ISER Working Paper Series

Completing Web Surveys on Mobile Devices: Does Screen Size Affect Data Quality?

Alexander Wenz
Institute for Social and Economic Research
University of Essex

No. 2017-05
April 2017

INSTITUTE FOR SOCIAL & ECONOMIC RESEARCH

www.iser.essex.ac.uk

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

University of Essex

Non-Technical Summary

An increasing number of survey respondents complete web questionnaires on their smartphone or tablet instead of using laptops or desktop computers. This creates new challenges for survey researchers. One of the primary concerns is that certain features of mobile devices, such as the smaller screen size or the touchscreen interface, make it harder for survey respondents to complete the survey, especially if the questionnaire was not designed for mobile devices. As a result, respondents who use smartphones or tablets might provide survey data of lower quality compared to those using laptops or desktop computers. Survey researchers, however, have not yet fully understood which features of mobile devices contribute to the lower data quality among mobile device users. Although it has been speculated that a small screen size may be the major factor, this assumption has not been directly tested.

In this paper, I examine how the diagonal screen size of a mobile device affects survey data quality by comparing quality indicators between four screen size groups: small smartphones vs. large smartphones vs. small tablets vs. large tablets. The analysis is based on data from an online survey in the UK that provides mobile respondents with a questionnaire that was not originally designed for mobile devices, what I refer to as *non-mobile-optimised* questionnaire.

The results show that data quality mainly differs between small smartphones, that have a screen size of below 4.0 inches, and larger mobile devices. Survey respondents who use small-screen smartphones are more likely to break off the survey and they provide shorter answers to open questions compared to respondents with larger devices. Users of small smartphones are also more likely to select the same item in grid questions instead of providing more differentiated responses. However, I did not find any evidence that respondents with small-screen mobile devices need more time for survey completion, that small screens affect the responses to questions with a long list of response categories, nor that small-screen users select fewer items in check-all-that-apply questions.

Completing Web Surveys on Mobile Devices: Does Screen Size Affect Data Quality?

Alexander Wenz

Abstract

Using data from a non-mobile-optimised survey in the UK, 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 straight-line in grid questions. There are no significant differences between screen size groups in completion times, response distributions, and number of items selected in check-all-that-apply questions.

Keywords: mobile web survey, smartphone, tablet, screen size, measurement error, data quality

JEL classification: C81, C83

Acknowledgements: This work was supported by a +3 PhD grant awarded by the Economic and Social Research Council [grant number ES/J500045/1]. I would like to thank Peter Lynn, Tarek Al Baghal, Annette Jäckle, Mick Couper and Joel Williams for comments on earlier drafts of this paper, and Kantar Public for providing me with the data used in this paper.

Corresponding address: Alexander Wenz, Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester, Essex, CO4 3SQ, UK, awenz@essex.ac.uk.

Introduction

Mobile technology has become an integral part of people's daily life. In 2016, around 70 percent of Britons used a smartphone and around 60 percent used a tablet (Ofcom, 2016). On average, British people spend one and a half hours on their smartphone each day, those aged 16-24 even three and a half hours. Tablets are used daily for around half an hour across all age groups (Ofcom, 2014). 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 (Lugtig & Toepoel, 2015; Peterson, 2012; Poggio, Bosnjak, & Weyandt, 2015; Revilla, Toninelli, Ochoa, & Loewe, 2014). 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% in March 2012 to 11% in September 2013 (de Bruijne & Wijnant, 2014). More recently, Struminskaya, Weyandt, and Bosnjak (2015) report that in 2014, 18% 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% 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 assumed 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 provide survey data of lower quality compared to respondents who use desktop computers or laptops. Survey managers have various options to handle mobile devices: they can offer an optimised web questionnaire for mobile browsers, provide the questionnaire as a mobile application, or administer the conventional web questionnaire and discourage the use of mobile devices (Callegaro, Lozar Manfreda, & Vehovar, 2015). Although 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 question with horizontal rating scales on mobile devices.

Therefore, some large-scale social surveys still adopt the more conservative approach of discouraging mobile survey completion or even blocking mobile device access. There is, however, little agreement yet as to which types of mobile devices should be discouraged from being used to maintain high data quality. Different thresholds have been applied that were either based on diagonal screen size (Hanson, Matthews, & McGee, 2015) or on screen resolution (TNS BMRB, 2014).

