The Effects of Questionnaire Completion Using Mobile Devices on Data Quality. Evidence from a Probability-based General Population Panel

Bella Struminskaya, Kai Weyandt & Michael Bosnjak
GESIS – Leibniz Institute for the Social Sciences

Abstract

The use of mobile devices such as smartphones and tablets for survey completion is growing rapidly, raising concerns regarding data quality in general, and nonresponse and measurement error in particular. We use the data from six online waves of the GESIS Panel, a probability-based mixed-mode panel representative of the German population to study whether the responses provided using tablets or smartphones differ on indicators of measurement and nonresponse errors from responses provided via personal computers or laptops. We follow an approach chosen by Lugtig and Toepoel (2015), using the following indicators of nonresponse error: item nonresponse, providing an answer to an open question; and the following indicators of measurement error: straightlining, number of characters in open questions, choice of left-aligned options in horizontal scales, and survey duration. Moreover, we extend the scope of past research by exploring whether data quality is a function of device-type or respondent-type characteristics using multilevel models. Overall, we find that responding with mobile devices is associated with a higher likelihood of measurement discrepancies compared to PC/laptop survey completion. For smartphone survey completion, the indicators of measurement and nonresponse error tend to be higher than for tablet completion. We find that most indicators of nonresponse and measurement error used in our analysis cannot be attributed to the respondent characteristics but are rather effects of mobile devices.

Keywords: mobile phone surveys, panel survey, mobile devices, nonresponse, measurement

1 Introduction

In web surveys and online panels, it can no longer be expected that respondents participate using desktop computers and laptops only. Survey researchers have reported a growing share of unintended mobile respondents – respondents who use their mobile devices such as smartphones or tablets to access and participate in surveys that were originally designed to be taken on PCs or laptops (de Bruijne & Wijnant, 2014b; Peterson, 2012; Toepoel & Lugtig, 2014; Wells, Bailey, & Link, 2014). In the Dutch online probability-based LISS Panel, the proportion of unintended mobile respondents increased from 3% in 2012 to 11% in 2013, in the CentERpanel, another probability-based general population online panel in the Netherlands, the proportion of unintended mobile respondents increased from 3% in 2012 to 16% in 2013 (de Bruijne & Wijnant, 2014b). In the German mixed-mode GESIS Panel, in 2014 about 17.9% of online respondents completed the questionnaires using mobile devices with 9.2% using smartphones and 8.7% using tablets. In 2015, about 15.6% of online respondents name tablets and 8.1% name smartphones as the preferred mode to answer the questionnaires.[1]

Responding to surveys using various devices, that increasingly become heterogeneous with regard to size and functionality, raises concerns about data quality. Differences between PCs/laptops and mobile devices in screen size and input method as well as the possibility to participate in surveys via mobile devices from a variety of locations and situations where distractions are possible can affect respondents' cognitive processing, increasing the risk of errors (Peytchev & Hill, 2010). Nonresponse error and measurement error are of particular concern.

Respondents using mobile devices for survey completion have demonstrated lower response rates (Buskirk & Andrus, 2014; de Bruijne & Wijnant, 2013), lower completion rates (Mavletova, 2013; Mavletova & Couper, 2013), and higher break-off rates[2] (Callegaro, 2010; Cook, 2014; Mavletova, 2013; McClain, Crawford, & Dungan, 2012; Poggio, Bosnjak, & Weyandt, 2015; Stapleton, 2013). Item-nonresponse has been found to be more pronounced when completing the survey on a mobile device in open-ended questions (Peytchev & Hill, 2010). However, more recent studies did not replicate this result: de Bruijne and Wijnant (2014a) show

1 GESIS (2015): GESIS Panel - Standard Edition. GESIS Datenarchiv, Cologne. ZA5665 Data file version 8.0.0, doi:10.4232/1.12245. Own calculations.

2 We use the term response rate for studies based on a probability samples and completion rate for studies that are not based on probability samples. For studies that focused on break-offs we do not divert from the original terminology used by the authors.

Direct correspondence to
Bella Struminskaya, GESIS – Leibniz Institute for the Social Sciences,
B 2,1, 68159 Mannheim, Germany
E-mail: bella.struminskaya@gesis.org

that respondents using mobile devices are not more likely to provide a half-open “other” answer than to choose a closed “other” option; Wells et al. (2014) find that mobile respondents are not more likely to skip the half-open or open questions. Nevertheless, mobile web respondents have been shown to provide shorter answers to open-ended questions than PC respondents (Mavletova, 2013; Peterson, 2012; Wells et al., 2014).

The second major concern in mobile web surveys is the risk of more pronounced measurement errors. Comparing the responses provided by mobile web respondents to the record data, Antoun (2015) shows that smartphone respondents provide fewer accurate answers when reporting age and date of birth than PC respondents. Cases when validation data is available to the researchers to study measurement errors are an exception rather than a rule. Hence, most researchers use indicators of satisficing behavior that suggests reporting with measurement error. Krosnick (1991) defines satisficing as respondents’ failure to consecutively and carefully execute the cognitively demanding stages that precede producing accurate and valid survey responses. These stages include interpreting the meaning of the question, retrieval of relevant information from memory, formation a summary judgement, carefully integrating this information, and clear report of the summary judgement (Tourangeau, 1984; Tourangeau, Rips, & Rasinski, 2000). Satisficing behavior is the result of the interplay of three factors: respondents’ ability, motivation and difficulty of the task (Krosnick, 1991, p. 225). Using a mobile device for survey completion can be a difficult task due to technical reasons such as a small screen, a touchscreen, as well as situational characteristics if respondents are outside of home. Providing satisfactory answers instead of accurate answers is indicative of measurement error.

In past studies, the following indicators of satisficing have been used when studying mobile web responses: number of “don’t know” answers, non-differentiation (straightlining), primacy effects, rounding, measures of superficial cognitive processing (e.g., answers to cognitive reflection tests), avoiding half-open questions, length of answers to open-ended questions, and answers to sensitive questions (Antoun, 2015; Buskirk & Andrus, 2014; Lugtig & Toepoel, 2015; Mavletova, 2013; Mavletova & Couper, 2013; Wells et al., 2014). Lugtig and Toepoel (2015) find that mobile web respondents report with higher measurement error than PC respondents showing more item missing responses, higher item-nonresponse in open-ended questions, more primacy effects, and fewer response options selected in check-all-that-apply questions. Conversely, in other studies little evidence is found: mobile web respondents are not more likely to demonstrate primacy effects (Buskirk & Andrus, 2014; Mavletova, 2013; Toepoel & Lugtig, 2014; Wells et al., 2014), do not differ from PC respondents in providing socially desirable answers (Antoun, 2015; Mavletova, 2013), do not show increased rounding or superficial cognitive processing (Antoun, 2015)[3]. Mixed results have been obtained on using the horizontal scales in mobile web surveys. Peytchev and Hill (2010) found that horizontal scrolling generally did not affect responses but a small proportion of respondents failed to scroll and see all possible answer options. De Bruijne and Wijnant (2014a) find that horizontal scale format produces slightly more item missings than the vertical format even when the horizontal scales are fully visible on screen with no need to scroll.

Survey duration, another indicator of satisficing behavior in web surveys with shorter duration being associated with more primacy effects (Malhotra, 2008), has been shown to produce opposite results for mobile web surveys. Using smartphones for survey completion is associated with longer completion times (Antoun, 2015; Cook, 2014; de Bruijne & Wijnant, 2013; Mavletova, 2013; Mavletova & Couper, 2013; Peterson, 2012; Wells et al. 2014). However, the longer duration can be explained by other factors such as connection speed, scrolling, familiarity with the device, or distractions due to respondents' multitasking. Couper and Peterson (2015) show that the connection speed accounts for a small proportion of the difference between PC and smartphone completion. They further argue that multitasking and familiarity with the device are less plausible explanations than the display size and the need for scrolling.

