	*0.406*	*0.018*	*0.415*	*0.118*
Proportion of full interviews	Less than 1	64.9	66.0	60.3	60.6	48.1	47.6	51.0	34.0
	1	64.9	59.5	54.7	54.6	50.9	47.0	35.5	25.8
	Prob>F	*0.992*	*0.079*	*0.149*	*0.106*	*0.492*	*0.887*	*0.000*	*0.034*

Table 3.9. Questionnaire.

	Variable	Question wording
Q1	Access to mobile technologies	Which of the following devices do you use to connect to the Internet? *Desktop computer; Laptop; Smartphone; Tablet; Feature phone / non-touchscreen mobile phone; E-book reader (e.g., Kindle); Smartwatch; Other*
Q2	WiFi access	Do you have WiFi access at home? *Yes; No*
Q3	Type of smartphone contract	Do you have a fixed data plan or a pay-as-you-go contract to get mobile Internet on your smartphone? *Fixed data plan; Pay-as-you-go contract; No fixed data plan or pay-as-you-go contract (use WiFi only)*
Q4	Frequency of mobile device use	How often do you use a smartphone for activities other than phone calls or text messaging? *Every day; Several times a week; Several times a month; Once a month or less*
		How often do you use a tablet? *Every day; Several times a week; Several times a month; Once a month or less*
Q5	Activities carried out on mobile devices	Do you use your smartphone for the following activities? *Yes; No* Browsing websites Email Taking photos Looking at content on social media websites/apps (e.g., looking at text, images, videos on Facebook, Twitter, Instagram) Posting content to social media websites/apps (e.g., posting text, images, videos on Facebook, Twitter, Instagram) Making purchases (e.g., booking train tickets, buying clothes, ordering food)

	Online banking (e.g., checking account balance, transferring money) Installing new apps (e.g., from iTunes, Google Play Store) Using GPS/location-aware apps (e.g., Google Maps, Foursquare, Yelp) Connecting to other electronic devices via Bluetooth (e.g., smartwatches, bathroom scales) Playing games Streaming videos or music Other
	Do you use your tablet for the following activities? *Yes; No* Browsing websites Email Taking photos Looking at content on social media websites/apps (e.g., looking at text, images, videos on Facebook, Twitter, Instagram) Posting content to social media websites/apps (e.g., posting text, images, videos on Facebook, Twitter, Instagram) Making purchases (e.g., booking train tickets, buying clothes, ordering food) Online banking (e.g., checking account balance, transferring money) Installing new apps (e.g., from iTunes, Google Play Store) Using GPS/location-aware apps (e.g., Google Maps, Foursquare, Yelp) Connecting to other electronic devices via Bluetooth (e.g., smartwatches, bathroom scales) Playing games Streaming videos or music Other

Q6	Self-reported level of skill	Generally, how would you rate your skills of using a smartphone on a scale from 1 = Beginner to 5 = Advanced? *1 Beginner; 2; 3; 4; 5 Advanced*
		Generally, how would you rate your skills of using a tablet on a scale from 1 = Beginner to 5 = Advanced? *1 Beginner; 2; 3; 4; 5 Advanced*
Q7	Willingness to participate in mobile data collection	How willing would you be to carry out the following tasks on your smartphone for a survey? *Very willing; Somewhat willing; A little willing; Not at all willing* Complete an online questionnaire on your mobile phone. Download a survey app to complete an online questionnaire. Download an app which collects anonymous data about how you use your smartphone. Answer a couple of questions sent via text messaging. Use the camera of your smartphone to take photos or scan barcodes. Allow built-in features of your smartphone to measure the frequency and speed at which you walk, run or cycle. Share the GPS position of your smartphone. Connect your smartphone via Bluetooth to other electronic devices (e.g., wearables such as Fitbit).
		How willing would you be to carry out the following tasks on your tablet for a survey? *Very willing; Somewhat willing; A little willing; Not at all willing* Complete an online questionnaire on your tablet. Download a survey app to complete an online questionnaire. Download an app which collects anonymous data about how you use your tablet. Use the camera of your tablet to take photos or scan barcodes. Connect your tablet via Bluetooth to other electronic devices (e.g., wearables such as Fitbit).

