Device effects on survey response quality. A comparison of smartphone, tablet and PC responses on a cross sectional probability sample

Sanne Lund Clement | Majbritt Kappelgaard Severin-Nielsen | Ditte Shamshiri-Petersen

Abstract : The increasing use of web surveys and different devices for survey completion calls for the examination of device effects on survey response quality. Whereas most existing studies are based on web panels, subgroups (e.g., students), or short questionnaires designed for device experiments, which compels participants to respond through specific devices, this study is based on two large, nationally representative cross-sectional samples (ISSP 2018 and ISSP 2019) in which the completion device was chosen by the respondent. Seven indicators of response quality are applied, which allows comparison among survey participants answering the questionnaire on a smartphone, tablet, or PC. The results are in line with previous findings: Respondents' self-evaluated engagement in survey completion does not differ across devices, and only small, non-systematic differences between devices on satisficing indicators, such as the tendency to agree regardless of question content (acquiescence), non-substantive answers, selection of mid-point response options and primacy effects, and straightlining are identified. Controlling the associations between response device and response quality indicators for self-selection biases did not change the overall result.

Introduction

Over the past decade, technological developments have allowed survey participants to complete web surveys on a wider range of devices. Internet-enabled, handheld mobile devices such as smartphones and tablets are increasingly used for this purpose, contesting the use of personal computers (PCs) (e.g., (Revilla, 2017: 267; Couper, Antoun & Mavletova, 2017). The increasing number of devices for data collection potentially provides participants with greater availability and flexibility, which could in turn also potentially help to stave off declining response rates. However, questionnaire completion on mobile devices has also triggered concerns over negative implications for data quality (Couper, Antoun & Mavletova, 2017: 133-134; Antoun, 2015: 100; Groves et al., 2009: 186-188). The most prominent concerns in this regard relate to mobile, handheld devices promoting uncommitted survey response behaviour, such as satisficing (cf. Krosnick, 1991). First, survey design and layout that is not optimized for mobile devices with smaller screens may render survey completion far more burdensome. Second, the

same applies to unfamiliarity with mobile devices; survey participants who are unaccustomed to operating smartphones or tablets may find survey response difficult, leading to incorrect answers or incomplete data. While being able to adapt survey design to multiple devices and allowing participants to select the most suitable device for themselves, survey researchers have no influence on the context of the survey completion. There are significant differences in the patterns of device use (Antoun, Katz, Argueta & Wang, 2018; Couper, Antoun & Mavletova, 2017; Deng et al., 2019; Helles, 2016; Wells, Bailey & Link, 2013). Whereas PCs are used most often at home or work, handheld mobile devices are also often used while in the public space and 'on the go', which possibly involves additional distractions. Thus, a third element of concern is that mobile devices are applied in contexts not conducive to committed, thorough response behaviour.

Whereas device effects regarding nonresponse are well-documented (Couper, Antoun & Mavletova, 2017: 240), most studies do not find consistent support for differences in the quality of participants' answers (Couper, Antoun & Mavletova, 2017: 142-143). Some studies do, however, point to mobile devices causing poorer response quality than PCs regarding the inclination to provide short answers to open-ended questions and question skipping (Mavletova, 2013: 737; Struminskaya, Weyandt & Bosnjak, 2015: 281), and other studies find that PCs elicit poorer responses regarding the tendency to straightline (Keusch & Yan, 2017: 759; Lugtig & Toepoel, 2016: 88; Mavletova, 2013: 737); that is, the tendency to 'give identical (or nearly identical) answers to items in a battery of questions using the same response scale' (Kim et al., 2019: 214).

However, most device effect studies have been based on web panels (e.g., Struminskaya, Weyandt & Bosnjak, 2015; Lugtig & Toepoel, 2016; de Bruijne & Wijnant, 2013; Antoun, Couper & Conrad, 2017) and short questionnaires designed for the purpose of testing device effects (e.g., Andreadis, 2015; Keusch & Yan, 2017; Schlosser & Mays, 2017; Toepoel & Lugtig, 2014). Furthermore, most recent studies apply the experimental design in order to eliminate self-selection effects in results (e.g., Antoun, Couper & Conrad, 2017; Keusch & Yan, 2017; Mavletova, 2013; Tourangeau et al., 2018).

Based on two nationally representative cross-sectional samples of the Danish adult population conducted as part of the International Social Survey Programme (ISSP), the main aim of this study is to examine whether web surveys completed on handheld mobile devices are of lesser quality compared to surveys completed on PCs. The study separates itself from most existing studies in that the samples are both nationally representative and free of potential panel effects. Furthermore, having an autonomous scientific purpose in addition to studying device effects, the questionnaires are authentic with respect to theme and length. Finally, in both samples respondents could choose a device on which to complete the survey. Thus, a high level of ecological validity is a strength of the present study; it contributes to knowledge of device effects in a setting that resembles the setting for most social research surveying and prompts authentic survey respondent behaviour. Moreover, the replication of results using two sets of data improves the reliability.

Previous studies on device effects in web surveys

Due to novelty, studies of device effects in web surveys are rather sparse, most having been conducted in recent years (e.g., Couper, Antoun & Mavletova, 2017; Antoun, Couper & Conrad 2017; Keusch & Yan 2017; Schlosser & Mays 2017; Tourangeau et al., 2018). The established knowledge regarding device effects concerns how different population groups are prone to prefer certain devices. Conversely, knowledge of participants' survey response behaviour when using the devices is more limited, and the results vary somewhat.