In light of the mixed results about the data quality in mobile web surveys outlined above it is noteworthy that few studies on mobile responding are based on probability-based online panels; and from those that are, several studies are based on the LISS Panel (cf. Antoun, 2015; de Bruijne & Wijnant, 2013; de Bruijne & Wijnant, 2014; Lugtig & Toepoel, 2015), other studies are based on the CentER-panel in the Netherlands (de Bruijne & Wijnant, 2014b) or the Knowledge Panel of GfK Knowledge Networks in the USA (Wells et al., 2014). Mobile web respondents in probability-based panels can differ from mobile respondents in nonprobability panels. Respondents in nonprobability panels can be more technologically sophisticated and able to answer surveys on mobile devices, thereby compensating mea-

3 It can be assumed that finding adverse effects on data quality can be caused by some studies being optimized for survey completion while others are not. Indeed, studies mentioned in this paragraph with the exception of Antoun (2015) were optimized for mobile completion or included experimental conditions that were optimized for mobile devices. However, it does not seem that mixed results presented in this section can be fully explained by mobile optimization as providing shorter answers in open-ended questions, lower completion and response rates are found in both optimized and non-optimized studies. In this review, studies with optimized design (i.e., where special programming for mobile devices was performed), including experimental conditions are: Buskirk & Andrus 2014, de Bruijne & Wijnant 2013, Mavletova 2013, Mavletova & Couper 2013, Peytchev & Hill 2010, Stapleton 2013, Toepoel & Lugtig 2014, and Wells, Bailey & Link 2014. Non-optimized studies are: Antoun 2015, Callegaro 2010, Cook 2014, de Bruijne & Wijnant 2014, 2014a, Lugtig & Toepoel 2015, McClain et al. 2012, Peterson 2012, and Poggio, Bosnjak, & Weyandt 2015.

surement errors with their experience and motivation. For example, in a Russian non-probability panel, Mavletova (2013) finds that more experienced mobile users wrote significantly longer answers to open questions than less experienced mobile users. Furthermore, learning effects can play a role if respondents in nonprobability panels are more experienced than respondents in probability-based panels. It has been shown that professional respondents in nonprobability panels are not more likely to produce data of lower quality (Hillygus, Jackson, and Young, 2014; Matthijsse, de Leeuw, and Hox (2015), but this aspect has not been studied for mobile device vs. PC survey completion.

It is important to investigate the consequences of responding via mobile devices in probability-based general population panels to fully understand whether mobile web response is something survey researchers should be concerned about, given the mixed results provided by the literature reported above. In this article, we concentrate on nonresponse and measurement using several measures of satisficing behavior as indicators of possible measurement errors. We follow an approach chosen by Lugtig and Toepoel (2015) for the LISS Panel data using the data from the GESIS Panel, a probability-based mixed-mode (online and mail) panel of the general population in Germany.

If preferences to answer surveys using a particular device are correlated to the propensity to satisfy, selection and measurement effects will be confounded (Lugtig & Toepoel, 2015). Indeed, past studies have found that respondents answering online surveys via mobile devices differ at least in their demographic characteristics from those who answer online surveys via laptops and PCs (Cook, 2014; de Bruijne & Wijnant, 2013; de Bruijne & Wijnant, 2014b; Toepoel & Lugtig, 2014). Cook (2014), who uses the U.S. data, finds that demographic composition of device groups differ: those who take surveys on tablets are significantly younger, more likely to be female; smartphone respondents are lower educated and have lower income than tablet and PC respondents, both smartphone and tablet use is higher for Hispanics and African-Americans. For the Netherlands, de Bruijne and Wijnant (2013) find small differences in gender between smartphone and PC users with smartphone users more likely to be men; the proportion of those higher educated is significantly higher among smartphone users. Consistent with other studies, mobile web use is highest among young respondents. Toepoel and Lugtig (2014) demonstrate that income, household size, and household composition are predictive of mobile survey completion. Furthermore, de Bruijne and Wijnant (2014b) find that in the LISS Panel sex and age are predictive of unintended access to online surveys via smartphones and tablets. Women and younger respondents are more likely to use mobile devices for survey access. Additionally, living alone is negatively associated with accessing online surveys via tablets while respondents in paid work are more likely to use tablets to access online surveys.

Therefore, it is important to study whether certain respondent behaviors are attributable to a respondent (response style) or are a result of survey completion using mobile devices. This conceptual extension to past approaches involves disentangling device-level and respondent-level determinants of data quality indicators using a multilevel perspective. Overall, our analyses have two goals: (1) to find out to which extent the findings of Lugtig and Toepoel (2015) can be replicated in the GESIS Panel, that is, generalized across different countries and panel configurations, and (2) disentangle the effect of respondent characteristics and device characteristics on measurement-related and nonresponse-related data quality indicators.

2 Data, Measures, and Hypotheses

We use data from six waves of the GESIS Panel – a face-to-face recruited mixed-mode probability-based panel, which is representative of the general population in Germany aged 18 to 70 years at the time of recruitment. About 65 percent of respondents participate online and about 35 percent participate offline via postal mail questionnaires. The recruitment for the GESIS Panel took place in 2013. The first regular wave was fielded in the beginning of 2014. Respondents receive invitations to participate in self-administered surveys every two months. The recruitment rate for the GESIS Panel is 31.6% (AAPOR RR5), the response rate for the profile survey is 79.4%. For 2014 surveys, the completion rates per wave vary between 88.7% and 92.0% for the online questionnaires and between 76.7% and 84.6% for the offline questionnaires. All active panel members receive unconditional incentives of five euros with questionnaire invitations for every wave per post. For our analysis, we use the data for online respondents only. Overall, 3041 online respondents were invited to participate in the first regular GESIS Panel wave in 2014. From those, we exclude 127 persons who did not participate in any of the waves in 2014 as well as one person who switched modes from online to offline. This leaves us with a sample size of 2913 respondents.

The online questionnaires in GESIS Panel are not programmed in a mobile device optimized way, that is, questions are not adjusted for a particular device. For the identification of the device used by a respondent to complete the questionnaire we use the user agent strings (UAS) provided by the panel software. The user agent strings are recoded into the device-variables using a Stata code "parseuas" developed by Rossmann and Gummer (2014). The script distinguishes between mobile phones, tablets and other devices used to complete the questionnaire. The category "other devices" includes desktop computers, laptops and possibly a small proportion of the devices with browser versions that cannot be classified as mobile phones or tablets. Thus, the proportion of PC-completions might be somewhat overestimated in our analyses.

The contents of the questionnaires fielded in the GESIS Panel vary from wave to wave. In order to eliminate the influence of varying questionnaire content on nonresponse and measurement error indicators, our analyses are based on an (mostly) invariant set of questions that are asked in each survey wave. This approach was chosen by Lugtig and Toepoel (2015) for the analyses based on the LISS Panel. The questions that are invariant in every wave are concerned with survey evaluation as they are in the LISS Panel. However, the indicators for the GESIS Panel are slightly different. The evaluation includes various types of questions: a grid question, an open question, and several singe-choice questions. The evaluation part includes overall 14 items about the questionnaire itself, the device used to fill out the questionnaire, whether the respondent completed the questionnaire without a time break, and if not, how long the break lasted, whether the questionnaire was completed at home or outside of the home, whether others were present, and an open field for remarks about the questionnaire.

We use the following indicators of measurement error (ME) and nonresponse error (NR): item-nonresponse (NR), item-nonresponse to an open question (NR), length of answers to an open question (ME), straightlining (ME), choice of left-aligned answer options in horizontal scales (ME), and survey duration (ME). The indicators are operationalized as follows.

Item-nonresponse: We use all of the items for questionnaire evaluation, reported device and conditions under which the respondent filled out the questionnaire to count the number of item missings. We exclude the remark as well as the open question about the duration of the time break if the respondent indicates that he or she did not complete the survey without a break. Thus, the indicator for the number of missing values ranges from 0 to 13. We expect respondents who use smartphones for survey completion to show higher number of item missings. However, we expect no differences in item missings between PC and tablet respondents (Hypothesis H1).

Straightlining: The first question about the questionnaire evaluation is a grid question that contains six items: whether the survey was interesting, diverse, and important for research, long, difficult, or too personal, each measured with a five-point labeled scale. We define straightlining as providing the same answer to all of the items of the grid a respondent answered if the respondent answered at least two items from the evaluation grid. Lugtig and Toepoel (2015) find that straightlining is surprisingly higher for PC respondents. However, the questions they used for analysis were not arranged in a grid. For grid questions, straightlining has been shown to be higher for respondents using mobile phones than for those using tablets and PCs (McClain et al., 2012). Since we use the grid question, we expect to find more straightlining for respondents who answer the questionnaire via smartphones (Hypothesis H2a). For tablets, we expect to find no differences to PCs given the larger screen size (Hypothesis H2b).

Response to an open question: At the end of each questionnaire respondents have the opportunity to provide additional verbal feedback about the questionnaire. We use a binary variable whether a respondent has provided feedback or not. We expect respondents who use smartphones or tablets for survey completion to provide answers to an open question at a lower rate than respondents who complete the survey using PCs (Hypothesis H3).

