Article

Social Science Computer Review
2016, Vol. 34(1) 78-94

DOI: 10.1177/0894439315574248
ssc.sagepub.com
SAGE

The Use of PCs, Smartphones, and Tablets in a Probability-Based Panel Survey: Effects on Survey Measurement Error

Peter Lugtig[1,2] and Vera Toepoel[2]

Abstract

Respondents in an Internet panel survey can often choose which device they use to complete questionnaires: a traditional PC, laptop, tablet computer, or a smartphone. Because all these devices have different screen sizes and modes of data entry, measurement errors may differ between devices. Using data from the Dutch Longitudinal Internet Study for the Social sciences panel, we evaluate which devices respondents use over time. We study the measurement error associated with each device and show that measurement errors are larger on tablets and smartphone than on PCs. To gain insight into the causes of these differences, we study changes in measurement error over time, associated with a switch of devices over two consecutive waves of the panel. We show that within individuals, measurement errors do not change with a switch in device. Therefore, we conclude that the higher measurement error in tablets and smartphones is associated with self-selection of the sample into using a particular device.

Keywords

mobile phones, tablets, measurement error, panel survey, mixed device

Introduction

How surveys are displayed and completed in online panel research has changed in recent years. Nowadays, people do not only complete surveys on desktop PCs or laptops, but also on subnotebooks, tablets, or smartphones (Bosnjak et al., 2013; Callegaro, 2013a). The variance in screen size has increased rapidly, now ranging from 4 inches for a small smartphone to 27 inches for a large desktop PC screen. In addition, touchscreens are now used instead of keyboards for most smartphones and tablets. These new technologies bring new challenges to respondents and survey designers. Internet surveys are now completed using a mix of devices and this can introduce selection and measurement effects.

[1] University of Essex, Colchester, United Kingdom
[2] Utrecht University, Utrecht, the Netherlands

Corresponding Author:
Peter Lugtig, Utrecht University, Padualaan 14, 3508 TC, Utrecht, the Netherlands
Email: p.lugtig@uu.nl

There have been only a few studies studying the effects on survey measurement error of different Internet-enabled devices. Most of these studies have concentrated on comparing mobile phones to desktop PCs and have found no clear differences between them with regard to survey measurement error (Couper, 2013; de Bruijne & Wijnant, 2013; Lynn & Kaminska, 2013; Mavletova, 2013; Mavletova & Couper, 2013; Toepoel & Lugtig, 2014; Wells, Bailey, & Link, 2014).

Almost all of these studies have relied on experiments of prescreened mobile phone users within an online panel. Such a situation does, however, not correspond to the daily practice of Internet panels. In reality, respondents may at every wave consider what particular device they will use to complete a questionnaire. Some will use the same device consistently, while others may use different devices over the course of the panel study.

Even if there is no overall effect of device on measurement error, this may not be true for all individuals, with some respondents perhaps being better respondents at PCs and others at mobile phones. This is important as some respondents may switch devices over time within a panel. This can lead to changes in measurement error within the individual due to a switch of device. The fact that not everyone owns a smartphone or a tablet imposes selection effects. Although biases seem to decline over time (Fuchs & Busse, 2009), there are still important differences between owners and non-owners of various devices.

It is difficult to separate measurement effects from selection effects, as they interact when respondents self-select into using a particular device. This article aims to overcome the selection problem by analyzing over time (1) whether respondents in a probability-based Internet panel use different devices to complete surveys and (2) assess whether a device switch over time affects survey measurement error. Using six waves of data from the Longitudinal Internet Study for the Social sciences (LISS) in the Netherlands, we investigate the effects on different aspects of measurement error and find that measurement errors are lowest in PCs, followed by tablets and smartphones. Within individuals, we find that a switch of devices over time is not associated with a change in measurement error. This implies that the differences in measurement error between devices can be explained by self-selection of respondents into particular devices.

Background

Bosnjak et al. (2013) compared seven independent samples from market research access panels and one probability-based panel (GESIS Online Panel Pilot) in Germany and found that on average 6% of respondents accessed the survey via their mobile phone. Couper (2013) discusses similar rates for different populations such as students and consumers in the United States. We have reasons to believe that the proportion of people in the Netherlands is higher than this given the fact that smartphone and tablet penetration is higher in the Netherlands than other countries (Eurostat, 2012). In previous research (Toepoel and Lugtig, 2014), we found that 57% of panel members with a smartphone used it when prompted to complete the online survey on a smartphone.

The Nature of Differences Between Internet-Enabled Devices

There are a great number of different devices that can be used to complete Internet surveys. The distinction between PCs, tablets, and smartphones is useful, but the boundaries between these devices have blurred in recent years. In our view, all devices can be classified along two dimensions: (1) screen size and (2) method of data entry.

In as recent as 2007, Internet surveys were mostly displayed on a screen of relatively large size, and respondents entered answers to those surveys using a combination of keyboard and mouse. With the arrival of small laptops, the screen size became an important consideration for designing web surveys. Surveys had to be designed in such a way that questions and response scales would be clearly visible on screens ranging from about 10 to 21 inches. The size of the screen on which a survey is completed may affect measurement errors mainly if the survey is not programmed to dynamically change the size of questions. Desktop PCs normally have a screen size of 15 inches or larger, which implies that respondents can see item batteries or grouped questions in one go. If a survey is not dynamically programmed, individual survey questions will appear very small, and respondents are required to manually zoom in. Dynamically programmed questionnaires will zoom in automatically. This implies, however, that a respondent will only see one question at a time and manually has to scroll from question to question taking extra time and effort. Peytchev and Hill (2010) found that measurement error in a smartphone survey using a very small screen (2.2 inch) was not affected by the number of questions on a page, nor by the need to scroll (horizontally and vertically).

Probably the most important difference between devices lies in the method of data entry. PCs rely on a combination of mouse movements and character entry through a keyboard. Smartphones and tablets use touchscreens, where answers are "indicated" by finger-touches on the screen, and an on-screen keyboard is mostly used to type in answers. De Bruijne and Wijnant (2013) found among a sample of experienced tablet users randomly assigned to tablets, PCs or smartphones, that smartphone respondents were slower than PC and tablet respondents. There was no difference between tablet and PC respondents, implying that it is screen size, or differences in Internet connection speeds, rather than the touchscreen that affects interview length. We should note that finger navigation is less precise than mouse navigation. This could result in frustration on the respondent's part. The larger necessity to scroll can bias ratings (see Couper, Tourangeau, & Conrad, 2004) and can lower respondents' evaluation of the questionnaire (Toepoel, Das, & van Soest, 2009).