Q8	Security concerns	In general, how concerned would you be about the security of providing information in the following ways? *Not at all concerned; A little concerned; Somewhat concerned; Very concerned; Extremely concerned* Complete an online questionnaire in your mobile browser. Download a survey app to complete an online questionnaire. Download an app which collects anonymous data about how you use your [smartphone/tablet/smartphone or tablet]. Answer a couple of questions sent via text messaging. Use the camera of your [smartphone/tablet/smartphone or tablet] to take photos or scan barcodes. Allow built-in features of your smartphone to measure the frequency and speed at which you walk, run or cycle. Share the GPS position of your smartphone. Connect your [smartphone/tablet/smartphone or tablet] via Bluetooth to other electronic devices (e.g., wearables such as Fitbit).

Conclusion

In this thesis, I examined two potential sources of error in mobile survey data collection: measurement error and nonresponse. While Chapter 1 and 2 studied how two features of mobile data collection, the small screen size of mobile devices and the potentially more distracting environment of mobile device users, affect measurement error, Chapter 3 investigated potential barriers to participation in mobile data collection tasks that might affect nonresponse error.

I hope that the research presented in the three chapters will contribute to the growing survey methodological literature examining error properties in mobile data collection. Chapter 1 extends earlier research on measurement error in mobile web surveys by comparing devices with different screen size rather than categories of devices. Chapter 2 is the first study, to the best of my knowledge, that investigates the impact of distractions on measurement error in surveys using a laboratory experimental setup. Chapter 3 proposes a theoretical framework of how respondent characteristics, characteristics of the data collection task, and task-respondent interactions might affect willingness to participate in mobile data collection; the chapter also tests the framework by using data from a general population survey, thereby extending earlier research that focused on non-probability samples.

To summarise, these are the main empirical findings of this research:

- Survey completion on small smartphones with a screen size of below 4.0 inches is detrimental to data quality if the questionnaire is not mobile-optimised (Chapter 1).
- Respondents who listen to music during survey completion or are in noisy environments where other people have a loud conversation do not provide data of lower quality (Chapter 2).

- Security concerns and low levels of device familiarity are the main reasons why respondents would not be willing to take part in mobile data collection (Chapter 3).
- Respondents are more willing to participate in data collection tasks where they are actively involved in the data collection process. They are less willing to complete tasks that require downloading an app or that measure highly private information (Chapter 3).

This thesis not only aims to contribute to the survey methodological literature, but also to help inform decisions that survey practitioners have to make when designing and implementing studies using mobile data collection. Based on the findings of Chapter 1, I would recommend survey managers to develop mobile-optimised questionnaires that adapt to small screens. Mobile optimisation seems to be particularly important for surveys that contain a considerable amount of check-all-that-apply questions questions or open-ended questions. I also hope that survey managers might find the theoretical framework and the results of Chapter 3 useful when deciding about the implementation of mobile data collection tasks. They might adjust their data collection request to the potential barriers of participation that the particular tasks entail, or tailor their request to characteristics of the respondent. The framework might also guide them when choosing between alternative designs. For example, they might consider administering mobile surveys in a browser rather than in a survey app since downloading and installing an app might be a potential barrier to participation.

There are many more challenges related to mobile data collection that need to be addressed. Three areas in particular might require further investigation. First, more research is needed on the scalability of mobile data collection to the general population (Couper et al., 2017). Most of the existing studies test the feasibility of mobile data collection on small samples of volunteers, but the question remains whether research involving mobile technologies can be extended to the general population and what might be potential barriers among different population subgroups. Second, survey methodologists need to get a better understanding of how to increase participation in mobile data collection, in one-off studies as well as in longitudinal studies that require the respondent's compliance over time. In addition to monetary incentives, future research might investigate the effectiveness of gamification approaches (Keusch & Zhang, 2015) or of providing feedback to participants based on the data that they provided. Third, the survey profession needs to comprehend the legal and ethical issues that arise with mobile data collection (Couper et al., 2017; Link et al., 2014), in particular with technologies that collect data passively on a large scale. Regardless of the "privacy paradox", according to which technology users report privacy concerns but do, in fact, very little to protect their privacy (Barth & de Jong, 2017), researchers need to develop careful measures to protect data security and privacy of their respondents when collecting data with mobile technologies.