Length of the answer to an open question: The second indicator that we use related to the open questions is the length of the answer provided by a respondent. In line with the findings from the literature reviewed in the previous section, we expect respondents who fill out their questionnaires via smartphones to provide shorter answers given the small screen size (Hypothesis H4a). We expect to find no differences in answers to open questions or length of these answers provided via PCs and tablets (Hypothesis H4b).

*Choice of left-aligned option*s: The measure of a higher proportion of left-aligned answer options selected is based on the items of the grid evaluation question as well as three single-choice evaluation questions with 5-point horizontal scales. One of these three items is the overall questionnaire evaluation, the other two items vary between the waves: for the first three waves the items ask whether the questions were understandable and whether they made the respondent think about things and in all the following waves the questions asked about how difficult it was to understand the questions and how difficult it was to find an answer. We count the number of times respondent chose the two answer options aligned to the left. Although the questions differ between the waves this should not affect the rate at which respondents using different devices provide options aligned to the left or not. We expect more left-aligned options for responses on smartphones than for PCs and tablets (Hypothesis H5a). No difference between PCs and tablets is expected due to the screen size (Hypothesis H5b).

Duration: The duration is measured in seconds for every wave. We truncated the extreme values of questionnaire duration longer than an hour to an hour. Since our surveys are not optimized for mobile devices, we expect longer completion times both for smartphones and tablets (H6). Note that the indicator for duration does not restrict the questionnaire to the non-changing evaluation part as do the other indicators that we use. For duration, we analyze the time it took respondents to complete the entire questionnaire.

In the first part of our analyses, we follow closely the procedure found in Lugtig and Toepoel (2015). First, we report the overall device use for questionnaire completion in the GESIS Panel in 2014. Second, we look at the indicators of measurement and nonresponse error associated with the usage of a particular device. Third, we concentrate on the longitudinal device use and measurement and nonresponse errors.

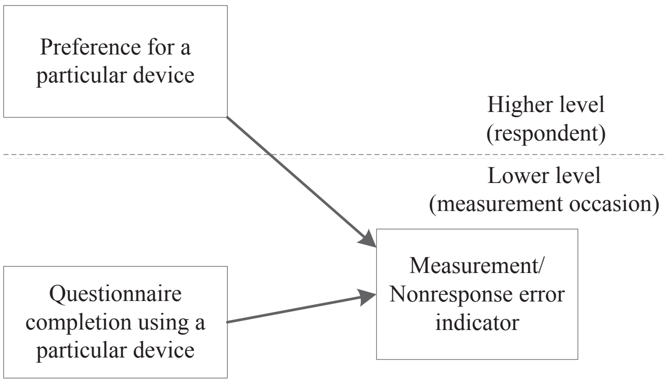


Figure 1 Graphical representation of the two measurement levels

In the second part of our analyses, we attempt to disentangle whether a particular indicator of measurement or nonresponse error is device-related or rather a characteristic of the respondent. For this purpose, for each measurement and nonresponse error indicator we estimate the intercept-only multilevel models, models with indicators of survey completion via tablet or smartphone, and lastly we add respondent characteristics. The intercept-only models do not explain any variance in our dependent variables (i.e., measurement or nonresponse error indicators) but decompose the variance into two independent components for each level (Hox, 2010, p. 15). Our lower level is the measurement occasion (operationalized as each singular survey wave) and respondent is our higher level (see Figure 1).

Measurement occasion is defined as a combination of characteristics of the device that is used to complete the questionnaire and situational characteristics that can be related to the use of this device. The situation characteristics can include distractions, multitasking, changing location, etc. Measurement occasions are nested within respondents. Since different indicators have different scales, we compute logistic models for binary indicators and multilevel regression models for continuous indicators. Adding the device indicators to the models allows us to tease out the device effects from other situational factors that form a measurement occasion. We add respondent characteristics in order to separate the device effects from selection effects. We compare the models based on the intra-class correlation coefficients (ICC), a proportion of the variation at the higher level (respondent) over the total variation (respondent plus measurement occasion).

3 Results

First, we present descriptive results on the device use in the GESIS Panel in 2014. Table 1 shows the absolute counts and proportions of respondents by device as well as transitions from one device to another over the six waves used in our analyses. Most respondents complete the surveys via PCs or laptops, the proportion decreases from 84% in the first wave to 79% in the sixth wave. This indicates an overall increase of mobile device use over time. This result is especially interesting since the online questionnaires in GESIS Panel are not optimized for the completion on mobile devices. The groups who complete the surveys using mobile devices are considerably smaller. The proportion of respondents who complete the surveys via tablets ranges from 7.9 to 10.5%. Smartphone completions have a similar range from 7.6 to 10.5%. The proportions of respondents using tablets for survey completion are about the same as reported by Lugtig and Toepoel (2015) for the LISS Panel, however, the share of respondents who use smartphones to complete panel waves is considerably higher in the GESIS Panel in 2014 than in the LISS Panel in 2013, where it ranged from 1.4 to 3.4%. However, in February 2015 about 6.6% of LISS respondents completed questionnaires via smartphones and about 15.5% of respondents used tablets (Wijnant, 2015). It seems that the differences between the proportions of those completing the surveys via mobile devices in the LISS Panel and in the GESIS Panel can be attributed to the differences in reference periods (i.e., 2013 vs. 2014) and can be explained, for example, by mobile devices becoming more affordable or the public learning to operate such devices.

Transitions from one device to the other are the lowest for PC respondents, ranging from 88.04% (fifth wave to sixth wave) to 90.27% (fourth wave to fifth wave). This result is similar to the results reported by Lugtig and Toepoel (2015) for the LISS Panel in 2013 with less than 5% of respondents switching from PC survey completion to smartphone or tablet.

We calculated the average consistency for each device type. For PC usage, the average device consistency is the highest with 89.09 percentage points. For tablet users, the average device consistency is 67.68 percentage points, ranging from 64.00% to 72.93%. The lowest device consistency is observed for smartphone users: overall, from 58.91 to 61.69% of respondents use smartphone to complete two consecutive waves. The average consistency for smartphone survey completion is 61.46 percentage points. Furthermore, respondents participating via smartphones have higher rates of nonparticipation in the following wave for initial waves. However, these rates become comparable between the devices at later waves (e.g., the fifth and the sixth waves), probably because respondents who participate via smartphones have a higher probability to attrite.[4]

4 In GESIS Panel, after not having participated for three consecutive waves due to either noncontact or nonresponse, participants are excluded from the panel (involuntary attrition). Respondents can also request to be removed from the panel (voluntary attrition).

Table 1 Devices used for questionnaire completion in the six waves of the GESIS Panel (in percent)

		The following wave: Wave x +1					
		PC	Tablet	Smart-phone	Not parti-cipated	N	% of wave respon-dents
First wave	PC	88.86	2.52	3.67	4.95	2342	84.25
2014	Tablet	23.11	64.00	8.89	4.00	225	8.09
(Feb/Mar)	Smartphone	23.94	3.29	61.03	11.74	213	7.66
	Not participated	64.66	4.51	9.77	21.05	—	—
Second wave	PC	89.16	2.73	3.13	4.98	2270	83.00
2014	Tablet	27.31	64.35	2.78	5.56	216	7.89
(Apr/May)	Smartphone	22.49	3.61	65.06	8.84	249	9.10
	Not participated	48.31	4.49	10.67	36.52	—	—
Third wave	PC	89.13	2.92	2.74	5.12	2225	82.38
2014	Tablet	20.18	70.18	4.59	5.05	218	8.07
(Jun/Jul)	Smartphone	23.64	6.98	58.91	10.47	258	9.55
	Not participated	36.79	4.25	6.13	52.83	—	—
Fourth wave	PC	90.27	2.12	3.37	4.24	2168	81.84
2014	Tablet	23.27	66.94	5.71	4.08	245	9.25
(Aug/Sep)	Smartphone	28.39	4.24	60.59	6.78	236	8.91
	Not participated	25.38	3.41	6.82	64.39	—	—
Fifth wave	PC	88.04	3.68	4.00	4.28	2148	81.83
2014	Tablet	14.41	72.93	6.99	5.68	229	8.72
(Oct/Nov)	Smartphone	24.60	7.66	61.69	6.05	248	9.45
	Not participated	22.57	2.43	6.25	68.75	—	—
Sixth wave	PC	—	—	—	—	2050	79.00
2014/2015	Tablet	—	—	—	—	272	10.48
(Dec/Jan)	Smartphone	—	—	—	—	273	10.52
	Not participated	—	—	—	—	—	—
Average	PC	89.09					
device	Tablet		67.68				
consistency	Smartphone			61.46			

N = 2913.