Designs to Investigate Measurement Errors in Different Devices

Estimating the exact differences between devices with regard to survey measurement is difficult because selection and measurement effects interact. This is shown in the current literature, since most of the experimental studies that have been conducted all had particular sample selection problems.

Experimental assignment to a particular device may lead to two problems. First, respondents may not be familiar with a device that is given to them. Peytchev and Hill's (2010) early study is an example of this. Survey researchers have tried to deal with this problem by restricting their sample to people who they know to be smartphone users, but that leads to a second problem. Wells, Bailey, and Link (2014) report that about 23% of respondents assigned to a PC control group did not adhere to the experimental condition and still complete the survey on a smartphone. Mavletova (2013) found the opposite. About 13% of respondents assigned to a mobile phone condition, completed the questionnaire on a PC, while de Bruijne and Wijnant (2013) found both types of experimental contamination. The only way to really conduct experimental studies is by strictly controlling the internal validity of the experiment and that may not be feasible with devices in Internet surveys. Mavletova and Couper (2013) used a cross-over design, where all respondents replied to two surveys, using mobile phones and PC. Even in this design, some respondents did not adhere to the device assigned to them at every measurement occasion.

Nonexperimental studies have the disadvantage that respondents may self-select into using a particular device. In a previous study (Toepoel and Lugtig, 2014), respondents could choose whether they wanted to complete the survey on their smartphone or on a regular desktop. There were no differences between devices on measurement error, but letting respondents choose their device themselves makes it impossible to separate measurement effects from selection effects. The effects on measurement errors may become biased if the correlates for self-selection are also related to measurement error. For example, if younger people are more likely to use tablets and smartphones, and if younger people also report with more measurement error, nonexperimental studies could falsely

Table 1. Indirect Indicators of Satisficing and Measurement Error.

Don't knows	The more "don't know" answers, the more respondents use cognitive shortcuts, and the larger measurement errors are
Length of open answers	The shorter (and less substantive) the answer, the more respondents satisfice, and the larger measurements errors are
Consistent answer patterns (straightlining)	The more consistent respondents respond to questions in an item battery, the more they satisfice. This may take the form of respondents agreeing to every Likert-scale answer option, or respondents choosing a consistent "extreme" or "middle" response category.
Primacy effect	First answer options chosen indicate satisficing and increased measurement error
Rounding	The more respondents round continuous answers to a round number, the more respondents satisfice
Answers to sensitive items	Lower reports of sensitive behaviors and attitudes indicate satisficing in a particular device.

conclude that surveys on tablets and smartphones are completed with more measurement error. It is therefore important to model measurement error without the confounding selection effect.

Measurement Error

In absence of validation data, most studies use indirect methods to assess measurement error in survey research. Indirect methods link measurement error in surveys to the process of answering a survey question. Measurement errors are caused by not conscientiously understanding the question, retrieving and judging information from memory, or giving an answer (Tourangeau, Rips, & Rasinski, 2000). For example, the more often a respondent uses the "Don't know" answer category, the more likely it is that this respondent does not put a lot of cognitive effort in one of the four stages of the question–answer process as described by Tourangeau, Rips, and Rasinski (2000). Krosnick (1991) has labeled such behavior as satisficing and has associated it with the occurrence of more measurement error. Measures that were used in the literature to detect measurement errors in mobile surveys are listed in Table 1 (see Lynn & Kaminska, 2013; Mavletova and Couper, 2013; Wells et al., 2014).

Bosnjak et al. (2013) found no differences when comparing the number of entries, the number of open-ended questions answered, and the number of characters entered on mobile phones in comparison to desktop computers. This is confirmed by Mavletova (2013). Bosnjak et al. (2013) found that dropout rate was higher for mobile (12% mobile vs. 6% desktop). Lynn and Kaminska (2013) used seven indicators of satisficing (a form of measurement error) and only found mean interview length to be longer for mobile interviews. Mavletova (2013) found no effect of questionnaire length on completion and break off rates. Guidry (2012) found fewer item missing on mobile phones but more straightlining. McClain, Crawford, and Dugan (2012) also found more straightlining in grids, but no evidence of less item nonresponse for smartphones. De Bruijne and Wijnant (2013) added evaluation questions to the questionnaire such as difficulty, clarity, and enjoyment, but they found no significant differences between smartphone and PC users.

On the whole, cognitive processing between PC and mobile surveys appears to be similar. However, to the best of our knowledge no existing study has investigated device switches over different waves in a panel and its effects on survey measurement error.

Methods

Our main aim is to compare survey measurement error over time across three major groups of devices: desktop PCs, tablets, and smartphones. We define a PC desktop as a computer with a large

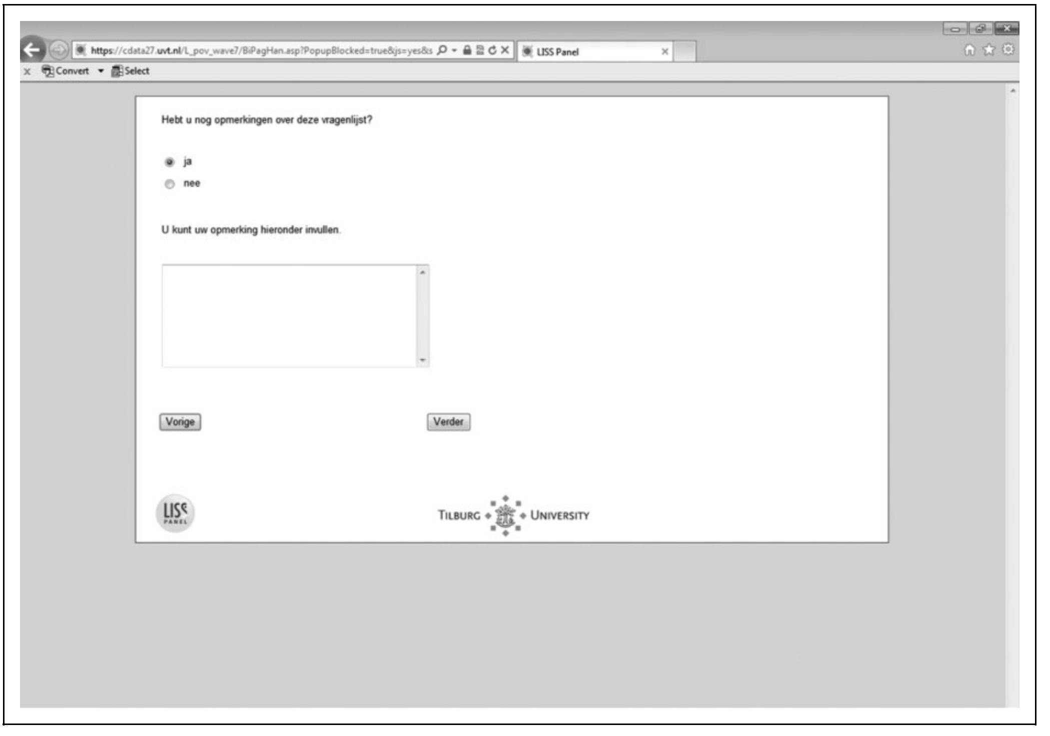

Figure 1. Screenshot for Longitudinal Internet Study for the Social sciences (LISS) questionnaire on desktop computer (HP Windows PC with Internet Explorer).

