An experiment comparing grids and item-by-item formats in web surveys completed through PC and smartphones

Melanie Revilla, Daniele Toninelli and Carlos Ochoa

RECSM Working Paper Number 46
November 2015

http://www.upf.edu/survey/_pdf/RECSM_wp046.pdf

An experiment comparing grids and item-by-item formats in web surveys completed through PC and smartphones

Melanie Revilla, *RECSM-Universitat Pompeu Fabra, Barcelona, Spain*

Daniele Toninelli, *University of Bergamo, Bergamo, Italy*

Carlos Ochoa, *Netquest, Barcelona, Spain*

Acknowledgments

We are very grateful to Netquest for providing us with the necessary data, to the University of Bergamo (this research has been partially supported by the 60% University funds), and to Aigul Mavletova and Mick Couper for their help and advices in setting up the experiment. The authors would also like to acknowledge the contribution of WEBDATANET (COST Action IS1004, http://webdatanet.cbs.dk/).

Abstract

Purpose: Some respondents already complete web surveys via mobile devices, even when this is unintended. However, these devices vary at several levels from PCs (e.g. screen size). Therefore, it is necessary to study the comparability of the data collected across devices. In particular, we expect differences when grids questions are used due to the device characteristics (lower visibility on mobile devices) and because in questionnaires that optimized to be completed through smartphones, grids are automatically split up into an item-by-item format.

Design/methodology/approach: This paper reports the results of a two-waves experiment conducted in Spain with the online fieldwork company Netquest in February-March 2015. In each wave, respondents were randomly assigned to three groups: PCs, smartphones not-optimized, or smartphones optimized.

Findings: Similar levels of interitem correlations are observed in all groups. Longer completion times are found in the case of grids for the smartphone respondents. Less non-differentiation is observed in the PC group for some questions. For smartphones, the non-differentiation is reduced by splitting grids into separate questions. In general, the results suggest that both the device used and splitting the grids affect the results.

Practical implications: Using the item-by-item format for both smartphones and PCs seems to be the most appropriate way to improve comparability.

Originality/value: Most online panels are still using grids in the surveys proposed to their participants. This research shows that this practice creates problem of comparability of the data when smartphones are used.

Keywords: Web surveys, smartphones, grids, interitem correlation, completion time, non-differentiation

1) Introduction

In web surveys, when a set of items share the same scale, it is common to present them in a grid. This mainly happens when the items aim to measure the same concept or at least are about a similar topic. "In grid questions, a series of items is presented (usually in rows), sharing a common set of response options (usually in columns), asking one or more questions about each item" (Couper et al., 2013, p.322).

Using grids presents both pros and cons. First, grids have the advantage of presenting the information in a more compact way, since the scale is shown only once. This reduces the need for scrolling or changing webpage. Several studies (Couper, Traugott, and Lamias, 2001; Tourangeau, Couper and Conrad, 2004) showed that the completion times are significantly shorter when grids are used instead of separate questions, both if the questions are proposed on the same webpage or on separate ones. In principle, this is attractive for the respondents (who usually prefer shorter surveys), but also for the researcher (when incentives are related to the expected survey length). Nevertheless, Thorndike et al. (2009) and Toepoel, Das, and van Soest (2009) showed that the satisfaction of the respondents with the survey experience is reduced when grids are used, even if the completion time is shorter.

Moreover, there is some evidence that using grids can increase the interitem correlation (as measured by Cronbach's α) between items measuring the same concept (Tourangeau et al., 2004). This is also considered as a positive aspect: Cronbach's α generally increases as the intercorrelations among test items increase (Cronbach, 1951). Therefore, it is often used as an internal consistency estimate of the reliability of a concept. However, other studies did not find this relationship (e.g. Bell, Mangione, and Kahn, 2001; Callegaro et al., 2009). In addition, an increase of the interitem correlation does not necessarily indicate a higher reliability of the obtained answers: it can result from common method variance (CMV) or a higher level of non-differentiation (Peytchev, 2005). CMV occurs when there are systematic measurement errors due to the use of a common scale. The common scale affects the answers of the respondents creating extra correlation between observed items that is not coming from the latent construct of interest. Because of the CMV, the observed correlations are over-estimated. Non-differentiation is a kind of strong satisficing (Krosnick, 1991), i.e. a tendency of not putting the maximum effort in answering the questions. In the most extreme form of non-differentiation (called straight-lining), the respondent effort is reduced by choosing always the same answer category, independently of his/her real opinion and of the item s/he is asked about. Long grids of questions in web surveys elicit particularly these kinds of undesirable behaviours. Couper et al. (2013) suggest that one reason explaining the increase of non-differentiation is that matrices the condensed information can give the impression to the respondents of a long, complicated task. This can discourage respondents from putting the maximum effort.