It is well-established that survey participants who complete surveys on mobile devices differ in socio-demographic characteristics from those who use PCs. Participants completing surveys on smartphones are usually younger than PC users and better educated, and more women and ethnic minorities tend to use smartphones for survey completion (Antoun, 2015: 114; Keusch & Yan, 2017: 751; Sommer, Diedenhofen & Musch, 2017: 378). This 'device divide' (cf. Antoun, 2015: 103) implies a self-selection bias when the response device is chosen by the respondent. It is also established that response rates are lower and break-off rates higher for smartphone respondents compared to PC respondents (Mavletova, 2013; Mavletova & Couper, 2013; Sommer, Diedenhofen & Musch, 2017). In a literature review based on 13 studies, Couper, Antoun, and Mavletova (2017) found the average break-off rate to be approximately 13% for mobile devices, whereas the average break-off rate was close to 5% for PCs (p. 140). In addition, in a recent experimental study among iPhone owners in which respondents who started the survey on an iPhone were allowed to complete the survey on it, whereas respondents who started the survey on a PC were free to choose either to proceed to complete the survey on the PC or switch to an iPhone, Keusch and Yan (2017) found that almost 10% of smartphone participants skipped at least one question, whereas the figure was only 3.6% for the PC participants (pp. 758-759).

However, indications of participants completing surveys on handheld mobile devices being more prone to satisficing behaviour are not consistently supported. Most studies point to marginal (or no) differences when comparing survey completions on smartphone and PC, respectively, particularly after controlling for self-selection and nonresponse (e.g., Andreadis, 2015; de Bruijne & Wijnant, 2013; Sommer, Diedenhofen & Musch, 2017; Toepoel & Lugtig, 2014). On response quality, Keusch and Yan (2017) also found that although smartphone participants had more missing data and took longer to complete the survey, they were also less inclined to straightline compared to respondents who completed the survey on a PC. Furthermore, based on a recent experimental study in which the respondents were randomly assigned to complete the survey on either a smartphone, tablet, or laptop, Tourangeau et al. (2018) conclude: 'data from the smartphone group seem, by most standards, to be just as good as those obtained on tablet or laptop computers' (p. 550).

Despite the varying results, it is relatively well-established that an important factor in enabling high response quality in surveys completed on mobile devices is the survey being optimized for mobile devices, as demonstrated by Antoun et al. (2017) in their systematic review of the existing literature. As regards response rates and break-off rates, studies have demonstrated that even when the questionnaire is optimized for the smaller screens of mobile devices, response rates are lower and break-off rates higher for those using a smartphone compared to a PC (Antoun, 2015; Buskirk & Andrus, 2014; Toepoel & Lugtig, 2014). Compared to 'non-optimized' designs, however, an optimized design is found to have positive effects on response quality (Antoun et al., 2017: 560). Interestingly, the design of the Tourangeau et al. (2018) study was not optimized for smartphones (2018: 545).

Previous studies have addressed device effects on response quality by comparing PC and smartphone completions, as these devices differ significantly as regards user context, screen size, and other technical features. While studies tend to argue that tablets resemble PCs more than smartphones (e.g., Couper, Antoun & Mavletova, 2017: 134; Peterson et al., 2017: 204; Wells, Bailey & Link, 2013: 2, 10), this question remains unsettled, which calls for further empirical examination.

Most device effect studies are based on web panels, the LISS Panel (e.g., Antoun, Couper & Conrad, 2017; Lugtig & Toepoel, 2016) the CentERpanel (de Bruijne & Wijnant, 2013), and the GESIS Panel (Struminskaya, Weyandt & Bonsjak, 2015). Although these panels are probability-based and have no indication of unrepresentativeness (cf. Scherpenzeel & Das, 2011; Bosnjak et al., 2018), the risk of panel

effects remains. Web panellists are more experienced than the general population, and, in contrast to common assumptions, tend to be less likely to satisfice (for a literature overview, see Hillygus, Jackson & Young, 2014).

Furthermore, most studies are based on questionnaires containing relatively few items designed for device experiments (e.g., Andreadis, 2015; Keusch & Yan, 2017; Schlosser & Mays, 2017; Toepoel & Lugtig, 2014). Genuine questionnaires for scientific purposes are most likely more challenging to participants, requiring greater effort and commitment (Peterson et al., 2017: 219) and possibly reinforcing device effects. Yet no empirical studies support this proposition. Finally, most of the recent studies of device effects apply an experimental design to eliminate self-selection bias. However, forcing participants to complete a survey using a specific device with which they may not be familiar and which they would not choose if given the choice of device creates an unnatural survey response context, and the results will therefore likely exaggerate the device effects (e.g., discussed by Lugtig & Toepoel, 2016: 80).