screen (> 6.0 inches) without a touchscreen. Tablets have a similarly large screen, but use a touchscreen, whereas smartphones are devices with a small screen (≤ 6.0 inches) with or without a touchscreen and a high-resolution screen. This definition implies that we define a laptop as a desktop PC. Also, we do not distinguish between feature phones or smartphones and label any mobile phone with Internet access as a smartphone.

This article studies measurement errors using longitudinal data from 6.226 respondents in a probability-based Internet panel in the Netherlands. Using six waves of data, we:

1. study whether respondents use different devices to complete Internet surveys,
2. analyze measurement errors between different devices, and
3. analyze measurement error within individual respondents over time, to answer the question whether a device switch affects measurement errors.

Respondents complete six questionnaires over a 6-month period (April–September 2013) using a device of their choosing. In our data, one question was normally displayed per screen (this can be a matrix question), and images were not used. In other words, the questionnaires were designed so that they look the same on different devices. Figures 1 to 3 show screenshots of the final question in each survey that asks the respondent "whether you have any remarks about the questionnaire," along with a textbox where the respondents can list these remarks. These screenshots illustrate that in the LISS panel, the questions are displayed in a similar way across devices and that the main difference between devices is the screen size. Because the topic of the questionnaires in LISS changes every month, the indicators for measurement error should be largely independent of the topic of the survey,

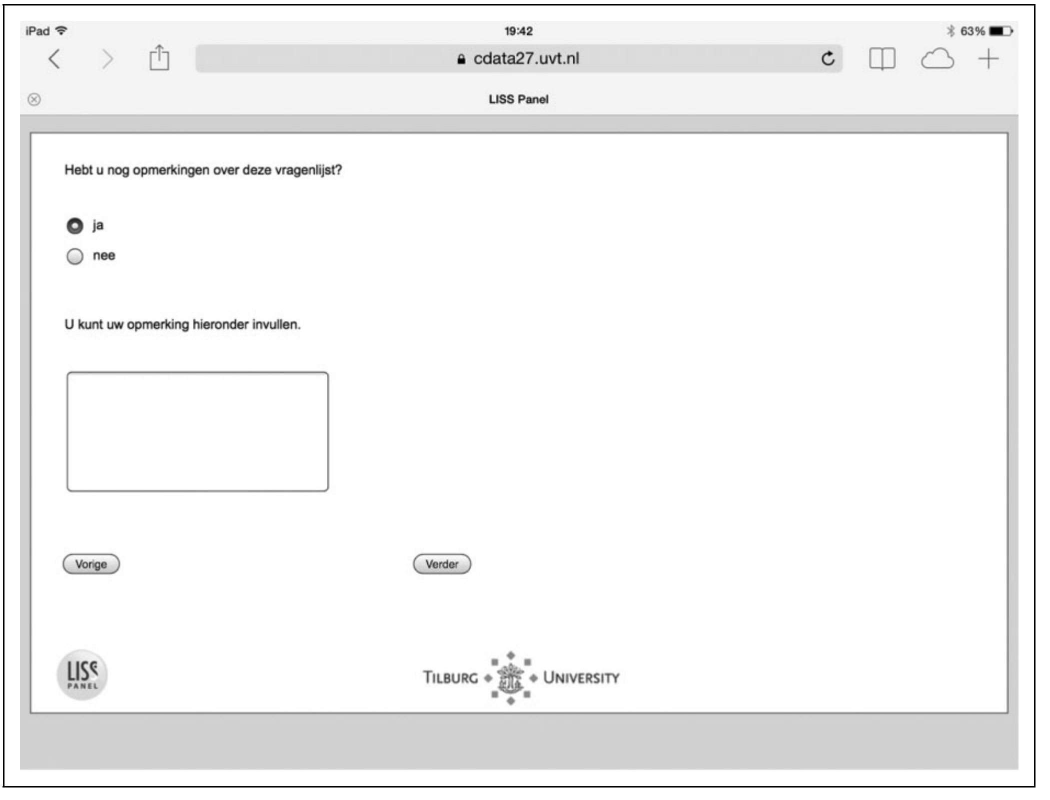

Figure 2. Screenshot for Longitudinal Internet Study for the Social sciences (LISS) questionnaire on tablet (Ipad 3 with safari browser).

implying that methods to model measurement errors using reinterviewing methods using the same variable (Alwin, 2007) are out of scope. We also lack validation data, meaning that only indirect indicators of measurement error can be used.

Indicators for Measurement Error

Item missing. In every wave, we use five Likert-scale questions to count the number of item missings. The Likert scale questions ask respondents, always at the end of the questionnaire, whether the respondent found the survey (1) difficult, (2) clear, (3) whether it got the respondent thinking, (4) interesting, and (5) enjoyable. These questions are asked every wave of the LISS panel and are non-obligatory, so they may be susceptible to satisficing or item missing. The more item missings we find, the higher we believe measurement error to be. Based on earlier studies, we expect no differences between devices on the proportion of item missings.