Previous studies on response quality in non-mobile-optimised surveys found larger quality differences between smartphones and PCs than between tablets and PCs. Therefore, it is speculated that a small screen size may be the major reason why mobile respondents provide data of lower quality compared to PC users, if they are not offered an optimised questionnaire or a mobile survey app (de Bruijne & Wijnant, 2013; Guidry, 2012; Lugtig & Toepoel, 2015). However, this assumption 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 quality. The results 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 optimization be triggered?

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 are not optimised for mobile devices (Bosnjak et al., 2013; Callegaro, 2010; Couper, Antoun, & Mavletova, 2017; 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? Website content is displayed proportional to screen size if a website is not adapted to mobile devices, for example using a responsive web design. The survey page is therefore 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 with devices with larger screens. The screenshots in Appendix Figure A1 from 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 (Figure A2). 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 the survey page may not fit entirely on a small screen. Survey participants may need to scroll to see elements of the page that are initially not visible, such as parts of the question text or response options (Figure A3).

These usability problems 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 and Peterson (2016) examined question-level response times of web surveys taken on PCs and on 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.

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

Survey breakoff

Completing a non-optimised survey on a small mobile device requires more psychomotor effort as it is more difficult to record the answer using small buttons and a small keyboard, in particular when answering grid or open questions. If respondents perceive survey completion as too burdensome, they may decide to switch to larger device or may drop out of the survey. 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

Since non-mobile-optimised survey pages may exceed small screens, it can be expected that some response options are initially not visible on small devices and require the respondent to scroll down. Respondents 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 earlier 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

Responding to open-ended questions may be particularly burdensome on small devices because the keys of the digital keyboard are smaller, which makes it more difficult for respondents to type. To reduce their effort, users with small screens may try to minimise typing and give shorter answers to open-ended questions.

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) and may select fewer items in check-all-that-apply questions. Extant research has mixed findings with regard to straight-lining (Guidry, 2012; Lugtig & Toepoel, 2015; McClain et al., 2012) and no 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.

Data

The analysis is based on data from the web component of the Community Life Survey 2013-2014 which were collected from October 2013 to April 2014 (Cabinet Office, 2014). 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% was achieved for the web survey component. The standard PC version of the questionnaire was delivered to mobile devices, i.e. the questionnaire was not mobile-optimised. In the invitation letter, respondents were not particularly encouraged to use mobile devices but mobile access was not blocked for any device.

In total, N = 4,698 respondents took part in the web survey: 3,638 respondents (= 77.4%) completed the survey on a desktop PC or laptop whereas 1,060 respondents used a mobile device (= 22.6%). 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 were only available for one of the four fieldwork quarters of the survey. In this reduced 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 summarises the sample size available for each of the six data quality indicators used in the analysis.

Table 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	---	945	109
Straight-lining	---	950	109
Check-all-that-apply questions	---	951	109
Community Life web survey sub-sample	**887**	**260**	**48**
Completion time	---	237	35
Survey breakoff	---	260	48
Length of open responses	---	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; the following analyses are therefore based on mobile devices only.

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 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 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 are operationalised as follows:

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 above 150 minutes were removed from the analysis. Respondents who completed the survey in less than 10 minutes were also removed.

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 ask whether the respondent has donated any money to the listed charities. 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. 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) 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. Survey respondents are defined as straight-liners if they give the same response to all grid items. One tablet user refused to answer the question and was 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. These questions ask about different aspects of community involvement. For each respondent, the average number of selected items was calculated across all applicable questions.

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 four 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 four screen size categories in the regression model, three dummy variables were created by setting the group of large tablets as baseline category. A separate indicator for 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 (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. The model predicting completion times furthermore includes a count variable indicating how many items the respondent was asked in the survey because respondents may get 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 modelling count data (length of open responses).

The socio-demographic variables considered in the multivariate analysis contain a considerable amount of missing data. In the full sample, around 27% 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 are imputed using multiple imputation with n = 5 imputations. The imputation is conducted in SPSS using the fully conditional specification (FCS) algorithm.