Thus, prior studies call for further empirical examination of device effects on survey response quality (Antoun et al., 2018; Antoun, Couper & Mavletova, 2017). The main aim of this study is to determine whether mobile devices cause poorer survey responses than PCs in web surveys. We aim at answering the research question: *Are responses on smartphones and tablets of poorer quality than responses on PCs?*

Examining device effects on cross-sectional samples and having respondents answer authentic questionnaires while they choose the response device themselves distinguishes the present study from most of the recent studies and possibly leads to different results. On one hand, we would expect device effects in survey responses to be reinforced: The study is based on cross-sectional samples representative of the general population, which is more inclined to satisfice than trained panellists, as demonstrated above, and the length of the questionnaires used surpasses most previous studies (respectively, the questionnaires included 119 and 142 items), resulting in an increased respondent burden. On the other hand, enabling the respondents to select their preferred response device provides the optimal conditions for high-quality survey response, and is expected to reduce device effects on response quality.

The following section presents the data and methods applied in the study.

Data and methods

The current study utilized two datasets that originate from the Danish ISSP Programme in 2018 and 2019. Both surveys were collected using a self-administered web questionnaire with respondent self-selection of response device and were optimized for mobile devices. In both surveys, the response device was detected by a question: 'How did you answer this survey?' (PC, smartphone, tablet, and other) and validated by paradata on the screen size of the response device (see Severin, Clement & Shamshiri-Petersen, 2019).

In both cases, the surveys were based on representative population samples drawn from the Danish Civil Registration System (CPR), and the target population was all Danish adults aged 18–79. In 2018, 1,865 out of 5,000 adults completed or partially completed the survey (AAPOR RR2: 37%). In 2019, 1,139 out of a gross sample of 3,004 adults answered the survey completely or partially (AAPOR RR2: 38%). Both samples are largely representative of the target population except for a somewhat larger proportion of males than females in 2018 (population: 50%, sample: 55%, 95% CI for difference: [3.0;7.7]) and a somewhat smaller proportion of participants in the youngest age group (18–25 yrs.) in 2019 (population:

14%, sample: 9%, 95% CI for difference: [3.3;6.6]). Despite these deviations, the results presented are based on unweighted data, as the ambition of the article is to ensure high ecological validity. For more details on sample characteristics, see Appendix 1.

Two samples were included for the purpose of validating the results. If findings differ significantly, device effects presumably result from particular circumstances in the individual surveys, such as the topic of the survey or the mode of contact. Device effects found across the two samples, however, are considered valid. In 2018, the topic of the survey was Danish attitudes towards religion and religious practices, and in 2019 the main topic was attitudes to and experiences with social inequality in Danish society. As regards the means of contact, digital letters of invitation were sent together with a personalized link to the web survey via e-Boks, the online digital mailbox through which Danish authorities have been sending all personal correspondence to residents of the country since 2014. The service is linked to respondents' civil registration number (the so-called CPR system), which enabled us to contact them digitally without an e-mail address. If necessary, follow-up letters of invitation (identical to the digital version) were sent via traditional post in 2019.

Measures and data analysis

The survey methods literature has typically drawn a distinction between representation errors and measurement errors when assessing data quality (cf. Groves et al., 2009: 48). Data quality by measurement revolves around how well data captures the concept or phenomenon of interest. In this case, the engagement of survey participants in the completion is crucial. Satisficing (Krosnick & Alwin, 1987; Krosnick, 1991; Narayan & Krosnick, 1996; Krosnick, 1999) refers to participants shortcutting important steps in the cognitive process of survey responding. Instead of careful considerations, they provide merely satisfactory answers with the least possible psychological cost. Thus, in line with previous studies on device effects (e.g., Andreadis, 2015; de Bruijne & Wijnant, 2013; Keusch & Yan, 2017; Lugtig & Toepoel, 2016; Mavletova, 2013; Struminskaya, Weyandt & Bosnjak, 2015; Tourangeau et al., 2018), this study applies seven indicators, which are both direct and indirect measures (Baumgartner & Steenkamp, 2001) of participant engagement in the survey response process (e.g., Tourangeau et al., 2000). The two direct measures of survey response quality are: (1) participants' own evaluation of their engagement measured in terms of how carefully they considered and answered the questions asked (response categories ranging from 'strongly agree' to 'strongly disagree') and (2) completion time. Spending time completing the survey is mostly considered an indicator of participant engagement, where the higher the completion time, the more committed the survey response behaviour. However, the interpretation of completion time is somewhat ambiguous, as a high completion time may also reflect difficulties in answering the survey. We therefore use the neutral concept 'completion time' (Couper, Antoun & Mavletova, 2017: 140-141; Couper & Peterson, 2017). Either way, varying completion times indicate differences in participant engagement across devices. Five additional, indirect measures are also included; that is, the inclination to select: (3) 'Strongly agree' regardless of question content in Likert-scale variables (often denoted acquiescence) (Baumgartner & Steenkamp, 2001: 145), (4) non-substantive responses ('Can't choose', 'Refuse to answer', and item nonresponse included), (5) the midpoint in Likert scales and similar scales, (6) choosing the first possible response option (primacy effects), and (7) nearly identical response options in batteries (straightlining).

The indirect measures are constructed as simple additive indexes and summarize the number of items in which the respondent answered 'Strongly agree', provided a non-substantive answer, etc. As the number of items varied across surveys (and to enable comparisons), all of the measures were rescaled to a 100-point scale using the following formula:

$$\left(\frac{X - X_{min}}{X_{range}}\right) \cdot n$$

where X = value on original variable, X_{min} = min. on original variable, X_{range} = range on the original variable, and n = upper limit of the new variable (Giannoulis 2020). The closer to 100, the more pronounced the satisficing behaviour. For example, a total 19 items are used to calculate the acquiescence index in 2018, meaning that the highest possible respondent score before rescaling the index is 19, the lowest possible score is 0 (x_{min}), and the range is 19 (x_{range}). The selected upper limit of the rescaled index is 100 (n). All of the items used to construct the measures are single-choice questions with radio buttons (for a more detailed list of items used, see Appendix 2).