Open questions. We code two aspects of open questions. First, whether respondents answer a nonobligatory open question at the end of each questionnaire asking for "other remarks" about the questionnaire. Along with the fact whether any answer was given, we also code the length of the answer in characters. Short or no answers are a proxy for more measurement error. We expect no differences in answer length between the different devices. We expect respondents who select the smartphone or tablet to be skilled at typing a relatively short text on these devices. Reversely, those respondents

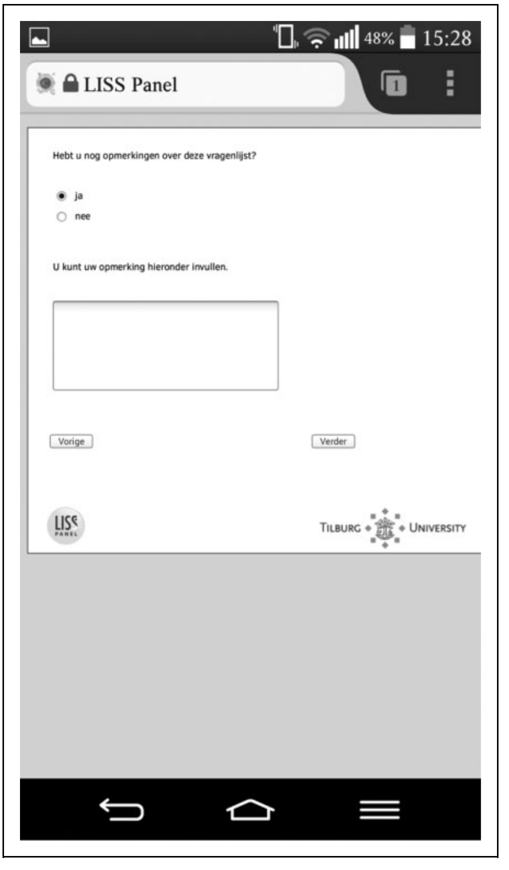

Figure 3. Screenshot for Longitudinal Internet Study for the Social sciences (LISS) questionnaire on smartphone (LG G2 with Firefox browser).

who feel unskilled to type or browse on a mobile device will probably not use such devices to respond to survey requests at all. Bosnjak et al. (2013) found no differences on this variable between smartphone and desktop survey respondents.

Straightlining. We define straightlining as five consistent answers for the evaluation questions discussed earlier (under the heading "item missing"). That is, a respondent is flagged as satisficing when he or she consistently gives the same answer to the five evaluation questions. This operationalization excludes forms of satisficing where the respondent exhibits different response behavior, for example, random answers. We see the occurrence of straightlining as an indicator for more measurement error, but because there were no visual differences for these questions apart from screen size, we expect no differences between the devices.

Primacy effect. In three of the questionnaires, check-all-that apply questions ask respondents to indicate the number of response options that apply to them. In April, a question asked respondents about the activities they were doing just before going to bed with six response options: (1) watching TV, (2) being on a computer, (3) household/administrative tasks, (4) activity outside home, (5) social activity in home, and (6) other. In June, respondents were asked about any income they receive from work outside of being an employee. The response options were (1) freelancer, (2) freelancer next to normal job, (3) self-employed, (4) company owner, (5) participating in partnership, (6) partner at company, (7) own a private limited company, (8) another type of business, and (9) none of these. Finally, in August, respondents were asked which of their direct family members collects things as a hobby, with nine possible response options: (1) father, (2) mother, (3) brother, (4) sister, (5) son, (6) daughter, (7) partner (male), (8) partner (female), and (9) someone else.

A primacy effect occurs when respondents select the first answer option listed. Although it is entirely possible that this choice is a viable choice for some respondents, the difference in the proportion of first answers chosen across devices indicates the difference in primacy effect across the devices we study.[1] We expect the primacy effect to be stronger for smartphones than tablets or desktops. On some smartphones, some check-all-that apply questions in the questionnaire may not have fitted onto the screens, causing the respondent to be more likely to pick the first answer category.

Number of answers checked in a check-all-that-apply question. Using the check-all-that-apply questions listed earlier (under "primacy effect"), we count the total number of answer boxes checked. More answers chosen indicate increased cognitive effort and signal lower measurement error. We again expect a small difference between smart phones and tablet and desktop PCs. On some smartphones, some check-all-that apply questions may not have fitted onto the screen. In this case, we expect the respondent to check fewer answers on a smartphone.

We use two further forms of paradata: interview duration and respondents' evaluation of the questionnaire. Although these do not serve as direct proxies for measurement error, they are important for evaluating the effects of mixed-device use in a panel survey.

Duration (interview length). For every wave, the overall time it took respondents to complete the questionnaire was recorded in seconds. Durations were skewed in every wave. This was caused by respondents leaving the questionnaire open for a long time, without answering any questions or answering the questionnaire in several phases. Because of this, we have decided to trim all durations higher than 3600 s to 1 hr. This occurred for 1.1% of all cases in July and September to 11.1% in June. We found such "break-offs" to be generally somewhat higher in smartphones than in tablets and desktop PCs. The differences are never significant, however, because of the very small sample sizes.

The reason for the large number of outliers in duration in June is probably related to the questionnaire. In that month, respondents were asked about household income and were urged to use their own administration to answer the survey questions. This then probably encouraged respondents to leave the computer, resulting in long interview durations. We expect, because of differences in Internet-speed and differences in navigation, the response duration to be shorter on desktop PCs as compared to tablets and smartphones.

Evaluation of questionnaire. The substantive evaluation of the questionnaire by respondents. See above under the heading Item missing. Factor analysis showed that the five evaluation questions can be summarized into either a one-factor or two-factor solution. For ease of interpretation, we computed one factor score in each month to evaluate whether respondents appreciate the questionnaire more depending on the device they used to complete the questionnaire.

Sample

The data for our study were collected in the LISS panel that started in 2007. This panel is the principal component of the MESS project, operated by CentERdata (a research institute located at the campus of Tilburg University, the Netherlands). The LISS panel consists of almost 8, 000 individuals who complete online questionnaires every month. Panel members were recruited based on a simple random sample of addresses from community registers, in cooperation with Statistics Netherlands. Potential respondents were contacted by letter, telephone or visit, and after an initial interview ("recruitment stage") were asked to become a member of the online panel (which they start with a "profile interview"). Although the LISS panel is Internet based, it was not necessary to own a personal computer with an Internet connection to participate in the panel, as CentERdata provided the equipment if required. Using the response metrics of Callegaro and Disogra (2008), the recruitment rate (or RECR, similar to AAPOR RR3, defined as the number of people that agree to join the panel, relative to all people invited) for the LISS panel is 63%. The profile rate (or PROR; defined as the number of people who complete the profile interview, relative to all people invited) is 48%. Retention is about 90% a year (Binswanger, Schunk, & Toepoel, 2013). Questionnaires are programmed in Blaise and are programmed dynamically, implying that the visual layout of the questionnaire will adapt itself to the device being used.

It is noteworthy that respondents in the LISS panel are paid 15 euros per hr for completing questionnaires (payments are based on an estimate of interview time, needed to fill in the questionnaires). For a more detailed description of the panel, the sample, recruitment and response, see the website www.lissdata.nl. In this article, we use an anonymized version of the data set, meaning that response data could not be linked to sociodemographic variables.