In the second step of our analyses, we report the indicators of measurement and nonresponse error separately for each device type (Table 2). Overall, we observe similar results as Lugtig and Toepoel (2015) that PC respondents report with least measurement and nonresponse error, followed by tablet respondents, and smartphone respondents report with highest measurement and nonresponse error. On average, responses via smartphones are characterized by higher item-nonresponse and a higher percentage of straightlining in a grid question. Those who respond via smartphones respond to an open question at a lower rate and enter fewer characters when they do answer an open question. Also, smartphone respondents demonstrate longer completion times than PC and tablet respondents.

Our hypothesis concerning item-nonresponse predicted higher levels of item-nonresponse for smartphone respondents and no difference for tablet respondents when compared to PC respondents. We indeed observe higher levels of item-nonresponse for smartphones, which is significantly different from PC and tablet respondents. No statistically significant difference is found for the comparison of item-nonresponse between PCs and tablets.

For straightlining, we also expected to find higher levels for smartphones and no differences between tablets and PCs. Straightlining is highest for smartphone completion and the differences to smartphones and tablets are statistically significant (Table 2), again there are no significant differences between tablets and PCs.

In line with our expectations, both smartphone and tablet respondents provide fewer answers to the open question than PC respondents (about 6% for mobile devices vs. 14% for PCs). There is no difference between providing an answer to the open question when using a smartphone or a tablet for survey completion. The length of the answers to an open question is shortest for smartphones and is followed by tablets, although the difference between tablets and smartphones is not statistically significant. The highest number of characters is provided by respondents who complete the surveys via PC or laptop. This finding can be attributed to the absence of the keyboard to type an answer (although we cannot control whether tablet users have used keyboards, it seems a likely explanation).

Regarding the tendency to choose left-aligned answer options in horizontal scales, smartphone respondents do not show a higher rate than PC or tablet respondents. On the contrary, left-aligned options are chosen more by PC and tablet respondents. Our explanation for this finding is that possibly horizontal scrolling is less of an issue with touch screens of smartphones, and zooming might prompt those who respond via smartphones to choose middle categories at a higher rate. However, this hypothetical explanation deserves further investigation. Concerning survey duration, we find the longest completion times for smartphones, followed by tablets. The differences between each pair of devices in survey duration are statistically significant.

To summarize, we find the highest measurement and nonresponse error indicators levels for smartphones. Although some differences between tablets and PCs are found (e.g., in answering an open question and duration), these differences are rather small and for most of measurement and nonresponse error indicators they are not pronounced. It is noteworthy, that although we find several statistically significant differences between PCs and tablets, and all indicators differ on a statistically

Table 2 Measurement and nonresponse error indicators by device in the six waves of the GESIS Panel in 2014

	PC	Tablet	Smart-phone	Total	ANOVA
Mean count of item nonresponse [b,c]	.189	.177	.472	.213	$F(2, 16047)=41.96$, $p<0.001$, $\eta^2=.005$
% Straightlining [b,c]	1.47	1.80	3.86	1.71	$F(2, 15911)=22.04$, $p<0.001$, $\eta^2=.003$
% Answered open question[a,c]	10.10	5.61	5.93	9.33	$F(2, 15937)=25.81$, $p<0.001$, $\eta^2=.003$
Mean number of characters in open question[a,c]	13.925	6.410	4.910	12.458	$F(2, 15937)= 16.53$, $p<0.001$, $\eta^2=.002$
Mean number of chosen left-aligned options[b,c]	2.470	2.418	2.248	2.445	$F(2, 16085)=23.01$, $p<0.001$, $\eta^2=.003$
Mean duration in seconds[a,b,c]	1445.46	1500.27	1862.79	1488.56	$F(2, 16085)=190.65$, $p<0.001$, $\eta^2=.023$

N pooled = 16085, N persons = 2913. Pairwise contrasts are *t*-tests for continuous variables and tests of proportions for percentages with $p<0.01$. a – significant difference PC-Tablet; b – significant difference Tablet-Smartphone; c – significant difference Smartphone-PC.

significant level for the comparison smartphones with PCs, the effect sizes for overall comparisons (in Table 2) are relatively small.

Results presented in Table 2 showing that mobile devices are associated with higher measurement and nonresponse errors can be attributed either to the characteristics of the devices or to the characteristics of the respondents. Those respondents who are more likely to use mobile devices for survey completion might be also more likely to cause higher measurement error. In this case, selection effects and measurement effects are intermingled. Following Lugtig und Toepoel (2015), we compare measurement and nonresponse error indicators for respondents who complete the surveys using one device consistently with measurement and nonresponse error indicators of respondents who switch between devices. If the indicators of measurement and nonresponse errors for those who constantly use tablets or constantly use smartphones for survey completion are larger than for those who switch between the mobile devices, it would indicate that measurement and nonresponse errors are more likely device-related than respondent-related. Table 3 presents the indicators of measurement and nonresponse error for groups of respondents who consistently used one device for survey completion, who switched between two devices, and who used all three device types of devices for survey participation. We restrict the sample to respondents who took part in at least two waves of the panel and thereby had a chance to switch between the devices. From respondents who participated in at least two waves, 67.7% did not switch between the devices and always participated using a PC or a laptop. The proportions of continuous use of a mobile device for survey completion are quite low: 3% of respondents always used tablets and 3.5% always used smartphones for survey completion. About ten percent of respondents used PCs and tablets and about 11.8% used PCs and smartphones. The group of respondents using all three types of devices to complete the surveys was with 2.9% the smallest group.

Table 3 shows that respondents who always use smartphones for survey completion have the highest level of item nonresponse. For those groups that switch between the devices, item nonresponse is highest in groups that involve smartphone completion. Switches between PC and tablet have similar levels of item nonresponse. These findings indicate that item nonresponse is rather device-specific. The indicator for straightlining shows a similar pattern as the indicator for item nonresponse: if switching between devices to complete the surveys involves smartphones or surveys are completed on smartphones exclusively, measurement and nonresponse error indicators are higher than in cases of tablet and PC completion.

Surprisingly, the proportion of respondents who answer the open question is the lowest for those who always complete the surveys using tablets or switch between PCs and tablets. The number of characters entered in an open question is the highest for the groups involving a PC and lowest for groups involving tablets and smartphones. The choice of left-aligned options does not vary much between the groups, and the duration is the highest for groups involving smartphone, except the group in which respondents switch between all three devices to complete the questionnaires.

Overall, from Table 3 we can conclude that as long as survey completion involves smartphones, measurement and nonresponse error indicators are generally higher. However, we cannot draw a conclusion from these results whether reporting with measurement error is due to using a particular device or due to respondent characteristics, since for some indicators (e.g., item nonresponse and straightlining) device properties seem to be one plausible explanation for the decreased data quality and for other characteristics this does not apply.

Following the analysis of Lugtig and Toepoel (2015), we concentrate on cases where respondents participated in two consecutive waves and code the device transitions as well as changes in error indicators for each transition for each respondent. Then we standardize the distributions of changes in wave-to-wave error indicators, because the indicators have different scales. If the device is the cause of higher nonresponse and measurement error, then for transitions involving device switches the standardized changes in measurement and nonresponse error indicators would not

Table 3 Measurement and nonresponse error indicators across groups of device use patterns

	No device switches			Switch between two devices			Switch between three devices	
	Always PC	Always Tablet	Always Smart-phone	PC & Tablet	PC & Smart-phone	Tablet & Smart-phone	PC, Tablet & Smart-phone	Total
Mean count of item nonre-sponse	.192 (.013)	.149 (.051)	.560 (.142)	.151 (.023)	.448 (.065)	.387 (.206)	.264 (.090)	.234 (.014)
Mean % straightlining	1.49 (.001)	1.76 (.010)	3.62 (.010)	1.31 (.005)	2.97 (.006)	4.02 (.030)	2.47 (.009)	1.78 (.001)
Mean % An-swered open question	10.55 (.005)	4.55 (.012)	8.10 (.018)	8.49 (.011)	6.03 (.007)	4.41 (.019)	5.99 (.015)	9.34 (.003)
Mean number of characters in open question	14.498 (1.061)	4.722 (1.694)	5.258 (1.477)	10.689 (1.647)	7.918 (1.572)	3.275 (1.472)	5.029 (1.577)	12.323 (.767)
% Choice of left-aligned options	.275 (.002)	.260 (.010)	.260 (.010)	.287 (.005)	.264 (.005)	.260 (.019)	.255 (.010)	.273 (.002)
Mean duration	1825 (28.9)	1960 (189.8)	3069 (227.6)	1718 (64.4)	2045 (79.1)	2096 (266.6)	1848 (90.3)	1891 (25.3)
Sample size	1918	85	99	282	333	34	81	2832