To examine the device effects, a series of multivariate logistic regression analyses are conducted. The data analysis is carried out in seven main steps, one for each data quality indicator. In each step, hierarchical logistic regression analyses are conducted, with Odds Ratio and Nagelkerke R^2 applied to measure device effects and model fit. In each sample, two regression models are computed. The first model examines the uncontrolled association between response device and the given response quality indicator; that is, the device effect when respondents choose the device themselves. The second model includes gender, age, and education to examine the controlled association; that is, controlling for self-selection bias. To conduct this analysis, all of the response quality indicators are dichotomized, generally with the 75th percentile as the cut-off: (0) scores below the 75th percentile and (1) scores above the 75th percentile. A score above the 75th percentile indicates that the respondent is more inclined to satisfice than most. Despite satisficing not being dichotomous in its core, we find this approach suitable, as it is the relative difference between devices (not necessarily the *actual level* of satisficing) that is of interest.

Results

First of all, examining the self-selection of response devices across different population subgroups, we find results largely in line with previous research (see Appendix 3). In both samples, we find a larger proportion of males in the PC group, whereas a larger proportion of females answered on smartphones. In addition, the PC and tablet is the preferred response device for the two oldest age groups (56–65, 66–79 yrs. old), whereas a larger proportion of the younger and middle-aged respondents prefer a smartphone (18–25, 26–35, 36–45 yrs. old). The age effects are, however, not linear. Finally, no systematic differences are identified among different educational groups.

Turning to the association between response device and response quality, seven measures were applied in this study altogether. They do not support any assumptions on mobile handheld devices consistently causing poorer responses. When comparing smartphone, tablet, and PC, there is no systematic evidence of device effects on survey response quality.

As to the direct measures of participant engagement in completion, no consistent device effects were found – neither in the uncontrolled nor the controlled logistic regression analyses. Table 1 shows respondents' self-evaluated engagement in the survey response process by device. As demonstrated, no significant effects were identified at the 0.05 alpha level in the final, controlled regression model in both samples (2018 and 2019).

Table 1. Logistic regression with self-evaluated engagement as dependent variable. Odds Ratio, Standard Error (SE), and Nagelkerke R^2

	2018		2019	
	Model I	Model II	Model I	Model II
Response device				
PC (ref)				
Smartphone	1.034 (0.14)	1.319* (0.19)	0.455 (0.48)	0.352 (0.38)
Tablet	1.349 (0.25)	1.261 (0.23)	2.101 (1.27)	2.382 (1.51)
Age	-	1.021*** (0.00)	-	0.977 (0.02)
Gender	-		-	
Female (ref)	-		-	
Male	-	1.212 (0.16)	-	0.685 (0.37)
Years of education		1.010 (0.01)	-	0.995 (0.04)
Nagelkerke R^2	0.003	0.032	0.019	0.036

Note: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.
Based on the question: "Agree or disagree. I spend a lot of time considering my responses in order to answer as precisely as possible". 2018: Strongly disagree (1), Disagree (2), Agree (3), and Strongly Agree (4). 2019: Strongly disagree (1), Disagree (2), Neither Agree Nor Disagree (3), Agree (4), and Strongly Agree (5). Data was dichotomized as follows: (0) Positive self-evaluation (strongly agree or agree), and (1) Negative self-evaluation (strongly disagree or disagree).
n=2018: 1,556, and 2019: 689.

Similarly, the examination of survey completion time does not reveal significant overall device effects. In 2018, the smartphone respondents generally completed the survey faster than the PC group (OR = 1.501, $p < 0.001$). The effect is, however, not significant when controlled for gender, age, and educational status.

Table 2. Logistic regression with survey completion time as dependent variable. Odds Ratio, Standard Error (SE), and Nagelkerke R²

	2018		2019	
	Model I	Model II	Model I	Model II
Response device				
PC (ref)				
Smartphone	1.501*** (0.20)	1.190 (0.17)	0.978 (0.27)	0.718 (0.21)
Tablet	0.713 (0.15)	0.839 (0.19)	0.667 (0.21)	0.967 (0.32)
Age	-	0.970*** (0.00)	-	0.963*** (0.01)
Gender	-		-	
Female (ref)	-		-	
Male	-	0.992 (0.13)	-	1.411* (0.28)
Years of education	-	0.999 (0.01)	-	0.997 (0.02)
Nagelkerke R²	0.018	0.074	0.004	0.10

Note: *p < 0.1, **p < 0.05, ***p < 0.01.
Based on paradata: The survey completion time was extracted from paradata and was calculated as the difference between the start and finish time for each completed survey response. Values below 5 minutes and 60 minutes were recoded to missing as they were considered to reflect responses in several sittings or long breaks rather than the actual interaction time with the survey. Data was dichotomized with the 25th percentile as cutpoint: (0) Completion time above the 25th percentile and (1) Completion time below the 25th percentile.
n=2018: 1,324, and 2019: 570.