Devices

To code the device that respondents use in responding to the six questionnaires in the LISS panel, we use User Agent Strings (UAS). UAS contain information on the device, operating system, and browser being used (Callegaro, 2013b). During the coding, we experienced one case which we were unable to code the device for. Similarly, we had some problems assigning hybrid laptops that combine keyboards with a touchscreen. We chose to delete these cases ($n = 4$) from our analyses.

Analysis

Our first, descriptive objective is to see how many respondents use a particular device to complete the surveys in the LISS panel. Second, we investigate measurement error per device. Third, we look at patterns of longitudinal device use and measurement error. We compare respondents who use the same device at two consecutive waves (PC, tablets, or smartphone) to groups who switch devices over time. For each of the nine possible switch patterns, we compute the associated change in measurement error between two waves and pool the results across the 5 transitions respondents can make.

Results

In every of the six waves, between 88 and 95% of all LISS respondents complete the surveys using a PC, including laptops. Only a small minority of respondents use a tablet (between 4 and 9%) or a smartphone (between 1 and 3%).

Device Use Over Different Waves

Over time, there is no obvious sample-level change in which devices are being used to complete the surveys. Table 2 shows the transitions respondents make over the course of the waves we analyze. For example, we see that 77.4% of PC respondents in April again use a PC in May. Only 1.5% of April's PC respondents switch to either a tablet or smartphone to complete a questionnaire in May.

The proportion of respondents switching a PC for either a tablet or smartphone is similarly low in the other months and is never more than 5%. This stability in device use for PCs is, however, not

Table 2. Devices Used Between April and September 2013 by LISS Panel Respondents.

% Device used in Month (t)		% Device Used in Subsequent Month (t+1) PC	Tablet	Phone	No Participation	*N*	% of Wave Respondents
April	PC	77.4	1.1	0.4	21.1	4,966	90.2
	Tablet	19.3	24.4	0.2	56.1	435	7.8
	Phone	33.9	4.6	30.3	31.2	109	2.0
	No participation	37.6	2.1	0.8	59.4	715	—
May	PC	71.6	2.1	0.6	25.8	4,234	94.5
	Tablet	24.7	47.8	1.6	25.8	182	4.1
	Phone	16.4	4.9	32.8	45.9	61	1.4
	No participation	37.9	6.5	1.3	54.3	1,749	—
June	PC	77.4	3.1	1.6	17.9	3,749	91.2
	Tablet	23.3	52.7	1.0	22.9	292	7.1
	Phone	29.0	10.1	44.9	15.9	69	1.7
	No participation	53.6	5.8	3.0	37.7	2,116	—
July	PC	84.3	1.9	1.1	12.7	4,122	88.1
	Tablet	21.8	63.9	1.8	12.5	399	8.5
	Phone	41.1	6.3	29.1	23.4	158	3.4
	No participation	55.4	7.9	1.9	34.8	1,547	—
August	PC	90.3	1.8	0.8	7.1	4,482	88.3
	Tablet	27.6	64.0	0.4	7.9	467	9.2
	Phone	46.0	4.0	35.7	14.3	126	2.5
	No participation	49.8	5.2	1.5	43.5	1,151	—
September	PC	—	—	—	—	4,806	89.9
	Tablet	—	—	—	—	443	8.3
	Phone	—	—	—	—	101	1.9
	No participation		—	—	—	876	—

Note. LISS = Longitudinal Internet Study for the Social sciences. $n = 6{,}226$.

found for tablets and smartphones. Once people are using a smartphone in particular, they are not very likely to use a smartphone in the next waves of LISS. Only 29% of smartphone users in July 2013 again uses a smartphone in August, for example. The consistency of tablet usage increases over the course of the panel; 24% of respondent is a consistent tablet user in April–May, but this increases to 64% in July–August.

Finally, it is worth to note that the use of either a smartphone or a tablet is more likely to lead to nonparticipation in the next wave of the survey. This may, however, be a sample selection effect. More loyal panel members may favor the PC to complete the questionnaires.

Measurement Error per Device

Now we turn to a discussion of the measurement error that is associated with each of the devices used to complete questionnaires. In Table 3, the pooled results for the amount of measurement error split by device are shown.

It is important to remember that any measurement differences that we find between the devices are not necessarily caused by the device being used. Rather, it could be that people who generally report with high measurement error have different device preferences from people who report with low measurement error. The differences in measurement error that we observe between the devices are large, as can be seen in Table 3. For example, pooled over the six waves, 4.1% of evaluation questions are missing for PC respondents, 7.4% for tablet respondents, and 12.2% for smartphone

Table 3. Measurement Error Indicators per Device in Six Waves of the LISS Panel.

	PC	Tablet	Smartphone	Total	ANOVA
% missing	4.10	7.35	12.18	4.52	$F(2, 29198) = 68.65$, $p < .01$, $\eta^2 = .005$
% completed open question	3.99	2.49	3.61	3.87	$F(2, 28429) = 5.99$, $p < .01$, $\eta^2 = .000$
Mean length of open answers (in characters)	6.90	3.29	4.27	6.57	$F(2, 28429) = 8.68$, $p < .01$, $\eta^2 = .000$
% Straightlining	10.02	8.12	9.29	9.86	$F(2, 29198) = 4.30$, $p < .01$, $\eta^2 = .000$
% primacy effect	4.43	4.82	14.29	4.64	$F(2, 5195) = 9.85$, $p < .01$, $\eta^2 = .004$
Mean number of answers in check-all-that-apply	.38	.36	.32	.38	$F(2, 5195) = 1.32$, $p = .27$
Mean duration of questionnaire	803	726	860	799	$F(2, 28392) = 8.36$, $p < .01$, $\eta^2 = .001$
Mean evaluation (factor score)	.01	−.17	−.37	−.01	$F(2, 22450) = 50.52$, $p < .01$, $\eta^2 = .004$

Note. LISS = Longitudinal Internet Study for the Social sciences; ANOVA = analysis of variance. $N = 29{,}901$ (pooled). For actual n at every wave, see Table 2. The "% primacy effect" and "mean number of answer checked" are based on $n = 5{,}198$, as they are only based on questions asked in April, June, and August and were only asked to a subset of respondents.
The data are clustered within respondents. For this reason, we have run a multilevel model, and found the Intraclass Correlation Coefficient (ICC) for the intercept-only model to be lower than .01 for all variables, except for the mean number of answers in the check all-that-apply question. The ANOVA results for the variable "mean number of answers in check-all-that-apply" may therefore be biased, and the p value we report underestimated.

respondents. We find similar patterns in all other indicators of measurement error—that PC respondents report with least measurement error, followed by tablet, and smartphone respondents. PC respondents are most likely to complete open questions, give the longest open answers, are the least likely to check the first answer box (primacy effect), and are most positive about the questionnaire.