N = 2832 since 81 observations who participated in only one wave were dropped, standard errors in parentheses.

be different from zero for the groups with transitions to the same device (PC-PC, tablet-tablet, and smartphone-smartphone), while we would expect to find significant differences for groups which involve device changes, especially smartphones. The results are presented in Table 4. Overall, there are 12,598 transitions with nonmissing indicators of measurement error. In line with our expectations, the transitions involving the same device (i.e., PC-PC, tablet-tablet, smartphone-smartphone) are not associated with significant changes in nonresponse and measurement error indicators. Moreover, the magnitude of the changes in standardized nonresponse and measurement error indicators for transitions without the device switches is small. For the groups involving device switches the most pronounced differences are found in duration: the differences are significant for all transitions with devices switches and for groups involving smartphones the magnitude of the change is considerably larger than for transitions between tablets and PCs. Significant effects are also found for the switches PC→tablet and PC→smartphone in providing answers to the open question and for groups PC→smartphone and smartphone→PC for the choice of left-aligned answer options. The magnitude of these changes, however, is rather small. The manner in which changes in standardized indicators are calculated makes them correspond to standardized mean difference effect sizes (Lipsey & Wilson, 2001, p. 198), so we use the benchmarks provided by Cohen (1992) to interpret the values from Table 4. Overall, we see moderate effects for duration in groups involving smartphones and small effects for duration, tendency to answer the open question and to choose left-aligned options in some groups. Our results are in line with Lugtig and Toepoel (2015), who find that transitions between tablets and PCs show small changes while transitions between smartphones and PCs show the largest changes in measurement indicators, although not significant possibly due to small group sizes. Significant changes were found by Lugtig and Toepoel (2015) for straightlining for transition tablet-tablet and the number of choices made in check-all-that-apply questions (for groups PC-PC, tablet-PC, and smartphone-PC) as well as questionnaire evaluation (for tablet-PC and smartphone-PC).[5]

The analysis presented in Table 4 is based on transitions between the waves, and it controls for respondent characteristics insofar that they stay the same over time while respondents switch between devices. We extend this analysis with multilevel modeling, in which we explicitly control for device effects and respondent characteristics. Since different indicators of nonresponse and measurement error are studied, ideally the models need to include the predictors of reporting with higher levels of item nonresponse, straightlining, or taking longer to complete the surveys, etc. This would make difficult comparing the models with each other. Thus, we use respondent characteristics that were shown to relate to the propensity of responding using a particular device. Since our goal here is not to explain which respondents produce higher nonresponse or higher measurement error but rather to tease out the device effects, this approach seems feasible.

In Table 5, the results of the stepwise procedure of calculating the multilevel models are presented. For this analysis we only include respondents who completed the survey without a break or completed after a break and have no missing values on the dependent variables to be able to compare the models with each other. First, intercept-only multilevel models are presented. The intra-class correlation coefficients (ICCs), the proportion of variance located at the level of the respondent to the total variance (i.e., respondent plus measurement occasion) for the empty models

5 The groups tablet-smartphone and smartphone-tablet were excluded by Lugtig and Toepoel (2015) due to small group sizes.

Table 4 Change in standardized indicators of nonresponse and measurement error associated with different device switches

Group/ Indicator	Item non-response	Straight-lining	Answered open question	Number of characters	Choice of left-aligned options	Duration	N
PC-PC	.000	.003	-.001	-.001	-.003	.001	9824
Tablet-Tablet	.023	-.047	.009	.005	.030	-.056	759
Smartphone-Smartphone	-.016	-.031	.032	.014	.031	.009	704
PC-Tablet	-.078	-.013	-.113*	-.044	-.105	.113*	308
Tablet-PC	.038	.041	.053	.044	.031	-.211**	238
Smartphone-Tablet	-.109	-.121	-.036	-.036	-.049	-.615***	59
Tablet-Smartphone	.114	.086	.085	.108	-.023	.658***	63
PC-Smartphone	.030	.093	-.105*	-.059	-.119*	.689***	358
Smartphone-PC	-.010	-.036	.110	.064	.190**	-.737***	285

N (person-waves) = 12598, N respondents = 2770 (only observations for respondents who took part in two consecutive waves are included); the values are predicted marginal means, significance tests against zero. *p<.05, **p<.01,*** p<.001

show that for some indicators the measurement occasion which includes but is not limited to a device, is more influential and for other indicators the differences are between-person differences.

For item nonresponse model, the intra-class correlation (ICC) of 0.172 means that item-nonresponse is a characteristic of the situation rather than a tendency of a respondent to skip questions. The differences in straightlining (ICC = .754) are rather individual-level differences than the characteristic of the survey situation: some respondents tend to straightline and some do not irrespective of the survey situation. We cannot definitely say that the differences in providing answers to the open question also are individual-level differences rather than the characteristic of the survey situation judging by the intra-class correlation of 0.551. Providing an answer to an open question seems to depend both on respondent preference and on the survey situation. The larger amount of variance for the choice of left-aligned options is located on the level of the measurement occasion, suggesting that choosing left-aligned options at horizontal scales is not a respondent-specific characteristic. Survey duration is as well situation specific, which is consistent with the results presented in Tables 2 and 3.

Table 5 Multilevel models for indicators of measurement and nonresponse error

	Item nonresponse	Straightlining	Answered open question	Choice of left-aligned options	Duration
Null models					
Constant	.124*** (.006)	-7.390*** (.608)	-3.510*** (.084)	2.469*** (.084)	24.736*** (.164)
Variance at higher level	.052 (.003)	10.782 (2.774)	4.038 (.302)	.406 (.016)	54.356 (2.028)
Variance at lower level	.251 (.003)	3.290† (—)	3.290† (—)	.999 (.012)	111.978 (1.398)
ICC	.172	.766	.551	.289	.327
Models with device dummies (reference: PC completion)					
Constant	.127*** (.006)	-6.570*** (.444)	-3.369*** (.085)	2.491*** (.015)	23.804*** (.171)
Tablet completion	-.033 (.018)	.031 (.339)	-.763*** (.177)	-.060 (.039)	1.552*** (.421)
Smartphone completion	.008 (.018)	1.255*** (.238)	-.711*** (.177)	-.187*** (.038)	8.716*** (.410)
Variance at higher level	.052 (.003)	6.220 (1.741)	3.947 (.297)	.405 (.016)	53.534 (1.993)
Variance at lower level	.251 (.003)	3.260†† (—)	3.254†† (—)	0.998 (.012)	108.611 (1.357)
ICC	.171	.656	.548	.289	.330
Models with device dummies (reference: PC completion) and respondent characteristics					
Constant	.279*** (.033)	-4.414*** (.497)	-3.303*** (.314)	2.141*** (.081)	29.369*** (.887)
Tablet completion	-.025 (.018)	.058 (.323)	-.694*** (.177)	-.060 (.039)	1.862*** (.417)
Smartphone completion	.028 (.018)	.911*** (.239)	-.410* (.179)	-.159*** (.039)	9.498*** (.414)
Gender (male)	-.015 (.012)	.319 (.205)	.051 (.112)	-.049 (.029)	.084 (.319)
Age (centered)	.003*** (.001)	-.038*** (.008)	.032*** (.004)	.004*** (.001)	.133*** (.012)
Education middle	-.070*** (.019)	-.549 (.281)	-.063 (.182)	.182*** (.047)	-1.028* (.515)
Education high	-.088*** (.018)	-1.399*** (.284)	.330 (.172)	.214*** (.045)	-1.556** (.492)
German	-.080** (.029)	-.856* (.397)	-.179 (.268)	.184** (.070)	-4.230*** (.766)

	Item nonresponse	Straightlining	Answered open question	Choice of left-aligned options	Duration
Living alone	.002	-.110	.308*	.045	.324
	(.017)	(.295)	(.152)	(.041)	(.449)
In paid work	.002	-.213	-.282*	.016	-.762*
	(.014)	(.231)	(.127)	(.034)	(.369)
Online survey experience	-.022	-.726*	.199	.010	-.229
	(.015)	(.295)	(.138)	(.037)	(.400)
Variance at higher level	.049	4.732	3.707	.396	48.926
	(.003)	(.800)	(.287)	(.016)	(1.865)
Variance at lower level	.252	3.260 ††	3.254††	.998	108.544
	(.003)	(—)	(—)	(.012)	(1.355)
ICC	.163	.620	.533	.284	.311

N (person-waves) = 15623, N (respondents) = 2793; coefficients are betas, ICC short for intra-class correlation, the ICC values higher than 0.5 mean that more variance is located at the higher level; standard errors in parentheses; **p*<.05, ***p*<.01,*** *p*<.001, † Note that for the logistic models the variance at the lower level is fixed at $\pi^2/3$ (Hox 2010: 128), which equals approximately 3.290. ††rescaled variance to compare logistic models with each other, coefficients are also rescaled – all using meresc Stata command, ICC for the models calculated with rescaled variances. Duration was rescaled to minutes to avoid estimation problems. We excluded the number of characters in open question since it is conditional on providing an answer to an open question and due to estimation problems.