Another disadvantage of grids is that they tend to increase the dropout rates (Jeavons, 1998; Puleston and Sleep, 2008). Some studies also found that grids increase the proportions of item missing values (Iglesias, Birk, and Torgerson, 2001; Manfreda, Batagelj and Vehovar, 2002; Toepoel et al., 2009). Since it is a common practice in web surveys to force respondents to give an answer, this can also be related with the increase in drop-out rates. However, these results are not confirmed by other studies (e.g. Couper et al., 2001). The differences between these findings might be related to the fact that the listed studies are based on grids with different characteristics (e.g. numbers of items, of answer categories, topics). Hardy (2009) found a lower data quality for grid questions. Grandmont et al. (2010), confirmed this finding, and investigated how to improve the design of grids in order to reduce these negative effects.

Going further in that direction, Couper et al. (2013) proposed to improve grids' design as much as possible by limiting the size of grids (splitting them into two or three smaller grids) and taking more advantage of the possibilities Internet offers to simplify the respondents' tasks. In particular, the authors suggest using dynamic shading, such that each row of the grid changes colour as soon as an answer has been selected.

But seeing that there seem to be more disadvantages than advantages, others recommend to avoid grids and to rather use separate questions (Poynter, 2001; Wojtowicz, 2001; Dillman, Smyth and Christian, 2009). However, in practice, grids have been and still are frequently used in web surveys (Couper, 2008; de Leeuw, Hox and Dillman, 2008; Tourangeau, Conrad. and Couper, 2013).

Nowadays, with the recent spread of mobile devices and their increased share in terms of Internet access (see, for example, Revilla et al., 2015), a new phenomenon appears: the ‘‘unintentional mobile response’’ (Peterson, 2012) or ‘‘unintended mobile response’’ (Wells, Bailey, and Link, 2013; de Bruijne and Wijnant, 2014). This concept defines the attempts to complete web surveys using mobile devices (especially tablets or smartphones) when the survey is designed for PCs and is not adapted for mobile browsers. This phenomenon increased very quickly in the last years in different countries (Callegaro, 2010; Bruijne and Wijnant, 2014; Revilla et al, forthcoming).

The fact that respondents access the surveys through mobile devices generates new challenges for designing web surveys (Fuchs and Busse, 2009; Baker-Prewitt, 2013; Couper, 2013). Indeed, mobile devices differ from PCs in several aspects, like the size and direction of the screen, the kind of keyboard, and the mobility. The screen and the keyboard characteristics can affect the visibility, making it necessary to zoom or scroll. It can also affect the difficulty of selecting an answer (e.g. because of the smaller buttons). All this affects the level of efforts respondents have to do. Mobile devices usually also show a lower loading speed of webpages and slower Internet connection (de Bruijne and Wijnant, 2013a). Moreover, the respondent's cognitive processing and comprehension of the questions can be affected by the location (Peytchev and Hill, 2010). Several studies (de Bruijne and Wijnant, 2013a; Revilla et al, forthcoming) show that a large majority of respondents answer web surveys from home or workplace even when using a mobile device. Nevertheless, part of the respondents participates to surveys using mobile devices from other places, which can decrease their concentration. In addition, respondents using mobile devices to complete the survey may be more prompt to multi-tasking (i.e. doing other activities while answering the survey), even when answering from home. All in all, this leads to longer perceived and objective completion times, but also to higher dropout rates for mobile web respondents. Finally, this can also affect the quality of the collected data, even if few differences have been found in some previous studies (e.g. de Bruijne and Wijnant, 2013a, and Mavletova, 2013). However, despite all these drawbacks and difficulties, given that getting people involved in surveys is becoming increasingly harder and for representativeness reasons, this problem cannot be solved by simply preventing respondents to participate through their mobile devices.