The only exception is for straightlining. PC respondents straightline more often than tablet and phone respondents. We also find an effect for duration. Here, tablet respondents are fastest, and smartphone users are slowest. Note that the effect sizes of the differences that we find are generally small.

Measurement Error per Device Over Time

Our findings from Table 3 can be caused either by selection effects or measurement error properties of the device. Table 2 shows that tablet and smartphone users were more likely to switch devices over the 6 waves we analyze. If a switch of devices is associated with increased measurement error, this may explain partly why we find relatively large differences in measurement error across the devices. For this, we identify four groups: (1) a group that always uses a PC in all waves, (2) a group that switches between using a PC and tablet, (3) a group that switches between using a PC and phone, and finally (4) a group that always uses a tablet. We would have liked to also look at a group who always uses a smartphone, but Table 2 already show that this group is small ($n = 19$) so we chose not to show results for this group.

Table 4 shows that the differences we found in Table 3 persist longitudinally. The group of respondents that always uses a PC to complete questionnaires report with the lowest measurement error, with the only exception being straightlining. Switching from a PC to a tablet, or reversely, does lead to higher measurement error. The primacy effect is most likely to occur for this group, but straightlining is on the other hand least likely. Respondents who always use a tablet give the shortest answers to open questions and report with most item missing. Whereas in Table 3, smartphone users performed worst on the percentage of item missing and the primacy effect, this is no longer the case. This suggests that the differences we found in measurement errors between PCs and smartphones in Table 3 are averaged out in Table 4, and that it is not the switch of devices as such, but rather

Table 4. Longitudinal Device Use and Measurement Error.

	Always PC	Switch PC <-> Tablet	Switch PC <-> Smartphone	Always Tablet
% missing	3.72	4.65	6.79	7.69
% completed open answers	3.98	3.74	2.90	2.83
Mean length of open answers (characters)	6.92	5.87	3.72	3.43
% Straightlining	9.49	6.44	8.16	8.58
% Primacy effect	2.90	8.00	5.70	4.30
# of answers checked	.33	.33	30	.30
Mean duration in seconds	790	815	766	654
Mean evaluation (factor score)	.01	−.09	−.21	−.19
Sample size	4,502	572	286	169

Notes: $n = 5{,}529$. This excludes 578 people who completed less than two waves of data and 19 people who completed the survey always on their smartphone. The summary statistics in Tables 3 and 4 do not exactly corrrespond, as Table 4 uses aggregate data from every respondent who completed at least two waves. Table 3 includes all available data pooled over all waves.

selection effects, or measurement error that can be attributed to the device being used, that is causing the differences in measurement error. We take a more formal look at this question in the next section.

Device Transitions and Measurement Error

Finally, we turn to an analysis of the effect of a switch of device on measurement error. As respondent characteristics are likely to affect both measurement errors and self-selection into using a particular device, the analyses from here on focus on a longitudinal analysis of device use on measurement. With six waves of data, each respondent can experience five switches of devices. Table 2 shows that many respondents do not always respond in all waves. Only using the data where respondents answer in two consecutive waves yields 19, 264 valid device transitions. For every transition, we can code both the device switch and the associated change in measurement error for each of the seven aspects of measurement error that are reported in Tables 3 and 4. Because the scale of measurement error varies across the seven indicators, the distributions for each monthly transition were standardized. After this, the changes in standardized measurement error were pooled across the five waves, split for each of the nine possible device switches that occur in the data, and plotted in Figures 4 and 5. Because of a small sample size, we have chosen not to show the switches between tablet→phone ($n = 26$) and phone→tablet ($n = 14$).

Figures 4 and 5 show changes in measurement error over time, associated with a device switch. The most frequently occurring transition in our data is the use of PC at two subsequent waves (PC→PC). We see that the change in whether an answer is given to an open question, the length of the answer to that question, and the number of item missing is about 0. This means that, as may be expected, using the PC consistently over time does not lead to more or less measurement error over time. Any difference we find between waves is close to zero and nonsignificant for six of seven indicators. However, we do find PC→PC respondents to check slightly more answers in a check-all-that-apply question in the subsequent wave. This is true as well for respondents who consistently use a tablet (tablet→tablet) or smartphone (phone→phone) over two waves, although the increase here is nonsignificant due to a smaller number of transitions. The fact that we find an increase in the number of boxes checked when the same device is used over time may be attributed to respondents "learning" how to complete the survey, also called panel conditioning. The effect is, however, very small.