Although this thesis was concerned with potential sources of error in mobile survey data collection, I want to conclude with a positive outlook on the future. Despite the methodological challenges that mobile data collection entails, many of which are yet to be addressed, mobile technologies are on the way to become promising data collection tools for social science research. They not only allow administering surveys in innovative ways, but also enable researchers to collect new forms of data that can supplement or potentially replace data collected with survey-based approaches. Technology evolves quickly and the potential measurement opportunities of mobile technologies and their use for social science research are likely to expand in the future. As more and more people use mobile devices, including older subgroups of the population, and these devices become an integral part of their daily life, the willingness to take part in mobile data collection is likely to increase in the general population. And as research as well as best practices in this area are emerging, the survey profession is likely to embrace the opportunities of mobile technologies and to adopt new mobile-based approaches of data collection. We live in an exciting time for survey methodology indeed.

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*(5), 676–703.

Anderson, M. (2015). Technology Device Ownership: 2015. Retrieved from http://www.pewinternet.org/2015/10/29/technology-device-ownership-2015

Ansolabehere, S., & Schaffner, B. F. (2015). Distractions: the incidence and consequences of interruptions for survey respondents. *Journal of Survey Statistics and Methodology*, *3*(2), 216–239.

Antoun, C. (2015). *Mobile web surveys: a first look at measurement, nonresponse, and coverage errors*. University of Michigan.

Antoun, C., Couper, M. P., & Conrad, F. G. (2017). Effects of mobile versus PC web on survey response quality: a crossover experiment in a probability web panel. *Public Opinion Quarterly*, 280–306.

Baker, R., Brick, J. M., Bates, N. A., Battaglia, M., Couper, M. P., Dever, J. A., … Tourangeau, R. (2013). Summary report of the AAPOR task force on non-probability sampling. *Journal of Survey Statistics and Methodology*, *1*(2), 90–105.

Barge, S., & Gehlbach, H. (2012). Using the theory of satisficing to evaluate the quality of survey data. *Research in Higher Education*, *53*(2), 182–200.

Barth, S., & de Jong, M. D. T. (2017). The privacy paradox - investigating discrepancies between expressed privacy concerns and actual online behavior - a systematic literature review. *Telematics and Informatics*, *34*(7), 1038–1058.

Belsham, R. L., & Harman, D. W. (1977). Effect of vocal vs non-vocal music on visual recall. *Perceptual and Motor Skills*, *44*(3), 857–858.

Berinsky, A. J., Margolis, M. F., & Sances, M. W. (2014). Separating the shirkers from the workers? Making sure respondents pay attention on self-administered surveys. *American Journal of Political Science*, *58*(3), 739–753.

Blom, A. G., Bosnjak, M., Cornilleau, A., Cousteaux, A.-S., Das, M., Douhou, S., & Krieger, U. (2015). A comparison of four probability-based online and mixed-mode panels in Europe. *Social Science Computer Review*, 1–18.

Bosnjak, M., Poggio, T., Becker, K. R., Funke, F., Wachenfeld, A., & Fischer, B. (2013). *Online survey participation via mobile devices*. Paper presented at the AAPOR Annual Conference, Boston, MA.

Burton, J. (2016). *Results for web/face-to-face linkage consent questions in the Innovation Panel*. Paper presented at the CLOSER Mixing Modes and Measurement Methods in Longitudinal Studies Workshop, London.

Buskirk, T. D., & Andrus, C. (2012). Smart surveys for smart phones: exploring various approaches for conducting online mobile surveys via smartphones. *Survey Practice*, *5*(1), 1–11.

Buskirk, T. D., & Andrus, C. H. (2014). Making mobile browser surveys smarter: results from a randomized experiment comparing online surveys completed via computer or smartphone. *Field Methods*, *26*(4), 322–342.