Some research focused on how to adapt web questionnaires to mobile devices (Boreham and Wijnant, 2013) in order to improve both the survey experience and data quality. In this framework, the case of grids is particularly challenging. If grids were already problematic in PC web surveys, moving to mobile web surveys, the disadvantages are even higher, mostly if surveys are completed through smartphones. The visibility is reduced and the complexity of answering increases. Therefore, different authors (Macer, 2011; Couper, 2013) and websites (e.g. http://www.survey.bris.ac.uk/support/creating-accessible-online-surveys or http://www.surveysystem.com/blog/scales-in-survey-questionnaires/) recommend not using grids for surveys completed through smartphones. Contrary to PC surveys, where many researchers do not follow the recommendation and continue using grids, with optimized mobile web surveys grids are often automatically split-up into separate questions (Macer, 2011; de Bruijne and Wijnant, 2013b; Lorch and Mitchell, 2014; Toepoel and Lugtig, 2014). Thus, in many current web surveys collected by online panels using optimized mobile questionnaires, respondents answering through PCs are asked to answer questions in grid formats, whereas respondents answering through mobile devices are automatically redirected into an optimized mobile web version of the questionnaire in which grids are replaced by separate questions.

In this context, our main goal is to investigate what is the impact of the used device (PC or mobile device), and of the questions layout (as a grid or as separate questions) on the following aspects: 1) the interitem correlations (measured by the Cronbach's α), 2) the completion time of a given set of questions, and 3) the non-differentiation. Peytchev and Hill (2010) did not find significant differences in mobile web surveys between a grid and a separate questions format. However, these findings were obtained studying data form a small experiment (92 adults), in which a smartphone was provided to respondents not owning one. This is a very different setting than the one we are interested in: mobile respondents within large international opt-in panels, who use their own mobile devices (they are hence familiar with the device).

Moreover, Maxl and Baumgartner (2013), studying a population of university students, highlighted how much the typology of mobile devices can vary across respondents. The different device characteristics can lead to different effects. This suggests that it may be interesting to consider additional information such as the size and the direction of the screen for respondents answering through mobile devices.

Section 2 presents further the main hypotheses. Section 3 describes the experiment implemented in order to test them, and the data used for this purpose. Section 4 shows the main results. Finally, Section 5 proposes some elements of discussion, the main limits of this work and ideas for further research.

2) Hypotheses

According to the literature, we expect that splitting grids into separate questions will have an effect on the answers, because it reduces the condensed presentation of the information. We focus on three effects. First, we hypothesize that it will have an effect on the interitem correlation. Even if the results are mixed in the literature, following Callegaro et al. (2009), we expect a lower Cronbach's α in grids with both positively and negatively formulated items: the reversed direction might be more confusing in grids than in an item-by-item format. However, when all items go in the same direction, we expect a higher correlation with grids. Second, we expect shorter completion time when grids are used, because the information is presented in a condensed way. Third, higher non-differentiation is expected with grids.

In summary, we want to test the following hypotheses about the effect of splitting grids into separate questions:

- **H1a**: it decreases the interitem correlation when all items have the same direction, and increases it when there are both positively and negatively formulated items

- **H1b**: it increases the completion time
- **H1c**: it reduces the non-differentiation

Moreover, we expect that the device will have an effect because of differences in the kind of keyboard, the screen size, and the level of mobility. Lorch and Mitchell (2014) mention that, on average, a smartphone screen is around 5% of the size of a desktop PC screen, and 10% of the size of a laptop screen. This smaller size of the screen is expected to be burdensome, particularly for grids, because horizontal scrolling becomes necessary in most of the cases and because it pushes the respondents to frequently rotate the direction of the screen. This should lead to longer completion times for smartphones than for PCs. In addition, if respondents only select from the categories they can see on the screen without scrolling horizontally, we expect less variance in the answers. This implies a higher level of non-differentiation. Finally, mainly when the items are formulated in the same direction, the interitem correlation can also increase if respondents choose only from the subset of answers visible on the smartphone screen without scrolling.