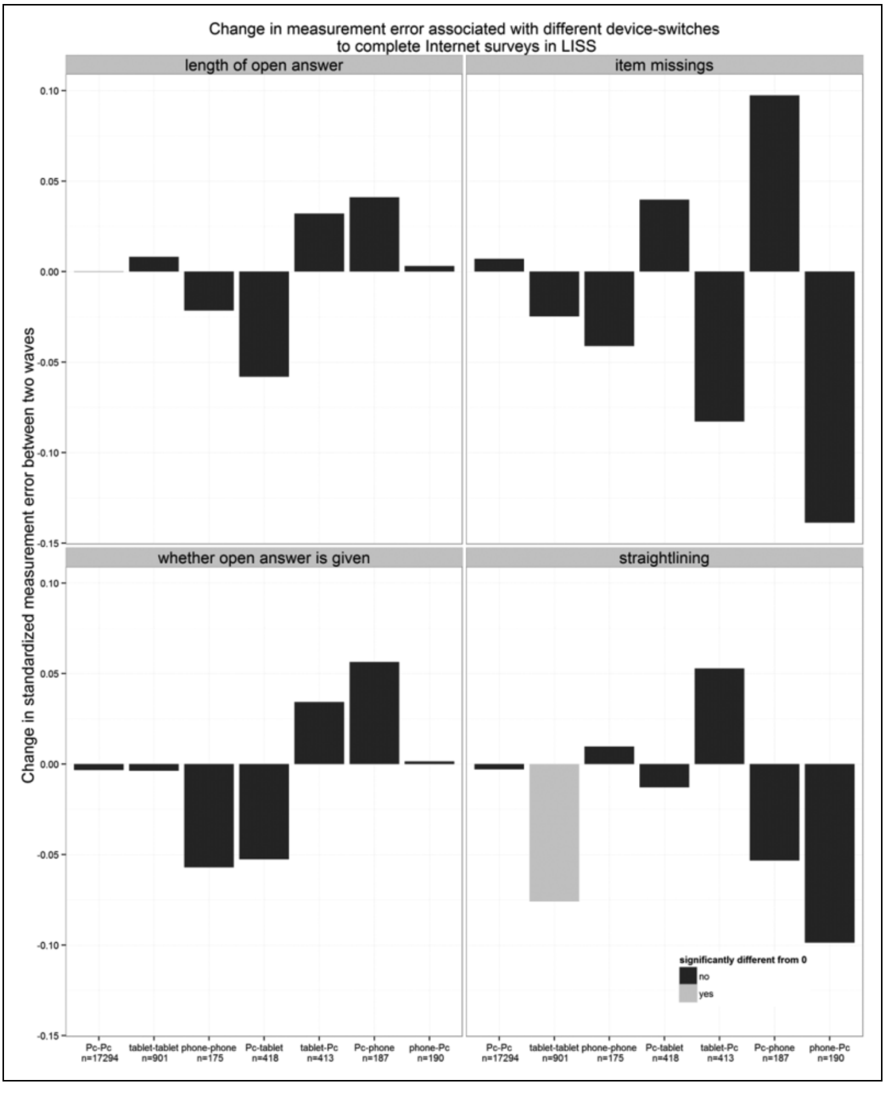

Figure 4. Changes in indicators of measurement (length of open answers, item missing, whether open answer is given and straightlining) associated with seven types of device switches observed in the LISS panel.

For respondents who either consistently use a tablet (tablet→tablet) or smartphone (phone→phone) in two consecutive waves, we find the changes in measurement error for all seven indicators to be either 0, or very small, and not significant. In short, we find that using the same device over time does not affect response behavior and that measurement errors stay about equal over time.

Rather than looking at respondents who use the same device over a two-wave transition, it is more interesting to look at response behavior that is associated with a device *switch* over time. A switch from PC to tablet (PC→tablet) or, reversely, from tablet to PC (tablet→PC) does not lead to a significant change in measurement error for five of seven indicators. The only significant effects are found for the number of answers given to the check-all-that-apply question and the evaluation of the questionnaire for a switch tablet→PC. A transition to a PC leads to more answers given, and a better evaluation of the questionnaire. The effects are of opposite sign for the reverse switch (PC→tablet), but nonsignificant.

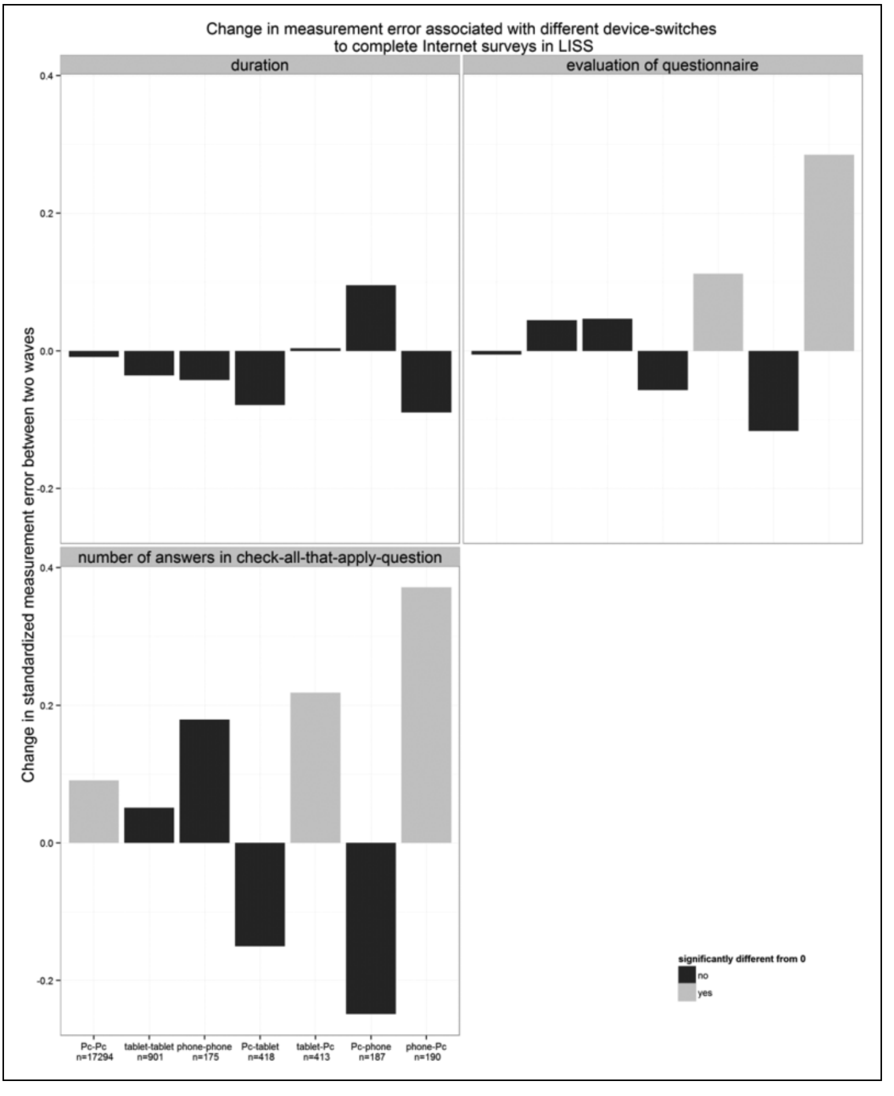

Figure 5. Changes in indicators of measurement (duration, evaluation, and number of answers in check-all-that-apply) associated with seven types of device switches observed in the LISS panel.

For transitions involving a smartphone and a PC, we find the largest effects, although most effects are again nonsignificant, due to the lower sample size of these transition groups. Similar to the switch tablet→PC, we find that the transition smartphone→PC leads to more answers being checked and a more positive evaluation of the questionnaire. Although we find large changes for many of the other indicators of measurement error, switches involving a PC and smartphone are not associated with significant changes in measurement error.

Conclusions and Implications

In this article, we have investigated measurement error in Internet devices over time in an Internet panel. Web surveys can nowadays be completed on different devices, such as desktop PC, tablets, and smartphones. Literature on device effects is still in its infancy. Moreover, it is difficult to disentangle selection effects (respondents can have their own preferred device) from measurement effects. Therefore, we looked at switches in devices that respondents use to complete questionnaires in the LISS panel longitudinally. In this way, we could rule out selection effects, and focus on measurement error associated with the type of device used.

Our results show that about 90% of the surveys are completed on PCs. This is similar to results found in the literature. At the time of this study, 26% of all respondents in LISS reported to own all three types of devices. If in future more people own Internet-enabled tablets and smartphones, it is likely that more questionnaires are completed on such devices. Panel members who complete surveys on their PC show less measurement error with regards to the number of item missing, open ended questions, primacy, and mean number of answers in a check-all-that-apply format. They also evaluate the questionnaire as more positive. On the other hand, they straightline more often than tablet and smartphone users.