Cabinet Office. (2014). *Community Life Survey: Technical Report 2013-2014.* Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/364511/Community_life_2013-14_Technical_report.pdf

Cacioppo, J. T., Petty, R. E., & Morris, K. (1983). Effects of need for cognition on message evaluation, recall, and persuasion. *Journal of Personality and Social Psychology*, *45*(4), 805–818.

Callegaro, M. (2010). Do you know which device your respondent has used to take your online survey? *Survey Practice*, *3*(6), 1–12.

Callegaro, M., Lozar Manfreda, K., & Vehovar, V. (2015). *Web Survey Methodology*. London: SAGE Publications.

Cernat, A. (2015). Using equivalence testing to disentangle selection and measurement in mixed modes surveys. *Understanding Society Working Paper Series*, *2015–1*, 1–13.

Chin, E., Felt, A. P., Sekar, V., & Wagner, D. (2012). Measuring user confidence in smartphone security and privacy. *Proceedings of the Eighth Symposium on Usable Privacy and Security*.

Couper, M. P. (2008). *Designing Effective Web Surveys*. New York, NY: Cambridge University Press.

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

Couper, M. P., Antoun, C., & Mavletova, A. (2017). Mobile web surveys: a total survey error perspective. In P. Biemer, S. Eckman, B. Edwards, E. de Leeuw, F. Kreuter, L. Lyberg, … B. West (Eds.), *Total Survey Error in Practice* (pp. 133–154). New York: Wiley.

Couper, M. P., & Peterson, G. J. (2016). Why do web surveys take longer on smartphones? *Social Science Computer Review*, 1–21.

Couper, M. P., Singer, E., & Tourangeau, R. (2003). Understanding the effects of audio-CASI on self-reports of sensitive behavior. *Public Opinion Quarterly*, *67*, 385–395.

Couper, M. P., Tourangeau, R., Conrad, F., & Crawford, S. D. (2004). What they see is what we get. Response options for web surveys. *Social Science Computer Review*, *22*(1), 111–127.

de Bruijne, M., & Wijnant, A. (2013). 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*(4), 482–504.

de Bruijne, M., & Wijnant, A. (2014). Mobile response in web panels. *Social Science Computer Review*, *32*(6), 728–742.

Firtman, M. (2010). *Programming the Mobile Web*. O'Reilly.

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

Furnham, A., & Bradley, A. (1997). Music while you work: the differential distraction of background music on the cognitive test performance of introverts and extraverts. *Applied Cognitive Psychology*, *11*(5), 445–455.

Furnham, A., Trew, S., & Sneade, I. (1999). The distracting effects of vocal and instrumental music on the cognitive test performance of introverts and extraverts. *Personality and Individual Differences*, *27*(2), 381–392.

Galesic, M. (2006). Dropouts on the web: effects of interest and burden experienced during an online survey. *Journal of Official Statistics*, *22*(2), 313–328.

Galesic, M., Tourangeau, R., Couper, M. P., & Conrad, F. (2008). Eye-tracking data: new insights on response order effects and other cognitive shortcuts in survey responding. *Public Opinion Quarterly*, *72*, 892–913.

Geurs, K., Veenstra, S., & Thomas, T. (2013). *The setup of a mobile mobility panel for the Netherlands*. Paper presented at the World Conference on Transport Research, Rio de Janeiro, Brazil.

Ginnis, S. (2017). *Mobile-based geo-triggered surveys: experiences from the field*. Paper presented at the CLOSER Workshop "New technologies to measure non-health topics in longitudinal surveys", London, UK.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-

Five personality domains. *Journal of Research in Psychology*, *37*, 504–528.

Greenfield, T. K., Bond, J., & Kerr, W. C. (2014). Biomonitoring for improving alcohol consumption surveys: the new gold standard? *Alcohol Research: Current Reviews*, *36*(1), 39–45.

Groves, R. M. (2011). Three eras of survey research. *Public Opinion Quarterly*, *75*(5), 861–871.