To summarize, our main hypotheses about the effect of answering through smartphones rather than PCs are the following:

- **H2a**: it increases the interitem correlation in grids
- **H2b**: it increases the completion times for grids
- **H2c**: it increases the non-differentiation in grids

In addition to our main hypotheses, we also study what is the effect of the device when the items are presented as separate questions rather than as grids. Indeed, instead of having the current mix of grids in PC and separate questions in smartphone optimized surveys, or instead of having grids in both devices, we could use separate questions in both devices. de Bruijne and Wijnant (2013b) propose to investigate two strategies: optimizing the survey layout for each type of device separately or optimizing a single design that is compatible with all devices. Lorch and Mitchell (2014) recommend the second option. They suggest to "start with the mobile design and size up" (slide 48). Thus, instead of starting by designing surveys for PCs and then adapting them for smartphones, researchers should start designing surveys for mobile devices and, then, use a similar design for PCs. In addition, grids are also problematic in PC surveys, and several authors recommend not using them. Therefore, we want to find out if it would be a good option for future data collection in web surveys to split grids into separate questions, not only in smartphones but also in PCs.

3) Method and data

3.1 Experimental design

In order to test our hypotheses, a two-wave experiment was carried out. This research focuses only on panellists that have Internet access on both a PC and a smartphone, since in an online access panel, they are the ones who can choose to switch devices. Moreover, it only focuses on smartphones because we expect more effects than for tablets.

The first wave of data collection lasted one week, the second one and a half. Between the two waves a one week break was introduced to limit as much as possible the possibility of changes in respondents' opinion and, at the same time, to avoid memory effects. The survey was identical in both waves, except for a more developed introduction in wave 1.

We focus on the measurement effects and not on the selection aspect. Thus, we try to maximize the proportion of respondents answering to both waves of the survey by doing the following: first, the introduction of wave 1 provides respondents with important information about the design of this experiment (e.g. that it includes two waves, or that a specific device should be used). In order to encourage panellists to answer both waves, a larger incentive is given for the participation in the second wave. Moreover, the introduction of wave 1 also included two questions to allow us selecting respondents of the population of interest: "Do you have access to Internet through a PC?", and "Do you have access to Internet through a smartphone?" Panellists who do not have access to both devices are filtered out.

Then, it also includes a question to select only respondents that are willing to commit themselves to complete both waves and with the required device: "Do you commit yourself to answer to the two waves of the survey and do it using the device we ask for?" Panellists who refuse to commit are filtered out. Even if respondents had to commit themselves to use the devices required, we expected that some of them could try to participate through a different device. It already happened in previous experiments: e.g. Mavletova and Couper (2013) had to "exclude from the analysis those respondents who were assigned to the mobile Web survey but attempted or completed the survey via PC" (p.195). Thus, in order to be sure that respondents are using the required device, we implemented an automatic check of the device. If the registered device was not the one they had to use, respondents got a message saying and had to connect through the required device in order to continue the survey.

Later, during the second wave, reminders were sent on days 3, 5 and 7, stressing the importance of the participation of the panellists in wave 2.

In the first wave, panellists who had access to Internet through both a PC and a smartphone were randomly assigned to one of the following three groups: PC, smartphone non-optimized, or smartphone optimized.

In the PC group, the survey is presented to the respondents in the usual way (i.e. the design was the most commonly used in the panel).

In the smartphone non-optimized group (SNO), the survey page is a smaller version of the PC webpage. It does not adapt to the screen size. Zoom-in and/or scrolling vertically and horizontally are usually necessary to see all the information, in particular when questions are presented in a grid format. Sometimes, even zooming-in, it is difficult to see well all the information. Selecting the desired option and going to the next page may be difficult too. Respondents can more easily make mistakes.