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 (Table 3). Surprisingly, respondents using large tablets have on average the longest completion times whereas the completion times of the other three screen size groups are on a similar level of around 31-34 minutes. The difference in mean completion times between screen size groups is not statistically significant as determined by a one-way ANOVA, $F(3, 268) = 0.968$, $p > 0.05$.

Table 3. Data quality indicators by screen size.

	Smartphone		Tablet	
	Small	Large	Small	Large
Mean completion time (in minutes)	34.2	31.0	30.7	37.3
	(15)	(20)	(48)	(189)
% Breaking off	31.8	15.4	7.8	4.3
	(22)	(26)	(51)	(209)
% Selecting response option in lower half of question	74.6	78.9	76.0	80.3
	(71)	(38)	(104)	(841)
Mean length of open responses (in characters)	49.4	85.5	74.8	93.3
	(21)	(19)	(43)	(200)
% Straight-lining	5.6	2.6	0.0	1.2
	(71)	(38)	(105)	(845)
Mean number of responses in check-all-that-apply	2.2	2.3	2.3	2.3
	(71)	(38)	(105)	(846)

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 4). The model shows no significant differences in completion times between screen sizes.

Table 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	-31.59	-3.77**	1.77***	3.89***	-3.78**	2.01***
Small Smartphone	-0.49	2.58***	-0.37	-0.59*	1.59*	-0.08
Large Smartphone	-0.78	1.62*	-0.13	-0.01	0.72	0.03
Small Tablet	-1.03	0.81	-0.25	-0.11	-16.84	0.01
Large Tablet	*-Baseline-*	*-Baseline-*	*-Baseline-*	*-Baseline-*	*-Baseline-*	*-Baseline-*
Age	0.25**	0.01	0.00	0.01	-0.01	0.00
Gender: Male	-1.21	-0.03	0.06	-0.20	-0.48	-0.03
Education: A-levels	3.90	0.17	-0.22	0.28	0.53	0.21***
Working status: paid work	-4.79	-0.38	-0.10	0.16	-0.88	0.09*
Income: High	-1.44	0.01	-0.13	0.09	0.15	0.02
Living alone in HH	4.32	-0.13	-0.05	-0.08	-0.39	0.00
Volunteering: Yes	-1.29	0.27	–	–	–	–
# Items completed	0.19**	–	–	–	–	–
N	272	308	1,054	283	1,059	1,060
Regression model	OLS	Logistic	Logistic	Negative binomial	Logistic	OLS

Note: * $p < .05$, ** $p < .01$, *** $p < .001$. 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 compared to respondents using larger screens.

However, the model does not allow us to draw 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 3). Users with smaller screens have a higher propensity to drop out of the survey: Almost one third (= 31.8%) of respondents using small smartphones and 15.4% of the large smartphone users failed to finish the survey. Among respondents with small tablets 7.8% dropped out and among users of large-screen tablets only 4.3%. The chi-square test for independence indicates that the relationship between screen size and breakoff rate is significant, $\chi^2 (3) = 23.296$, $p < 0.001$. 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. Only the standardised residuals for the small smartphone group are significant at $p < 0.01$ as the value of $z = 4.0$ is outside of the $\pm$ 2.58 threshold whereas the other three groups have non-significant standardised residuals. This implies that the significant association between screen size and survey breakoff is mainly driven by small smartphone 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 large tablet respondents, $p < 0.001$ (Table 4). Similarly, large smartphone users are significantly more likely to break off than users with large-screen tablets, $p < 0.05$. The effect of the small tablet group compared to large tablet users points in the expected direction but is not statistically significant. Interestingly, breakoff probability falls monotonically with screen size. It is largest for the small smartphone group, smaller for the large smartphone group and smallest for the small tablet group. None of the control variables are statistically significant at $p < 0.05$.

The findings of the bivariate and multivariate analysis support Hypothesis 2. Respondents with smaller screens, particularly the small smartphone and the large smartphone group, have a higher propensity to drop out of the survey. This suggests that a higher respondent burden experienced during survey completion on small-screen devices may have been one of the factors that made small-screen users more likely to drop out.