As shown in Table 2, the same overall tendency applies to the 2019 sample, as no significant device effects were revealed with regards to response time, neither in the uncontrolled nor the controlled analysis.

Although studies in the field continue to debate how to interpret completion time, the results from the present study indicate that smartphone respondents spend less time on completion than do PC respondents; however, this is due to self-selection effects more than 'pure' device effects.

As to the indirect measures of engagement, the inclination of participants to choose the 'Strongly agree' response option throughout the questionnaires was applied as an indicator of uncommitted survey response behaviour (acquiescence bias, e.g., Baumgartner & Steenkamp, 2001). As displayed in Table 3, however, we do not find systematic evidence for the existence of device effects.

Table 3. Logistic regression analysis with acquiescence as dependent variable. Odds Ratio, Standard Error (SE), and Nagelkerke R²

	2018		2019	
	Model I	Model II	Model I	Model II
Response device				
PC (ref)				
Smartphone	1.165 (0.16)	1.463** (0.22)	0.984 (0.21)	1.031 (0.23)
Tablet	0.875 (0.19)	0.832 (0.18)	0.756 (0.18)	0.707 (0.18)
Age	-	1.019*** (0.00)	-	1.006 (0.01)
Gender	-		-	
Female (ref)	-		-	
Male	-	1.577*** (0.22)	-	0.844 (0.13)
Years of education	-	0.992 (0.01)	-	0.996 (0.01)
Nagelkerke R²	0.002	0.033	0.003	0.007

*Note: *p < 0.1, **p < 0.05, ***p < 0.01.*
Calculation of the indexes: Respectively 19 items (2018) and 11 items (2019) from the topic section of the ISSP survey were used to calculate the index. Inspired by Keusch & Yan's approach (2017), all selected items used the Likert scale, and the respondent's score was determined by the number of items in which he/she answered "Strongly Agree". The scale was transformed to range from 0-100 to enable comparisons across surveys, and data was dichotomized by the 75th percentile: (0) Score below the 75th percentile and (1) Score above the 75th percentile.
n=2018: 1,595, and 2019: 746.

As Table 3 demonstrates, the smartphone group was more likely to answer 'Strongly agree', regardless of question content compared to the PC group in 2018 data (OR = 1.463, p < 0.05 in Model II). This effect was not found in the 2019 data, however, meaning that no consistent device effects were revealed.

A second indirect measure is the inclination of the participant to select non-substantive answers. Although these indeed reflect true opinions in some cases, it is usually a way to skip steps in the cognitive process when answering the survey (Krosnick et al., 2002). As evident in Table 4, no significant device effects were found in 2019. As for the 2018 data, the tendency to provide non-substantive answers was very low (mean: 0.062), and the variation was close to 0 (S.D. = 0.49). Logistic regression analyses were therefore not conducted in 2018. Following the low variation in the data, however, we do not expect device effects to be present.

Table 4. Logistic regression analysis with non-substantive answers as dependent variable. Odds Ratio, Standard Error (SE), and Nagelkerke R²

	2019	
	Model I	Model II
Response device		
PC (ref)		
Smartphone	1.448 (0.34)	1.316 (0.32)
Tablet	0.759 (0.23)	0.726 (0.22)
Age	-	0.995 (0.01)
Gender	-	
Female (ref)	-	
Male	-	0.640** (0.12)
Years of education	-	0.990 (0.01)
Nagelkerke R²	0.008	0.022

Note: *p < 0.1, **p < 0.05, ***p < 0.01.

Calculation of the index: 78 items (2019) from the topic section of the ISSP survey were used to calculate the index, and the respondent's score sums up the number of items in which he/she provided a non-substantive answer. The following responses were treated as non-substantive answers: "Can't choose", "Refused to answer", and item nonresponse. The scale was transformed to range from 0-100 to enable comparison across surveys. Data was dichotomized by the 75th percentile: (0) Score below the 75th percentile and (1) Score above the 75th percentile.

n=2019: 746.

A third indirect measure of response quality is the respondent's inclination to choose midpoint values, such as 'Neither agree nor disagree' in Likert or Likert-like scales. This may reflect their actual opinion but may also be a result of skipping cognitive steps when answering the surveys. As presented in Table 5, no significant device effects were identified in the 2018 or 2019 data.

Table 5. Logistic regression analysis with selection of midpoint values as dependent variable. Odds Ratio, Standard Error (SE), and Nagelkerke R²

	2018		2019	
	Model I	Model II	Model I	Model II
Response device				
PC (ref)				
Smartphone	0.935 (0.13)	0.940 (0.14)	0.915 (0.19)	0.914 (0.20)
Tablet	0.873 (0.17)	0.857 (0.17)	0.945 (0.22)	0.951 (0.22)
Age	-	1.002 (0.00)	-	0.998 (0.00)
Gender	-		-	
Female (ref)	-		-	
Male	-	0.922 (0.12)	-	1.085 (0.17)
Years of education	-	0.990 (0.01)	-	0.972** (0.01)
Nagelkerke R²	0.001	0.002	< 0.001	0.011

Note: *p < 0.1, **p < 0.05, ***p < 0.01.