In the smartphone optimized group (SO), the survey program recognizes the device and optimizes the survey for mobile participation. The survey page is adapted to the size of the screen, such that respondents do not need to zoom-in nor scroll horizontally. Unnecessary elements are limited. The size of the buttons is increased. All in all, the layout is intended to facilitate the respondents' task of reading and answering the questions. In addition, grids are transformed into a set of single questions. The questions are all presented on the same page, but one after each other, with a repetition of the scale.
Cross quotas for age and gender were used to guarantee the distribution for these variables in the sample was similar to the one observed in the panel.

All respondents that finished completing the first wave were invited to participate in the second wave. They were randomly assigned again to PC, SNO, or SO. Thus, combining the two waves, we can distinguish nine different groups, that are presented in Table 1.

Table1

In wave 1, the target sample size was 200 respondents finishing the survey per group, corresponding to a total of 1,800 complete surveys. In wave 2, the goal was to get as close as possible again to 1,800 complete surveys.

3.2 Fieldwork

The experiment was carried out by Netquest (www.netquest.com), an online fieldwork company accredited with the ISO 26362 quality standard present in Portugal, Spain and Latin America. Netquest invites its panellists through email, using a list of persons that agreed to receive emails after they answered a short satisfaction survey proposed in one of the numerous websites collaborating with Nequest. Panellists are rewarded for each survey completed, depending on the estimated length of the questionnaire.

The experiment was conducted using panellists in Spain. Data of the first wave were collected from the 23rd of February to the 2nd of March 2015, and data of the second wave from the 9th to the 18th of March. In wave 1, in total, 3,317 panellists were contacted. 2,720 of them (82.0%) got to the introduction page that contained the filter questions. 3 of these 2,720 panellists (0.1%) said they did not have access to Internet neither through a PC nor through a smartphone; 91 (3.3%) said they had Internet access only through a PC; and another 75 (2.8%) that they had Internet access only through smartphone. These 169 panellists (6.1%) were filtered out. Out of the 2,375 panellists with Internet access through both a PC and smartphone, 119 (5.0%) did not accept to commit themselves to answer both waves and to use the required devices. They were excluded too. From the remaining 2,256 ones, 296 (13.1%) were required to continue the survey from a different device than the one they started with but did not do the switch even if they just had committed themselves. Another 17 (0.7%) were filtered out because they did not access Internet with a smartphone in the past 30 days. 1,843 panellists answered the first survey question after all the filters. 1,800 finished the survey, which was the objective. They were randomly divided into three experimental groups.

The 1,800 panellists who finished the wave 1 survey were invited to participate in wave 2. Out of them, 1,610 (88.9%) completed the whole wave 2 survey. Two let almost all the questions empty (but continue until the end), mentioning in one open question that they were not willing to answer again the same questions one week after the first wave. Our analyses focus on the 1,608 respondents that completed wave 2, which are divided into 554 (34.4%) for the PC group, 518 (32.2%) for the SO, and 536 (33.3%) for the SNO group. The size of the nine experimental groups presented in Table 1 is quite similar, varying from a minimum of 165 panellists in the PC-SO group (10.3% of the 1,608 studied respondents) to a maximum of 188 in the PC-PC group (11.7%).

3.3 Questions used

The questionnaire used in both waves was mainly about sensitive behaviours. Contrary to what is the common habit in Netquest surveys, respondents were allowed to continue the survey without answering a question. This was announced in the introduction. However, to limit the missing answers, every time respondents skipped four questions, they got a message encouraging them to answer.

For this paper's analyses, we use the questions coming from two grids. The first one deals with attitudes toward immigrants. It contains 14 items[1] asked using a fully labelled, balanced

5-point agree-disagree scale. The grid is balanced with some positively and some negatively formulated items. Thus, we expect that a given respondent would provide different answers for the different items.

The second one consists of 14 items about attitudes toward alcohol consumption, asked on a scale from 0 to 10. Only the end points of the scale are labelled, from "Totally bad" to "Totally good". All items are formulated in the same direction. However, some situations are expected to be much more acceptable (e.g. drinking alcohol at a party) than others (e.g. drinking alcohol at work, or when you are pregnant). Thus, we expect again different answers to the different items from the same respondent.