In the second step of our analysis we add the device dummies for tablet and smartphone completion (reference: PC completion) to the models to tease out device effects from other factors forming measurement occasion.[6]

Adding device indicators does not considerably lower the intra-class correlation coefficients, however, significant device effects are found. Completing online surveys using tablets is associated with fewer answers provided in open-ended questions and longer duration, which are statistically significant. Smartphone completion shows significant effects for all indicators with the exception of item nonresponse. Completing surveys with smartphones is associated with higher straightlining, providing fewer answers to the open-ended question, providing fewer

6 Device dummies indicate whether a respondent completed a questionnaire via smartphone, tablet, or PC for each of the measurement occasions. We constrain the effects of the devices to be equal at each measurement occasion since we expect that the content of the survey evaluation items does not influence the indicators of nonresponse and measurement error that we use. One exception is the duration that is a measure for the whole questionnaire. Since the inclusion of measurement occasion dummies did not substantially change the effects of the devices or respondent characteristics, but led to difficulties in the rescaling process for the logistic models, the measurement occasion dummies are not included in the final analysis.

characters in the open-ended question, increased choice of the left-aligned options, and longer duration.

In the final step, we control for respondent characteristics. Adding the respondent characteristics reduces the intra-class correlation coefficients substantially, especially for the indicators of straightlining and providing answers to the open question. We do not observe any implausible results with respect to respondent characteristics. For example, gender shows no significant effects, and it is plausible to not expect differences in data quality indicators we use based on gender. The effects of age are also not contra-intuitive: older respondents tend to provide data of better quality, generating lower item nonresponse, showing lower rates of straightlining, providing answers to an open question and more characters in the responses to open-ended questions. Taking longer to answer online surveys is also plausible. The higher levels of education (lower education being a reference category) are associated with lower likelihood of item nonresponse, straightlining, providing longer answers to the open-ended question and shorter duration, all of which could be the result of higher cognitive abilities. One rather puzzling indicator is the choice of left-aligned options with higher educated and older respondents showing increased choice of left-aligned options.

However, our focus in this analysis is less on the respondent characteristics but rather on their influence on effects of devices used for survey completion. The device effects found in models with device dummies are significant in the models with respondent characteristics. Using tablets for survey completion is associated with lower likelihood to provide answers to an open questions and longer duration. Those who complete surveys on smartphones show more straightlining, are less likely to answer an open question, provide shorter answers to an open question, and show longer duration when controlling for respondent characteristics. Overall, the results of multilevel models signify that completion of the survey on a mobile device has adverse consequences for data quality, especially when smartphones are used. Some indicators of nonresponse and measurement error are more affected than others: for example, the effects are largest for duration, but item nonresponse does not show significant results. Furthermore, the effects of completion of the online surveys using a mobile device cannot be fully explained by the choice of this device by the respondents.

4 Conclusions and Discussion

In this article, we study whether survey completion of online surveys using smartphones and tablets leads to higher measurement and nonresponse errors than when surveys are completed using personal computers or laptops. The analyses replicate and extend the approach chosen by Lugtig and Toepoel (2015), who show that smartphone survey completion leads to a higher measurement error. In the GESIS Panel, a probability-based mixed-mode panel, the data source for our analyses, more respondents use smartphones for survey completion than in the LISS Panel, a probability-based online panel the data from which is used by Lugtig and Toepoel (2015).

We find that PCs prevail for completion of the online surveys, however, the average consistency for smartphones is about 60%, indicating that if the respondent completes one survey on a smartphone, on average, in 60% of the cases she will complete the next survey on a smartphone as well. For tablets, the average consistency is about 70%. Moreover, there is a slight increase in the proportion of respondents who use mobile devices for survey completion in the course of the six waves. Given that the GESIS Panel questionnaires are not optimized for mobile survey completion, studying the influence of mobile device use for survey completion on data quality is especially important.

We find that most of the indicators of measurement and nonresponse error are higher for mobile devices than for PCs. Online survey completion using smartphones shows higher item nonresponse, higher levels of straightlining in a grid question, lower rate of responding to an open question, and for those who do answer an open question providing shorter answers, as well as longer completion times compared to PC-completion. The differences found between smartphones and PCs are larger than the differences found between tablets and PCs, which is consistent with the results of previous research indicating that PCs and tablets lead to comparable results regarding data quality. For groups of respondents who switch between devices, the highest levels of measurement and nonresponse errors are found in groups, which involve smartphones. Nonetheless, the magnitude of the differences in measurement and nonresponse error indicators for various devices is rather small with the exception of survey duration with both tablet and smartphone respondents taking considerably longer to complete the surveys.

For the LISS Panel, Lugtig and Toepoel (2015) find that measurement errors do not increase when respondents switch from one device to the other. They conclude based on this finding that reporting with measurement error is a respondent-related characteristic. Our analysis of wave-to-wave device transitions shows significant effects in providing fewer answers to an open question for switches from PCs to smartphones or tablets, which is probably due to the absence of the keyboard, increased choice of left-aligned answer options in horizontal scales when switching from smartphone to PC, decreased choice of left-aligned answer options for the switch PC-smartphone, and longer duration for switches from PC to either mobile device. Changes in standardized nonresponse and measurement error indicators such as item nonresponse, straightlining, number of characters in open question are not significant. However, based on the multilevel analysis – an extension to the study we aimed to replicate – only item nonresponse is not predicted by tablet or smartphone completion. Other indicators of nonresponse and measurement error that we use are affected by the device on which the survey is completed and cannot be attributed to the respondent since we control for respondent characteristics. The results of the multilevel models with device indicators differ somewhat from the replication of Lugtig and Toepoel (2015) analysis. This may be due to the fact that for wave-to-wave transitions only those transitions between two consecutive waves are considered, whereas for multilevel models the basis for analysis are all observations for respondents who took part in at least two waves, meaning respondents could potentially switch devices but did not have to participate consecutively, While the main focus of wave-to-wave analysis is the replication of the strategy chosen by Lugtig and Toepoel (2015), we did not want to exclude respondents who did not switch consecutively between the waves thereby losing the information in multilevel models.

Other reasons why our results only partially align with the results of Lugtig and Toepoel (2015) can be multiple. First, although we also use the evaluation part of the questionnaire so that the content stays the same across the waves, the content of the questions varies between the LISS Panel and the GESIS Panel. Thus, using exact same indicators of nonresponse and measurement error based on the same questionnaire content would be desirable. Another reason for the differences we find might be that the LISS Panel exists longer than the GESIS Panel, so that panel attrition or panel conditioning might be causes of the differences. If respondents who prefer completing surveys on mobile devices attrite at a higher rate and/or respondents who are longer with the panel learn to use mobile devices to report with fewer errors, fewer negative effects on the data quality will be found in the LISS Panel than in the GESIS Panel. This point warrants further investigation. Ideally, two panels existing for the same amount of time should be compared, but this is difficult to realize in practice.