Calculation of index: Respectively 27 items (2018) and 16 items (2019) from the topic section of the ISSP survey were used to calculate the index. All items used Likert or Likert-like scales, and the respondent's score sums up the number of items in which he/she selected the midpoint. The scale was transformed to range from 0-100 to enable comparisons across surveys, and data was dichotomized by the 75th percentile: (0) Score below the 75th percentile and (1) Score above the 75th percentile.
n=2018: 1,595, and 2019: 746

The inclination to select the first response option is a fourth indirect measure of survey participant engagement when responding to survey questions. Choosing the first category offered on items regardless of the question asked or response options is indicative of satisficing behaviour, and the question pursued in this study is whether some response devices promote such behaviour more than others. We find no consistent device effects after summarizing the number of first response options selected and calculating mean scores.

Table 6. Logistic regression with primacy effects as dependent variable.
Odds Ratio, Standard Error (SE), and Nagelkerke R²

	2018		2019	
	Model I	Model II	Model I	Model II
Response device				
PC (ref)				
Smartphone	1.135 (0.14)	1.137 (0.15)	0.813 (0.20)	0.810 (0.20)
Tablet	0.780 (0.15)	0.814 (0.15)	0.511** (0.15)	0.544** (0.16)
Age	-	0.997 (0.00)	-	0.996 (0.00)
Gender	-		-	
Female (ref)	-		-	
Male	-	1.539*** (0.18)	-	1.343* (0.24)
Years of education	-	0.992 (0.01)	-	0.993 (0.01)
Nagelkerke R²	0.004	0.018	0.012	0.019

*Note: *p < 0.1, **p < 0.05, ***p < 0.01.*
Calculation of indexes: Respectively 70 items (2018) and 68 items (2019) from the topic section of the ISSP survey were used to calculate the index. The respondent's score sums up the number of items in which he/she selected the first response category. The scale was transformed to range from 0-100 to enable comparisons across surveys, and data was dichotomized by the 75th percentile: (0) Score below the 75th percentile and (1) Score above the 75th percentile.
n=2018: 1,595, and 2019: 746.

As demonstrated in Table 6, tablet respondents were less likely to select the first possible response option compared to the PC group in the 2019 data, and the effect was significant in both the uncontrolled and controlled analysis (OR = 0.544, p < 0.05). These findings were not confirmed in the 2018 data, however, where no statistically significant device effects were identified.

A fifth and final indirect measure of response quality is straightlining (cf. Herzog & Bachman, 1981; Krosnick & Alwin, 1988). As we were unable to construct comparable measures, only the 2019 data was included in this final analysis. The measure was based on a 10-item battery about the importance of different aspects for getting ahead in life (e.g., coming from a wealthy family, giving bribes, and a person's race).

Table 7. Logistic regression with straightlining as dependent variable. Odds Ratio, Standard Error (SE), and Nagelkerke R²

	2019	
	Model I	Model II
Response device		
PC (ref)		
Smartphone	1.022 (0.24)	1.191 (0.29)
Tablet	1.5* (0.35)	1.368 (0.33)
Age	-	1.015*** (0.01)
Gender	-	
Female (ref)	-	
Male	-	0.940 (0.16)
Years of education	-	1.003 (0.01)
Nagelkerke R²	0.006	0.019

*Note: *p < 0.1, **p < 0.05, ***p < 0.01.*

Calculation of index: Respondent's standard deviation across a battery of 10 items were calculated and were dichotomized with the 25th percentile as cutpoint: (0) S.D. above the 25th percentile, i.e. low degree of straightlining, and (1) S.D. below the 25th percentile, i.e. high degree of straightlining.
n=2019: 745.

As is evident from the table, tablet respondents were more inclined to straightline than the PC group, but the effect was not significant when controlled for core demographic variables. However, this finding should be interpreted with caution given how it was not possible to replicate findings.

Discussion and conclusion

The main aim of this study was to determine whether using mobile devices such as smartphones or tablets causes poorer survey responses than using a PC. Results from previous studies show no systematic differences. These studies are often based on web panels, however, allowing for possible 'panel effects' to affect the results. In addition, studies often rely on short questionnaires designed specifically for the purpose of testing device effects, and the response device is often randomly assigned to respondents to counter self-selection biases.

This study contributed to this growing body of literature on device effects by testing potential device effects in a setting with high ecological validity and increased reliability. To ensure high ecological validity, device effects were tested under three fundamental conditions: (1) The results are based on a cross-sectional probability sample outside of a panel context; (2) the results stem from a genuine scientific and authentic questionnaire that is significantly longer than previous studies, and (3) participants were given the opportunity to choose the response device themselves. To increase reliability, the device effects were tested in two different surveys.

The present study has adopted seven measures widely used to capture the quality of response behaviour: (1) how participants evaluate their own engagement in answering the survey and (2) response time as direct measures of participant engagement in the survey response process. The following indirect measures of satisficing were also used: (3) tendency to agree regardless of the question content (acquiescence), (4) inclination to provide non-substantive answers, (5) tendency to choose midpoint values in Likert or Likert-like scales, (6) inclination to choose the first response option presented (primacy effects), and (7) straightlining.