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

Guidry, K. R. (2012). *Response quality and demographic characteristics of respondents using a mobile device on a web-based survey*. Paper presented at the AAPOR Annual Conference, Orlando, FL.

Gummer, T., & Roßmann, J. (2015). Explaining interview duration in web surveys: a multilevel approach. *Social Science Computer Review*, *33*(2), 217–234.

Hamlyn, B., Fitzpatrick, A., & Williams, J. (2015). *Investigating the viability of moving from a face-to-face to an online/postal mode: evidence from a series of methodological studies 2012-2015*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/466921/Investigating_the_the_viability_of_moving_from_a_face-to-face_to_an_online_postal_mode_FINAL.pdf

Hanson, T., Matthews, P., & McGee, A. (2015). *Understanding Society Innovation Panel Wave 7. Technical Report.* Retrieved from https://www.understandingsociety.ac.uk/d/196/IP7_TechReport_v4.pdf

Hays, R. D., Liu, H., & Kapteyn, A. (2015). Use of Internet panels to conduct surveys. *Behavior Research Methods*, *47*(3), 685–690.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian*

Journal of Statistics, *6*(2), 65–70.

Hox, J. J., De Leeuw, E. D., & Zijlmans, E. a O. (2015). Measurement equivalence in mixed mode surveys. *Frontiers in Psychology*, *6*(FEB), 87.

Jäckle, A., Burton, J., Couper, M. P., & Lessof, C. (2017). Participation in a mobile app survey to collect expenditure data as part of a large-scale probability household panel: response rates and response biases. *Unpublished Manuscript*.

Jäckle, A., Gaia, A., Al Baghal, T., Burton, J., & Lynn, P. (2017). Understanding Society The UK Household Longitudinal Study Innovation Panel, Waves 1-9, User Manual. Colchester: University of Essex. Retrieved from https://www.understandingsociety.ac.uk/documentation/innovation-panel

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

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

Japec, L., Kreuter, F., Berg, M., Biemer, P., Decker, P., Lampe, C., … Usher, A. (2015). Big data in survey research: AAPOR task force report. *Public Opinion Quarterly*, *79*(4), 839–880.

Jerit, J., Barabas, J., & Clifford, S. (2013). Comparing contemporaneous laboratory and field experiments on media effects. *Public Opinion Quarterly*, *77*(1), 256–282.

Jones, M. S., House, L. A., & Gao, Z. (2015). Respondent screening and revealed preference axioms: testing quarantining methods for enhanced data quality in web panel surveys. *Public Opinion Quarterly*, *79*(3), 687–709.

Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kaminska, O., McCutcheon, A. L., & Billiet, J. (2010). Satisficing among reluctant respondents in a cross-national context. *Public Opinion Quarterly*, *74*(5), 956–984.

Kellogg, R. (2012). *Fundamentals of Cognitive Psychology*. Los Angeles: SAGE Publications.

Kennedy, C. (2010). *Nonresponse and measurement error in mobile phone surveys*. University of Michigan.

Kennedy, C., & Everett, S. E. (2011). Use of cognitive shortcuts in landline and cell phone surveys. *Public Opinion Quarterly*, *75*, 336–348.

Keusch, F., Antoun, C., Couper, M. P., Kreuter, F., & Struminskaya, B. (2017). *Willingness to participate in passive mobile data collection*. Paper presented at the AAPOR Conference, New Orleans, LA.

Keusch, F., & Zhang, C. (2015). A review of issues in gamified surveys. *Social Science Computer Review*, 1–20.

Kiger, D. M. (1989). Effects of music information load on a reading comprehension task. *Perceptual and Motor Skills*, *69*(2), 531–534.

Kooreman, P., & Scherpenzeel, A. (2014). High frequency body mass measurement, feedback, and health behaviors. *Economics and Human Biology*, *14*(1), 141–153.

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. (1987). An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly*, *51*, 201–219.

Krosnick, J. A., Narayan, S. S., & Smith, W. R. (1996). Satisficing in surveys: initial evidence. In M. T. Braverman & J. K. Slater (Eds.), *Advances in Survey Research* (pp. 29–44). San Francisco: Jossey-Bass.