Apart from the overall breakoff rate, I also analysed 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. However, findings 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% and 80.3% (Table 3). The maximum difference of 5.7 percentage points is between the small smartphone and the large tablet 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, $\chi^2(3) = 2.119$, $p > 0.05$. 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 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 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 open answers than those with large-screen devices as typing on small screens might be more burdensome (H4).

A bivariate analysis confirms the theoretical expectations (Table 3). Respondents with small smartphones provide on average the shortest responses to open-ended questions, with 49.4 characters on average, whereas users of large tablets have the longest open answers, with a mean length of 93.3 characters. The large smartphone and small tablet groups have a mean response length which lies in-between as they as they provide responses of around 74.8 to 85.5 characters. A one-way ANOVA indicates, however, that the difference in means across the four screen size groups is not significant, $F(3, 282) = 1.957$, $p > 0.05$. In addition, a Tukey post-hoc test which makes pairwise comparisons shows that none of the compared groups are significantly different from each other.

To control for selection effects, a negative binomial regression is fitted to predict the number of characters in the three open-ended questions (Table 4). The model shows that only the effect of the small smartphone group is significant at $p < 0.05$ when being compared to large-screen tablets.

The multivariate analysis provides evidence for Hypothesis 4. When controlling for selection effects, small smartphone users provide significantly shorter answers to open questions compared to those using large tablets. The answer length of respondents with large smartphones and small tablets, however, does not seem to differ from the length of large tablet users. This finding suggests that typing is more burdensome on small-screen devices with a screen size of less than four inches.

H5. Straight-lining

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

A bivariate analysis of straight-lining across the four screen size groups (Table 3) shows that there is a higher proportion of straight-lining respondents among the small smartphone group (5.6%) than among the large smartphone (2.6%) and the large tablet group (1.2%). 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 occurrence of straight-lining, $\chi^2(3) = 11.283$, $p < 0.05$. 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 ($z = 3.0$, $p < 0.01$).

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 4). The model indicates that the group of small-screen smartphone users is significantly more likely to straight-line compared to the large tablet group, $p < 0.05$. Large smartphone users are also more prone to straight-lining than respondents with large-screen tablets but the effect is not significant. Among the control variables, none of the effects are significant.

Bivariate and multivariate analyses provide support for Hypothesis 5. Users with small devices, particularly small smartphones, are more likely to straight-line than respondents with larger devices. This finding suggests that small-screen respondents are more prone to satisficing response behaviour in grid questions, presumably as they experience higher respondent burden when answering questions of this format.

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 around 2.2 to 2.3 items per check-all-that-apply question (Table 3). The small smartphone group has the lowest mean whereas the mean of the other three groups have the same value. A one-way ANOVA indicates that the difference in means between the four screen size groups is not significant, $F(3, 1056) = 0.968$, $p > 0.05$.

A linear regression which predicts the mean number of selected items in check-all-that-apply question confirms the bivariate findings (Table 4): none of the three dummy variables are statistically significant. Among the control variables, education ($p < 0.001$) and employment status ($p < 0.05$) are significant predictors of item selection in check-all-that-apply questions.

The bivariate and multivariate findings do not support Hypothesis 6. It seems that small-screen users do not take more cognitive shortcuts in check-all-that-apply questions than users with larger screens, presumably as they do not experience higher respondent burden in this question format.

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 the UK, 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 are more likely to adopt straight-lining response patterns in grid questions. 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 the number of selected items in check-all-that-apply questions.

The present study provides suggestive evidence that surveys ought to provide mobile-optimised questionnaires 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 grid questions because both question formats seem to be more burdensome on small-screen smartphones. The other main problem is the higher dropout rate of users of small smartphones, which may result in nonresponse error. The response quality of surveys taken on other mobile devices, however, seems to be 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.

References

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

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*, in press.

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.

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

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*. 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.

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., 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.

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.

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.

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

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

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.

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*.

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.

Ofcom. (2014). *The Communications Market Report*. Retrieved from http://stakeholders.ofcom.org.uk/binaries/research/cmr/cmr14/2014_UK_CMR.pdf

Ofcom. (2016). *The Communications Market Report*. Retrieved from https://www.ofcom.org.uk/__data/assets/pdf_file/0024/26826/cmr_uk_2016.pdf

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.