In addition, we also study three sets of questions about behaviours that are often considered as undesirable in Spain. The first set asks panellists how often they think each of 15 behaviours can be justified. The second set asks if they have ever done 15 behaviours. The third one asks about 9 behaviours related with alcohol consumption. Each set shares the same scale[2]. These sets are presented separate questions, with several questions on the same page. In the SO version, these sets of questions and the questions presented as grids in PC and SNO are presented in exactly the same way. By using these three sets of questions, we are able to study the effects of the device for a separate-questions layout.

Table 2 summarizes the information about the different questions: their main topics, the number of items in each set, the number of webpages on which these items were presented (e.g. for Set 2, the 15 items are divided into 2 pages, so we have 8 items on one page, and 7 on the other), and the range of the scale.

Table2

A list of all the items can be found in Appendix 1. The complete questionnaire proposed to the respondents in wave 1 is available at the following links: http://goo.gl/g9gAE4 for PC; http://goo.gl/5jF2vr for SO; and http://goo.gl/4c9d1C for SNO. The only difference between wave 1 and wave 2 questionnaires is the introduction.

4) Main results

4.1 Interitem correlation

In most surveys, grids are used to ask sets of questions that are about similar concepts. Thus, respondents may use the fact that items are presented in a grid format as an indicator that these questions are supposed to be related. This can push them to give the same answers to the different items. Therefore, presenting questions in a grid format can increase the interitem correlations. We investigate if the interitem correlation varies for grids and non-grids format, when answering the survey through PCs or smartphones.

The interitem correlation is measured by the Cronbach's alpha. Table 3 presents these Cronbach's alpha computed for each group in each wave. The negative items are automatically recoded into positive ones by Stata 12.1.

Table3

The interitem correlations are overall similar, both across groups (PC, SO, and SNO) and waves. The results do not support hypotheses **H1a** (splitting grids into separate questions in surveys completed through smartphones decreases the interitem correlation when all items go in the same direction, and increases it when there are positively and negatively formulated items) and also not **H2a** (answering through smartphone instead of PC increases the interitem correlation in grids).

4.2 Completion time

We are interested in the completion time for each of the five sets of questions described in Table 2, and how it may change depending on the device. For each respondent, in each wave, this paradata is registered. However, some respondents spent extremely long times on some pages. This usually indicates that they are interrupted and/or multitasking, i.e. performing different tasks at the same time. Thus, we use the median as our main indicator for the comparisons, which is more robust to outliers than the mean. Even if we are interested in the effect of distraction or multitasking on the completion time[3], if we have few respondents with really much longer completion times, this can affect too much the mean.

The two grids were presented on one page. Thus, Table 4 directly reports the median completion time (in seconds) for the corresponding page. The other three sets of questions were presented on several pages (respectively 4, 2 and 2, as indicated in Table 2). Therefore, we first computed for each respondent the total completion time for a given set of questions by taking the sum of the registered times for the different pages of this set. Table 4 presents the median of these total times for each group.

Table4

The completion times in wave 2 are in general shorter than in wave 1. This can be linked to the experience: the panellists already completed the same survey, thus the second time they go through it quicker. Nevertheless, the patterns of differences in medians across devices are very similar in both waves. The differences are all significant when comparing PC and SO or PC and SNO. Answering through PC is quicker both for grid questions and for item-by-item format. The device significantly affects the speed of answer. We should underline that we use client-side response time, so the observed differences cannot be caused by longer times of downloading in mobile devices. Comparing both smartphones groups, the non-optimized version usually leads to longer response times. Nevertheless, the differences are smaller than the ones observed between PCs and smartphones, and significant only in half of the cases. Besides, clear differences are not only present in the case of grids, but also for the item-by-item format. This suggests that splitting grids into separate questions is not the only reason explaining longer completion times: if panellists need to zoom and/or scroll horizontally, if the buttons are smaller and more difficult to select, the completion times might be longer, even when an item-by-item presentation is used.