Lathia, N., Sandstrom, G. M., Mascolo, C., & Rentfrow, P. J. (2017). Happier people live more active lives: using smartphones to link happiness and physical activity. *PLoS*

ONE, *12*(1), 1–13.

Lavrakas, P. J., Tompson, T. N., & Benford, R. (2010). *Investigating data quality in cell phone surveying*. Paper presented at the AAPOR Conference, Chicago, IL.

Lenzner, T., Kaczmirek, L., & Lenzner, A. (2010). Cognitive burden of survey questions and response times: a psycholinguistic experiment. *Applied Cognitive Psychology*, *24*, 1003–1020.

Link, M. W., Murphy, J., Schober, M. E., Buskirk, T. D., Hunter Childs, J., & Langer Tesfaye, C. (2014). *Mobile Technologies for Conducting, Augmenting and Potentially Replacing Surveys: Report of the AAPOR Task Force on Emerging Technologies in Public Opinion Research*. Retrieved from https://www.aapor.org/getattachment/Education-Resources/Reports/REVISED_Mobile_Technology_Report_Final_revised10June14.pdf.aspx

Lugtig, P., Lensvelt-Mulders, G. J., Frerichs, R., & Greven, A. (2011). Estimating nonresponse bias and mode effects in a mixed-mode survey. *International Journal of Market Research*, *53*(5), 669–686.

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*, 1–17.

Lynn, P. (2009). Sample design for Understanding Society. *Understanding Society Working Paper*, *2009–1*.

Lynn, P., & Kaminska, O. (2012). The impact of mobile phones on survey measurement error. *Public Opinion Quarterly*, *77*(2), 586–605.

Mavletova, A. (2013). Data quality in PC and mobile web surveys. *Social Science Computer Review*, *31*(6), 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*(3), 191–205.

McClain, C., Crawford, S. D., & Dugan, J. P. (2012). *Use of mobile devices to access computer-optimized web instruments: implications for respondent behavior and data quality*. Paper presented at the AAPOR Annual Conference, Orlando, FL.

McGaughey, R. E., Zeltmann, S. M., & McMurtrey, M. E. (2013). Motivations and obstacles to smartphone use by the elderly: developing a research framework. *International Journal of Electronic Finance*, *7*(3/4), 177.

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: part 2. Accounts of psychological refractory-period phenomena. *Psychological Review*, *104*(4), 749–791.

Millar, M. M., & Dillman, D. a. (2012). Encouraging survey response via smartphones: Effects on respondents' use of mobile devices and survey response rates. *Survey Practice*, *5*(3), 1–6.

Moskowitz, D. S., & Young, S. N. (2006). Ecological momentary assessment: what it is and why it is a method of the future in clinical psychopharmacology. *Journal of Psychiatry & Neuroscience*, *31*(1), 13–20.

Ofcom. (2017). *The Communications Market Report*. Retrieved from https://www.ofcom.org.uk/__data/assets/pdf_file/0017/105074/cmr-2017-uk.pdf

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*(4), 867–872.

Pashler, H. (1994). Dual-task interference in simple tasks: data and theory. *Psychological Bulletin*, *116*(2), 220–244.

Peterson, G. (2012). *Unintended mobile respondents*. Paper presented at the CASRO

Technology Conference, New York, NY.

Peytchev, A. (2009). Survey breakoff. *Public Opinion Quarterly*, *73*(1), 74–97.

Peytchev, A., & Hill, C. A. (2010). Experiments in mobile web survey design. Similarities to other modes and unique considerations. *Social Science Computer Review*, *28*(3), 319–335.

Pinter, R. (2015). Willingness of online access panel members to participate in smartphone application-based research. In D. Toninelli, R. Pinter, & P. de Pedraza (Eds.), *Mobile Research Methods: Opportunities and Challenges of Mobile Research Methodologies* (pp. 141–156). London: Ubiquity Press.