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., Toninelli, D., Ochoa, C., & Loewe, G. (2014). Do online access panels really need to allow and adapt surveys to mobile devices? *RECSM Working Paper, 41*.

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.

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

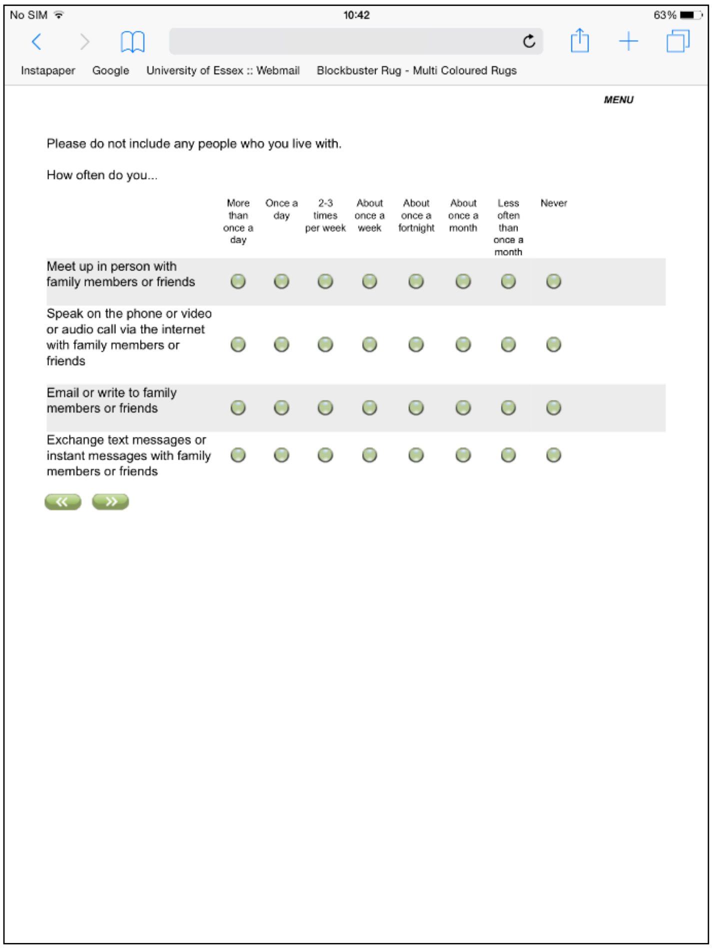
Toepoel, V., & Lugtig, P. (2014). What happens if you offer a mobile option to your web panel? Evidence from a probability-based panel of internet users. *Social Science Computer Review, 32*(4), 544–560.

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

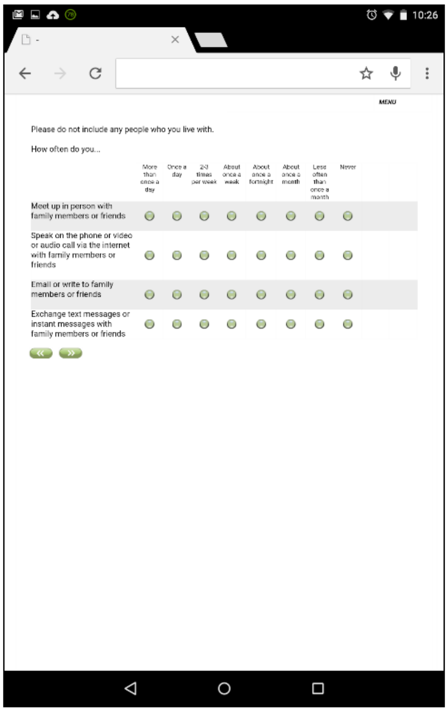
Appendix. Screenshots

Figure A1. 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*)