Furthermore, our study is not free from limitations. First, our study does not assign the respondents randomly to a device, which limits our possibilities in studying nonresponse to item nonresponse only. Second, we do not have validation data and can only assess measurement errors using indirect indicators (of satisficing) such as straightlining, choosing left-aligned answer options in horizontal scales, survey duration. Nonetheless, our study provides a robustness check for the results obtained in a probability-based online panel in the Netherlands and extends the replication by including the respondent characteristics. Ideally, to separate selection effects one would use an experimental design. However, in the context of large-scale population surveys it is practically not feasible and studies that assign respondents to the device are confronted with the issues of respondent noncompliance (de Bruijne & Wijnant, 2013; Mavletova, 2013; Wells, Bailey, and Link, 2014) when some respondents complete the survey on their preferred device rather than the device to which they were assigned. One solution to this problem would be to match respondents on a set of observable characteristics while the devices used for survey completion differ. We could not use this design as the groups completing the surveys on smartphones and tablets are still rather small, but given their rapid growth future studies should explore this option.

What are the practical implications of our analyses based on the results we obtained using the GESIS Panel data? The answer to the question whether survey completion using mobile devices is a problem that survey researchers should be concerned about is yes. Completing surveys with mobile devices, especially smartphones, is problematic. However, our analyses also indicate that for the most part the magnitude of these problems is not large: we find small to moderate effects. Although we cannot provide a definite answer to the question of how should survey designers deal with unintended mobile respondents since our findings are based on observational data, for the moment for GESIS Panel we do not see the need to address the issue of unintended mobile respondents based on the indicators that we use in this article. One notable exception is survey duration: for surveys in time-sensitive situations researchers need to investigate design options such as mobile optimization together with its consequences for data quality.

References

Antoun, C. (2015). *Effects of Mobile versus PC Web on Survey Response Quality: a Crossover Experiment in a Probability Web Panel*. Paper presented at the Annual Conference of the Americal Association for Public Opinion Research, Hollywood, FL.

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*. doi: DOI: 10.1177/1525822X14526146

Callegaro, M. (2010). Do You Know Which Device Your Respondent Has Used to Take Your Online Survey? *Survey Practice, 3*(6), 12.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 111, 155-159.

Cook, W. A. (2014). Is Mobile a Reliable Platform For Survey Taking? Defining Quality in Online Surveys From Mobile Respondents. *Journal of Advertising Research*. doi: DOI: 10.2501/JAR-54-2-141-148

Couper, M. P., & Peterson, G. (2015). *Exploring Why Mobile Web Surveys Take Longer*. Paper presented at the General Online Research Conference 19.03.2015, Cologne.

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. (2014a). Improving response rates and questionnaire design for mobile web surveys. *Public Opinion Quarterly, 78*(4), 951-962.

de Bruijne, M., & Wijnant, A. (2014b). Mobile Response in Web Panels. *Social Science Computer Review, 32*(6), 728-742. doi: 10.1177/0894439314525918

Hilligus, D. S., Jackson, N., & Young, M. (2014). Professional respondents in nonprobability online panels. In M. Callegaro, P. J. Lavrakas, J. A. Krosnick, R. Baker, J. D. Bethlehem, & A. S. Göritz (Eds.), Online panel research: A data quality perspective (pp. 219-237). New York: Wiley.

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications*. New York: Routledge.

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

Lipsey, M. W. & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, California: Sage.

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, Online first, February 26, 2015*. doi: 10.1177/0894439315574248

Malhotra, N. (2008). Completion Time and Response Order Effects in Web Surveys. *Public Opinion Quarterly, 72*, 914-934.

Matthijsse, S. M., de Leeuw, E. D., & Hox, J. J. (2015). Internet panels, professional respondents, and data quality. *Methodology, 11*(3), 81-88. doi: 10.1027/1614-2241/a000094

Mavletova, A. (2013). Data Quality in PC and Mobile Web Surveys. *Social Science Computer Review, 31*, 725-743. DOI: 10.1177/0894439313485201

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, V. C., Crawford, S. D., & Dungan, J. P. (2012). *Use of mobile devices to access computer-optimized web instruments: Implications for respondent behavior and data quality*. Paper presented at the Annual Conference of the American Association for Public Opinion Research, Orlando, FL.

Peterson, G. (2012). *Unintended mobile respondents*. Paper presented at the CASRO Technology Conference 31.05.2012, New York City.

Peytchev, A., & Hill, C. (2010). Experiments in Mobile Web Survey Design: Similarities to Other Modes and Unique Considerations. *Social Science Computer Review, 28*, 319-335.

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*(2).

Rossmann, J., & Gummer, T. (2014). *Stata-ado package "PARSEUAS: Stata module to extract detailed information from user agent strings"*. Retrieved from: http://fmwww.bc.edu/repec/bocode/p/parseuas.ado

Stapleton, C. E. (2013). The Smartphone Way to Collect Survey Data. *Survey Practice, 6*(2).

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*, 544-560.

Tourangeau, R. (1984). Cognitive sciences and survey methods. In T. Jabine, M. Straf, J. Taur & R. Tourangeau (Eds.), *Cognitive aspects of survey methodology: building a bridge between disciplines* (pp. 73-100). Washington DC: National Academy Press.

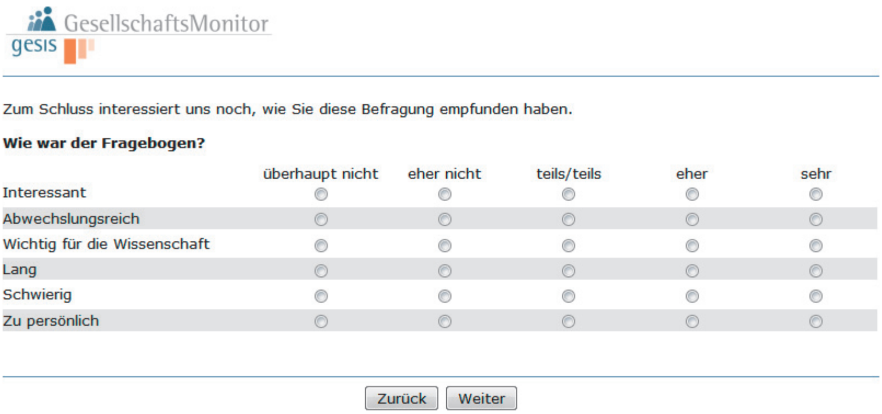
Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.

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. doi: 10.1177/0894439313505829

Wijnant, A. (2015). Mobile devices in a web panel: what are the results of adjusting questionnaires for smartphones and tablets. Paper presented at the 6th Conference of the European Survey Research Association 14.07.2015, Reykjavik, Iceland.

Appendix

Screenshots of the questions used for analysis with translations.


GesellschaftsMonitor
gesis

Zum Schluss interessiert uns noch, wie Sie diese Befragung empfunden haben.

Wie war der Fragebogen?

	überhaupt nicht	eher nicht	teils/teils	eher	sehr
Interessant					
Abwechslungsreich					
Wichtig für die Wissenschaft					
Lang					
Schwierig					
Zu persönlich					

Zurück Weiter

Question text: Finally, we are interested how do you feel about the questionnaire. How was the questionnaire? *Items*: interesting, diverse, important for science, long, difficult, too personal. Scale: not at all, rather not, partly, rather yes, very. (*)


GesellschaftsMonitor
gesis

Hat die Befragung Sie zum Nachdenken angeregt?

überhaupt nicht | eher nicht | teils/teils | eher | sehr

Zurück Weiter

Question text: Did the survey encourage you to think about things? *Scale*: not at all, rather not, partly, rather yes, very. (*)


GesellschaftsMonitor
gesis

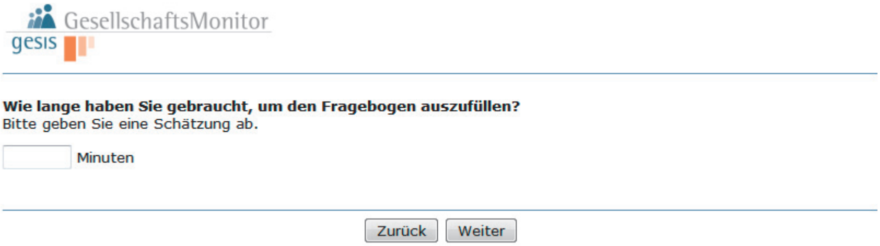
Waren die Fragen insgesamt verständlich?