Poggio, T., Bosnjak, M., & Weyandt, K. (2015). Survey participation via mobile devices in a probability-based online-panel: prevalence, determinants, and implications for nonresponse. *Survey Practice*, *8*(1), 1–7.

Poushter, J. (2016). Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies. Retrieved from http://www.pewglobal.org/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/

Revilla, M., Couper, M. P., & Ochoa, C. (2017). *Willingness of online panelists to perform additional tasks*. Paper presented at the GOR Conference, Berlin, Germany.

Revilla, M., Ochoa, C., & Loewe, G. (2016). Using passive data from a meter to complement survey data in order to study online behavior. *Social Science Computer Review*, 1–16.

Revilla, M., Toninelli, D., Ochoa, C., & Loewe, G. (2016). Do online access panels need to allow and adapt surveys to mobile devices? *Internet Research*, *26*(5), 1209–1227.

Sakshaug, J. W., Couper, M. P., Ofstedal, M. B., & Weir, D. R. (2012). Linking survey and administrative records: mechanisms of consent. *Sociological Methods &*

Research, *41*(4), 535–569.

Salamé, P., & Baddeley, A. (1989). Effects of background music on phonological short-term memory. *The Quarterly Journal of Experimental Psychology*, *41*(1), 107–122.

Salvucci, D. D., & Taatgen, N. A. (2011). *The Multitasking Mind*. New York, NY: Oxford University Press.

Schepers, J. M. (2007). The construction and evaluation of an attention questionnaire. *SA Journal of Industrial Psychology*, *33*(2), 16–24.

Scherpenzeel, A. (2017). Mixing online panel data collection with innovative methods. In S. Eifler & F. Faulbaum (Eds.), *Methodische Probleme von Mixed-Mode-Ansätzen in der Umfrageforschung* (pp. 27–49).

Schober, M. F., Conrad, F. G., Antoun, C., Ehlen, P., Fail, S., Hupp, A. L., … Zhang, C. (2015). Precision and disclosure in text and voice interviews on smartphones. *PLoS ONE*, *10*(6), 1–20.

Sendelbah, A., Vehovar, V., Slavec, A., & Petrovcic, A. (2016). Investigating respondent multitasking in web surveys using paradata. *Computers in Human Behavior*, *55*, 777–787.

Struminskaya, B., Weyandt, K., & Bosnjak, M. (2015). The effects of questionnaire completion using mobile devices on data quality. Evidence from a probability-based general population panel. *Methods, Data, Analyses*, *9*(2), 261–292.

Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers*. San Francisco: Jossey-Bass.

TNS BMRB. (2014). *Technical Report of the 1958 National Child Development Study: Age 55 survey (2013/2014).* Retrieved from http://www.cls.ioe.ac.uk/shared/get-file.ashx?id=1942&itemtype=document

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*, *32*(4), 544–560.

Toepoel, V., & Lugtig, P. (2015). Online surveys are mixed-device surveys. Issues associated with the use of different (mobile) devices in web surveys. *Methods, Data, Analyses*, *9*(2), 155–162.

Tourangeau, R. (1984). Cognitive sciences and survey methods. In T. B. Jabine, M. L. Straf, J. M. Tanur, & R. Tourangeau (Eds.), *Cognitive Aspects of Survey Methodology: Building a Bridge between Disciplines* (pp. 73–100). Washington DC: National Academy Press.

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

Trafton, J. G., & Monk, C. A. (2007). Task interruptions. *Reviews of Human Factors and Ergonomics*, *3*(1), 111–126.

University of Essex. Institute for Social and Economic Research. (2017). Understanding Society: Innovation Panel, Waves 1-9, 2008-2016. [data collection]. 8th Edition. UK Data Service. http://doi.org/http://doi.org/10.5255/UKDA-SN-6849-9

Wells, T., Bailey, J. T., & Link, M. W. (2014). Comparison of smartphone and online computer survey administration. *Social Science Computer Review*, *32*(2), 238–255.

Zahariev, M., Ferneyhough, C., & Ryan, C. (2009). *Best practices in mobile research*. Paper presented at the ESOMAR Online Research Conference, Chicago, IL.

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