Overall, the study's findings are largely in line with previous research, as we find no evidence of systematic device effects on survey response quality. The results demonstrate that a respondent's self-evaluated engagement in survey completion does not differ across devices, and only small, non-systematic differences between devices on the remaining satisficing indicators are identified across the two samples. As to both the non-substantive answers and selection of mid-point response options, no device effects were found at all. As to response time and the tendency to agree regardless of question content, the results indicate that smartphone respondents took less time to complete the survey and were more inclined to agree to survey questions. This was only the case in the 2018 sample, however, and as for completion time, the effect was no longer significant when controlled for self-selection bias. Results were not replicated in the 2019 sample. As to primacy effects and straightlining, results indicate that tablet respondents were less likely to choose the first response option on a range of items but more likely to choose the similar response category on a battery of items with the same response scale. This is only found in the 2019 sample, however, and the effect was no longer significant when controlled for self-selection bias.

Thus, the present study finds no support for mobile devices causing poorer survey response quality. The fact that the results are largely consistent across two different questionnaires with very different themes supports this conclusion. And as demonstrated, controlling the associations between response device and the response quality indicators for self-selection biases does not change the result.

The study has some notable limitations that should be taken into consideration when interpreting the results. First, device effects are tested on a broad range of measures based on single-choice questions with radio buttons, whereas open-ended questions were not included in the analysis. It is well-documented in the existing literature that mobile respondents typically provide shorter answers to open-ended questions compared to PC respondents (e.g., Mavletova, 2013: 738; Struminskaya, Weyandt & Bosnjak, 2015: 281). The inclusion of a measure capturing the length of answers might therefore have provided more nuance regarding the device effects on data quality. On the other hand, we considered the quality of answers to be a more complex matter than merely a matter of length, for which reason we did not include the measure in the article. Second, despite controlling for potential self-selection biases, we are not able to completely eliminate the presence of self-selection effects.

The findings are slightly contrary to initial expectations. Based on previous research on the quality of survey responses, we expected the two first conditions (cross-sectional samples in a non-panel setting and authentic questionnaire) to contribute to an increase in satisficing, whereas the last condition (choice of device for completion) was expected to decrease satisficing. The results showed that even under the first two conditions, no significant device effects were identified, whereas completion device being chosen by the respondent might have contributed to less satisficing. Self-selection effects will be considered a study limitation in almost every other case, as they weaken the validity of the study results. In this case, however, it is quite the opposite: Participants are free to select the device with which they are most comfortable, which gives them optimal conditions for providing high-quality answers. Against this

backdrop, an experimental setting poses an unnatural situation in which respondents may be forced to use a device they do not prefer or with which they are unfamiliar. This arguably produces poorer survey responses and may result in an overestimation of the size of device effects. It is, however, outside the scope of this study to draw any definitive conclusions on this matter, but future studies are encouraged to further explore the potential benefits of device self-selection on response quality.

Appendix 1

Table 8. Gender and age distribution in sample and population (%)

Appendix 2

Table 9. Overview of items used to construct indirect measures of satisficing

Appendix 3

Table 10. Response device by gender, age, and education (%)

References

1. Andreadis, I. (2015). Web Surveys Optimized for Smartphones: Are There Differences between Computer and Smartphone Users? *Methods, Data, Analyses*, *9*(2), 213-228.
2. Antoun, C. (2015). Who Are the Internet Users, Mobile Internet Users, and Mobile-Mostly Internet Users? Demographic Differences across Internet-Use Subgroups in the U.S. In: Toninelli, D., Pinter, R., & de Pedraza, P. (Ed.), *Mobile Research Methods: Opportunities and Challenges of Mobile Research Methodologies* (99-118). London: Ubiquity Press.
3. 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*, *81*(S1), 280-306.
4. Antoun, C., Katz, J., Argueta, J., & Wang, L. (2018). Design Heuristics for Effective Smartphone Questionnaires. *Social Science Computer Review*, *36*(5), 557-574.
5. Baumgartner, H., & Steenkamp, J-B. E. M. (2001). Response Styles in Marketing Research: A Cross-National Investigation. *Journal of Marketing Research*, *XXXVIII*(May 2001), 143-156.
6. Bosnjak, M. Dannwolf, T., Enderle, T., Schaurer, I., Struminskaya, B., Tanner, Angela, & Weyand, K. W. (2018). Establishing an Open Probability-Based Mixed-Mode Panel of the General Population in Germany: The GESIS Panel. *Social Science Computer Review*, *36*(1), 103-115.
7. Buskirk, T. D., & Andrus, C. (2014). Making Mobile Browser Surveys Smarter: Results from a Randomized Experiment Comparing Online Surveys Completed via Computer or Smartphone. *Field Methods*, *26*, 322–342
8. Couper, M. P., Antoun, C., & Mavletova, A. (2017). Mobile Web Surveys: A Total Survey Error Perspective. In: Biemer, P. P. et al. (Ed.), *Total Survey Error in Practice* (chapter 7, 133-154). New York: Wiley.
9. Couper, M. P., & Peterson, G. J. (2017). Why Do Web Surveys Take Longer on Smartphones? *Social Science Computer Review*, *35*(3), 357-377.
10. 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), 482-502.

11. Deng, T., Kanthawala, S., Meng, J., Peng, W., Kononova, A., Hao, Q., Zhang, Q., & David, P. (2019). Measuring Smartphone Usage and Task Switching with Log Tracking and Self-Reports. *Mobile Media & Communication*, *7*(1), 3-23.