überhaupt nicht | eher nicht | teils/teils | eher | sehr

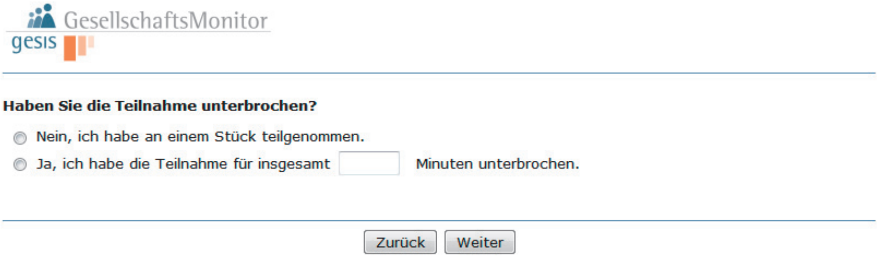
Zurück Weiter

Question text: Were the questions sufficiently clear? *Scale*: not at all, rather not, partly, rather yes, very.

Question text: Overall, how did you like the survey? *Scale*: not at all, not so good, moderately, good, very good.



Question text: How long did it take you to complete the questionnaire? Please provide an estimation. __ minutes.



Question text: Did you interrupt your participation?

No, I completed the survey at once.

Yes, I took a break for … minutes.

GesellschaftsMonitor
gesis

Waren Sie bei der Beantwortung der Fragen allein oder waren weitere Personen anwesend?

- Ich war allein.
- Andere Personen waren anwesend.

Von wo aus haben Sie an dieser Befragung teilgenommen?

- Von Zuhause
- An einem anderen Ort

Mit welchem Gerät haben Sie die Fragen beantwortet?

- PC bzw. Laptop
- Tablet-PC
- Smartphone
- Anderes Gerät, und zwar:

Zurück Weiter

Question text: Were you alone or were other persons present while you were answering the questions?

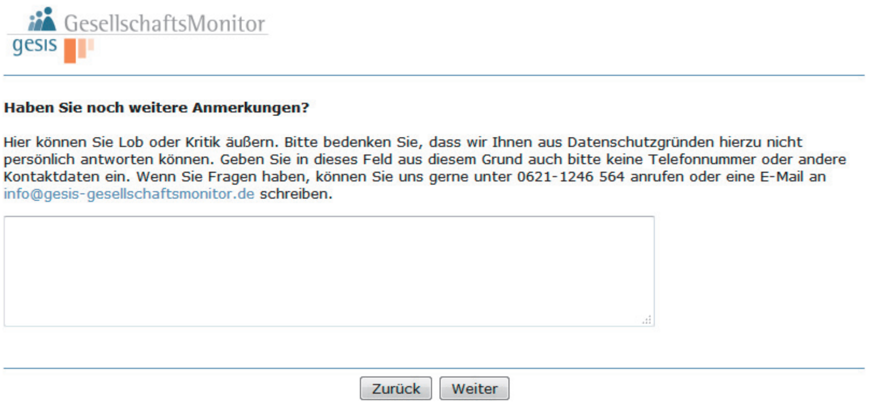
I was alone
Other persons were present

From what location did you participate in this survey?

From home
From another place

What type of device did you use to answer the questions?

PC or Laptop
Tablet-PC
Smartphone
Other device, namely:



GesellschaftsMonitor
gesis

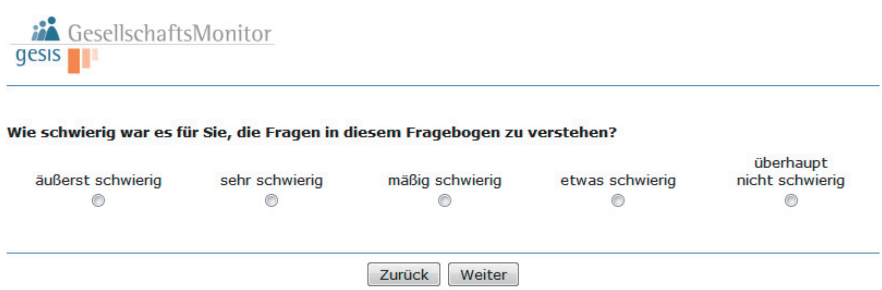
Haben Sie noch weitere Anmerkungen?

Hier können Sie Lob oder Kritik äußern. Bitte bedenken Sie, dass wir Ihnen aus Datenschutzgründen hierzu nicht persönlich antworten können. Geben Sie in dieses Feld aus diesem Grund auch bitte keine Telefonnummer oder andere Kontaktdaten ein. Wenn Sie Fragen haben, können Sie uns gerne unter 0621-1246 564 anrufen oder eine E-Mail an info@gesis-gesellschaftsmonitor.de schreiben.

Zurück Weiter

Translation: Do you have any further remarks?
Here you can express praise or critique. Please be aware, that we are not able to react to your comments due to data protection regulations. For these reasons, please do not write your telephone number or other contact information . If you have questions, you can call us on 0621-1246 564 or write us an email to info@gesellschafts-monitor.de.

(*) Two items that were used for waves 3 to 6 instead of the two items that directly follow the evaluation matrix (marked with an asterisk):



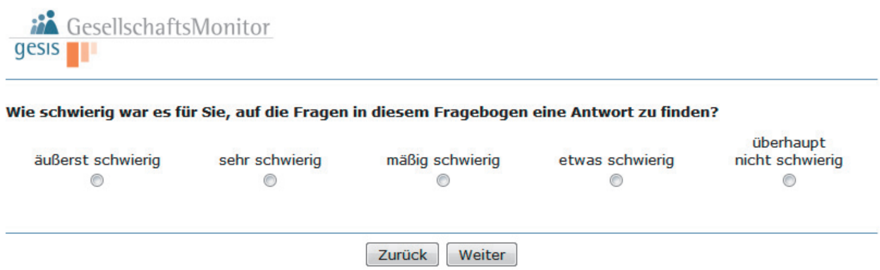
GesellschaftsMonitor
gesis

Wie schwierig war es für Sie, die Fragen in diesem Fragebogen zu verstehen?

äußerst schwierig	sehr schwierig	mäßig schwierig	etwas schwierig	überhaupt nicht schwierig
○	○	○	○	○

Zurück Weiter

Question text: How difficult was it for you to interpret the meanings of the questions in this questionnaire? *Scale*: Extremely difficult, very difficult, moderately difficult, slightly difficult, not difficult at all



Question text: How difficult was it for you to generate your answers to the questions in this questionnaire? *Scale*: Extremely difficult, very difficult, moderately difficult, slightly difficult, not difficult at all.

Equations for multilevel models:

Empty models for dependent variables "straightlining" and "answered open question":

$$\text{logit}(\pi_{ij}) = \gamma_{00} + u_{0j}$$

Empty models for dependent variables "item nonresponse", "choice of left-aligned options", "duration":

$$Y_{ij} = \gamma_{00} + u_{0j} + e_{ij}$$

Models with device dummies for dependent variables "straightlining" and "answered open question":

$$\text{logit}(\pi_{ij}) = \gamma_{00} + \gamma_1 tablet_{ij} + \gamma_2 smartphone_{ij} + u_{0j}$$

Models with device dummies for dependent variables "item nonresponse", "choice of left-aligned options", "duration":

$$Y_{ij} = \gamma_{00} + \gamma_1 tablet_{ij} + \gamma_2 smartphone_{ij} + u_{0j} + e_{ij}$$

Models with device dummies and respondent characteristics for dependent variables"straightlining" and "answered open question":

$$\begin{aligned} logit(\pi_{ij}) = {} & \gamma_{00} + \gamma_1 tablet_{ij} + \gamma_2 smartphone_{ij} + \gamma_3 gender_{ij} + \gamma_4 age_{ij} + \gamma_5 mid\,education_{ij} \\ & + \gamma_6 high\,education_{ij} + \gamma_7 german_{ij} + \gamma_8 living\,alone_{ij} + \gamma_9 in\,paid\,work_{ij} \\ & + \gamma_{10} online\,survey\,experience_{ij} + u_{0j} \end{aligned}$$

Models with device dummies and respondent characteristics for dependent variables "item nonresponse", "choice of left-aligned options", "duration":

$$
\begin{aligned}
Y_{ij} = {} & \gamma_{00} + \gamma_1 tablet_{ij} + \gamma_2 smartphone_{ij} + \gamma_3 gender_{ij} + \gamma_4 age_{ij} + \gamma_5 mid\, education_{ij} \\
& + \gamma_6 high\, education_{ij} + \gamma_7 german_{ij} + \gamma_8 living\, alone_{ij} + \gamma_9 in\, paid\, work_{ij} \\
& + \gamma_{10} online\, survey\, experience_{ij} + u_{0j} + e_{ij}
\end{aligned}
$$

where i is the lowest level (measurement occasion) and j is the highest level (respondent)