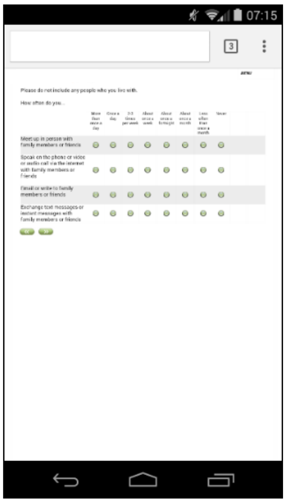
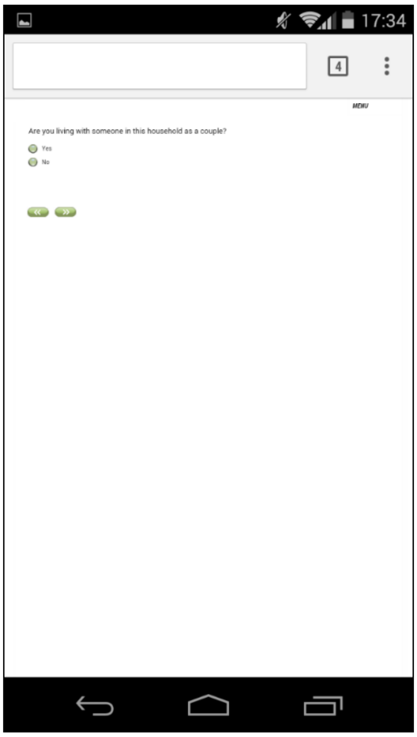


Figure A2. Zooming in on 4.5-inch smartphone (*Motorola Moto G*)

a) Not zooming in



b) Zooming in

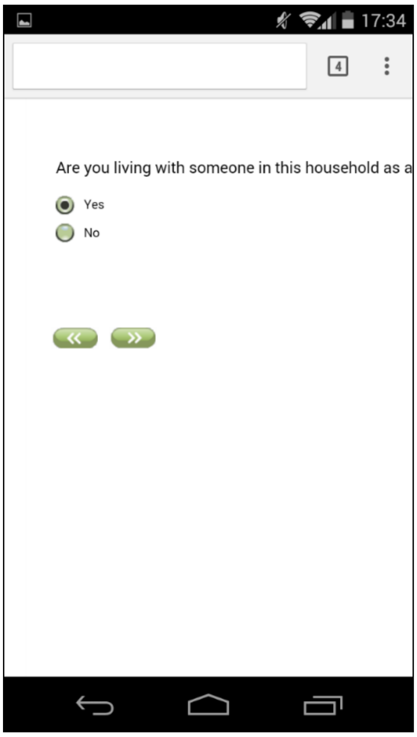
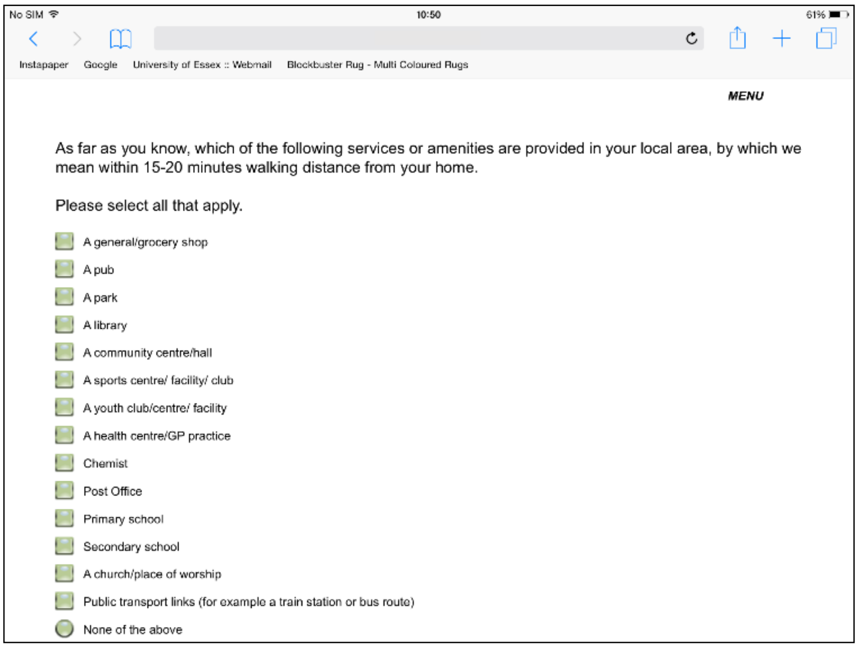


Figure A3. Question displayed in horizontal orientation on 9.7-inch tablet and 7.0-inch tablet

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



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