12. Giannoulis, C. (2020). Rescaling Sets of Variables to Be on the Same Scale. *The Analysis Factor*. https://www.theanalysisfactor.com/rescaling-variables-to-be-same/

13. Groves, R. M., Fowler Jr., F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology*. Hoboken: Wiley (2nd ed.).

14. Helles, R. (2016). Mobile medier. In: Jensen, K. B. (ed.). *Dansk mediehistorie 4: 1995-2015* (pp. 57-74). Frederiksberg: Samfundslitteratur (2nd ed.).

15. Herzog, A. R., & Bachman, J. G. (1981). Effects of Questionnaire Length on Response Quality. *Public Opinion Quarterly*, *45*, 549–559.

16. Hillygus, S. H., Jackson, N., & Young, M. (2014). Professional Respondents in Nonprobability online Panels. In: Callegaro, M. (ed.). *Online Panel Research: A Data Quality Perspective*. West Sussex: John Wiley & Sons.

17. Keusch, F., & Yan, T. (2017). Web Versus Mobile Web: An Experimental Study of Device Effects and Self-Selection Effects. *Social Science Computer Review*, *35*(6), 751-769.

18. Kim, Y., Dykema, J., Stevenson, J., Black, P., & Moberg, D. P. (2019). Straightlining: Overview of Measurement, Comparison of Indicators, and Effects in Mail-Web Mixed-Mode Surveys. *Social Science Computer Review*, *37*(2), 214-233.

19. Krosnick, J. (1991). Response Strategies for Coping with the Cognitive Demands of Attitude Measurement in Surveys. *Applied Cognitive Psychology*, *5*, 213-236.

20. Krosnick, J. (1999). Survey Research. *Annual Review of Psychology, 50*, 537-567.

21. Krosnick, J. A., Allyson L., Holbrook, M. K., Berent, R. T., Carson, W., Hanemann, M., Reymond J. K., Mitchell, R. C., Presser, S., Ruud, P. A., Smith, V. K., Moody, W. R., Green, M. C., & Conaway, M. (2002). The Impact of 'No Opinion' Response Options on Data Quality: Non-Attitude Reduction or an Invitation to Satisfice? *Public Opinion Quarterly, 66*, 371-403.

22. Krosnick, J., & Alwin, D. F. (1987) An Evaluation of a Cognitive Theory of Response-Order Effects in Survey Measurement. *Public Opinion Quarterly*, *51*(2), 201-219.

23. Krosnick, J. A., & Alwin, D. F. (1988). A Test of the Form Resistant Correlation Hypothesis: Ratings, Rankings, and the Measurement of Values. *Public Opinion Quarterly*, *52*, 526–538.

24. Lugtig, P., & Toepoel, V. (2016). The Use of PCs, Smartphones, and Tablets in a Probability-Based Panel Survey: Effects on Survey Measurement Error. *Social Science Computer Review*, *34*(1), 78-94.

25. Mavletova, A. (2013). Data Quality in PC and Mobile Web Surveys. *Social Science Computer Review*, *31*(6), 725-743.

26. Mavletova, A., & Couper, M. P. (2013). Sensitive Topics in PC Web and Mobile Web Surveys: Is There a Difference? *Survey Research Methods*, *7*, 191-205.

27. Narayan, S., & Krosnick, J. (1996). Education Moderates Some Response Effects in Attitude Measurement. *Public Opinion Quarterly*, *60*(1), 58-88.

28. Peterson, G., Griffin, J., LaFrance, J., & Li, J. (2017). Smartphone Participation in Web Surveys. Choosing between the Potential for Coverage, Nonresponse, and Measurement Error. In: Biemer, P. P. et al. (Ed.), *Total Survey Error in Practice* (chapter 10, 203-234). New York: Wiley.

29. Revilla, M. (2017). Are There Differences Depending on the Device Used to Complete a Web Survey (PC or Smartphone) for Order-by-click Questions? *Field Methods*, 29(3), 266-280.

30. Scherpenzeel, A. C., & Das, M. (2011) True Longitudinal and Probability-Based Internet Panels: Evidence from the Netherlands. In: Das M. et al. (ed.) *Social and Behavioral Research and the Internet: Advances in Applied Methods and Research Strategies.* New York: Routledge.

31. Schlosser, S., & Mays, A. (2017). Mobile and Dirty: Does Using Mobile Devices Affect the Data Quality and the Response Process of Online Surveys? *Social Science Computer Review*, *36*(2), 212-230.

32. Severin, M. C., Clement S. L., & Shamshiri-Petersen (2019). Device-effekter i web surveys. Har variationer i brugen af PC, smartphone og tablet betydning for surveykvaliteten? [Device Effects in Web Surveys: Do Variations in the Use of PC, Smartphone and Tablet Affect Survey Quality?]. *Metode & Forskningsdesign* [Methods & Research Design], *4*, 1-27.

33. Sommer, J., Diedenhofen, B., & Musch, J. (2017). Not to Be Considered Harmful: Mobile-Device Users Do Not Spoil Data Quality in Web Surveys. *Social Science Computer Review*, *35*(3), 378-387.

34. 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.

35. 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.

36. Tourangeau, R., Sun, H., Yan, T., Maitland, A., Rivero, G., & Williams, D. (2018). Web Surveys by Smartphone and Tablets: Effects on Data Quality. *Social Science Computer Review*, 36(5), 542-556.

37. Wells, T., Bailey, J., & Link, M. (2013). Filling the Void: Gaining a Better Understanding of Tablet-Based Surveys. *Survey Practice*, *6*(1), 1-13.