Our study also shows, however, that measurement error does generally not increase when a respondents over time switches from a PC to a tablet or smartphone. This implies that the measurement error differences that we find between the devices should not be attributed to the device being used, but rather to the respondents. Those respondents who are likely to respond with more measurement error in surveys are also more likely to use tablets or smartphones to complete questionnaires. These findings only apply to the situation where respondents are allowed to self-select into the device they use. It is conceivable that had we encouraged or forced some of the PC respondents to use a tablet or smartphone, we would have found different or larger effects on measurement error. Similarly, respondents who are pressed for time in a particular month may have been more likely to complete the questionnaire on a mobile device, possibly leading to larger errors on such devices. The fact that we find that measurement errors do not change when a respondent switches device indicates that this effect, if it exists, is probably small.

Also, this article leaves the question open whether some respondents are perhaps better at completing surveys on a specific device. Using the data in our article, we were not able to link the measurement errors to respondent characteristics, such as demographics or their tenure as a panel survey respondent. So although we find no average effect of device on measurement error, the interaction effects of device and respondent characteristics on measurement error should be a theme for further research.

Looking at transitions between devices over waves, we see that panel members who switch between smartphones and tablets/PC show somewhat larger measurement error. Most of these changes were not significant however, due to the fact that the sample sizes for these transitions are often small. This suggests that survey completion on a smartphone is something different than survey completion on PC or tablet. We conclude therefore that screen size, or perhaps the speed of the Internet connection used, is more important than method of data entry (touchscreen). Similar conclusions were drawn by De Bruijne and Wijnant (2013). Future research is needed to see how to design surveys for smartphones. Although research for mobile survey completion can draw on the visual design principles for web surveys (see Dillman, 2007; Toepoel & Dillman, 2011), it needs independent testing to see where and how differences in survey processing on small devices such as smartphones occur.

Our results show that respondents who use smartphones rated the survey less positive than when they used a PC. This suggests that respondents do not appreciate survey completion on a smartphone as much as on a computer and that the design of surveys for mobile completion should be a priority.

Declaration of Conflicting Interests

The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by a Future Leaders Grant by the UK Economic and Social Research Council [ES/K001027/1] awarded to the first author and supported by a VENI grant awarded to the second author by the Netherlands Organization for Scientific Research.

Note

1. The effects may differ under different operationalizations of the primacy effect. For example, the primacy effect could also be measured using the proportion of answers checked from the first three response options. We tested for this and did not find in any different results, implying that the primacy effect is consistent across the first three answer categories.

References

Alwin, D. F. (2007). *Margins of error - A study of reliability in survey measurement*. Hoboken, NJ: Wiley.

Binswanger, J., Schunk, D., & Toepoel, V. (2013). Panel conditioning in difficult attitudinal questions. *Public Opinion Quarterly*, *77*, 783–797.

Bosnjak, M., Poggio, T., Becker, K. R., Funke, F., Wachenfeld, A., & Fischer, B. (2013). *Online survey participation via mobile devices*. Paper presented at the Annual Meeting of the American Association for Public Opinion Research, Boston MA. Retrieved 29, July 2013, from http://www.psyconsult.de/bosnjak/pages/publications/conference-contributions.php

Callegaro, M. (2013a). Do you know which device your respondent has used to take your online survey? *Survey Practice*, *3*.

Callegaro, M. (2013b). Paradata in web surveys. In Kreuter (Ed.), *Improving surveys with paradata: Analytic uses of process information* (pp. 263–282). Hoboken, NJ: Wiley.

Callegaro, M., & Disogra, C. (2008). Computing response metrics for online panels. *Public Opinion Quarterly*, *72*, 1008–1032.

Couper, M. P. (2013). *Surveys on mobile devices: Opportunities and challenges*. Paper presented at the workshop ‘Web Surveys for the General Population: How, Why, and When?, London, 26 February 2013. Retrieved September 12, 2014, from http://www.websm.org/uploadi/editor/1364314923Couper_2013_Surveys_On_Mobile_Devices.pdf

Couper, M. P., Tourangeau, R., & Conrad, F. (2004). What they see is what we get response options for web surveys. *Social Science Computer Review*, *22*, 111–127. doi:10.1177/0894439303256555

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*, 482–504. doi:10.1177/0894439313483976

Dillman, D. A. (2007). *Mail and internet surveys - the tailored design method* (2nd ed.). Hoboken, NJ: Wiley.

Eurostat. (2012). *Statistics in focus*, *50*. Retrieved September 12, 2014, http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-SF-12-050/EN/KS-SF-12-050-EN.PDF

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

Guidry, K. R. (2012). *Response quality and demographic characteristics of respondents using a mobile device on a web-based survey*.” Paper presented at the Annual Meeting of the American Association for Public Opinion Research, Orlando, FL, 17-20 May 2012.

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

Lynn, P., & Kaminska, O. (2013). The impact of mobile phones on survey measurement error. *Public Opinion Quarterly*, *77*, 586–605. doi:10.1093/poq/nfs046

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*, 191–205.

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

Peytchev, A., & Hill, C. A. (2010). Experiments in mobile web survey design similarities to other modes and unique considerations. *Social Science Computer Review*, *28*, 319–335. doi:10.1177/0894439309353037

Toepoel, V., Das, M., & Van Soest, A. (2009). Design of web questionnaires: The effects of the number of items per screen. *Field Methods*, *21*, 200–213. doi:10.1177/1525822X08330261

Toepoel, V., & Dillman, D. A. (2011). Words, numbers and visual heuristics in web surveys: Is there a hierarchy of importance? *Social Science Computer Review*, *29*, 193–207. doi:10.1177/0894439310370070

Toepoel, V., & Lugtig, P. (2014). What happens if you offer a mobile option to your web panel? Evidence from a probability-based panel in the Netherlands. *Social Science Computer Review*, *32*, 544–560. doi:10.1177/0894439313510482

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

Wells, T., Bailey, J. T., & Link, M. (2014). Comparison of smartphone and online computer survey administration. *Social Science Computer Review*, *32*, 238–255. doi:10.1177/0894439313505829

Author Biographies

Peter Lugtig is a member of the Department of Methods and Statistics, Utrecht University, and is also a member of the Institute for Social and Economic Research, University of Essex. He may be contacted at p.lugtig@uu.nl.

Vera Toepoel is a member of the Department of Methods and Statistics, Utrecht University and may be contacted at v.toepoel@uu.nl.