Table 4 compares groups of respondents answering through different devices. Because of the random assignment to the experimental groups, we expect significant differences to be mainly due to the device used. However, our experimental design allows more precise analyses, since the same respondents participate in two waves. This enables the use of Mixed-Models (MM) with observations nested in individuals in order to study, for each set of questions, the effect on the completion time of the following factors: answering through PC instead of SNO ("PC" variable, in Table 5), splitting the grids in a SO version versus keeping them in SNO version ("SO" variable), and occasion, i.e. the effect of answering in wave 2 instead of wave 1 ("Wave 2"). Differences can occur in wave 2 for different reasons: because it is a different moment, because time has passed and people changed their mind, or because the same survey is repeated.

First, in order to check if the second level (individuals) presents enough variance, we compute the Interclass Correlation Coefficient (ICC). A rule of thumb is that about 10% of the total variance needs to be represented at a given level (Occhipinti, 2012, pp.4-5): if the ICC is lower than .10, then we do not need to consider the second level in representing the variance of the outcome (i.e., there is no design effect) and "we could go on to use OLS regression as usual".

In our analyses, 18.6% of the total variance is explained at the individual level for grid 1, and 66.5% for grid 2. Thus, MM seems appropriate for both grids. On the contrary, the ICCs are much lower than the threshold (10%) for the three sets of sensitive questions. Thus, in these three cases we use OLS regressions to study the completion time. Table 5 presents the coefficients and p-values of both analyses.

Table5

The results are mixed: if all explanatory variables are significant for grid 1, none of them is for grid 2: splitting the grid into separate questions in surveys completed through smartphones reduces the completion time for grid 1 (negative significant coefficient for SO versus SNO) but not for grid 2. Answering through PC instead of smartphone decreases the completion time for grid 1 only. Also the completion times are significantly reduced in wave 2, if compared with wave 1, only for grid 1.

For the three sets of questions, again, the results are mixed: PC and SO have no significant impact for set 1, whereas a significant impact is observed for sets 2 and 3. Moreover, the wave has a significant impact in set 3 (for $\alpha = 0.1$).

At this point, we can only make hypotheses: the mixed results might be linked to the topic of the grids (quite different), or to the kind of response scales used, or to the position of the grid in the questionnaire. Further research (e.g. a meta-analysis using many different grids with a variety of topics and scales) would be necessary to test the role of these factors.

4.3 Non-differentiation

The non-differentiation is measured by the variance of the answers observed for each individual across all items of a grid or set of questions (Krosnick, 1991). Thus, it is computed for each respondent and for each of the five topics described in Table 2. A variance of zero indicates the most extreme case of non-differentiation ("pure straight-lining").

Table 6 reports the average of the variances of each respondent's answers for each experimental group at both waves[4]. These results give a first idea about the non-differentiation depending on the device and questions format (grids of separate items). Few panellists with only missing values are excluded from the analysis. Since the panellists were randomly divided, significant differences across groups should mainly reflect the differences in treatments' settings.

We should note that a high non-differentiation can be really considered as a problematic behaviour only when positively and negatively oriented items were combined. In particular, for sets 2 and 3, which ask if people have ever done a series of socially undesirable behaviours ("yes/no" scale), we expect a significant proportion of respondents not to have done any of them. Thus, always saying "no" does not necessarily indicate satisficing. In addition, respondents can also choose the "no" answer because it is considered as the most socially desirable option. Nevertheless, the results are still interesting in order to compare the experimental groups.

Table6

For the grids, the results are mixed: if significant differences are observed for all groups in both waves for grid 2, there are only few significant differences for grid 1. Also, the SO group for both grids get a similar or higher variance than both other groups. For the three sets of sensitive questions, similar levels of non-differentiation are found across groups and waves.

Then, we use MM to estimate the effects on the non-differentiation of the same factors studied for the completion time. All the ICCs are much higher than the suggested 10% threshold, indicating that MM can be used for all five series of questions. 77.1% of the variance is explained at the individual level for grid 1, 59.8% for grid 2, 72.6% for set 1, 86.4% for set 2 and 84.7% for set 3. Table 7 shows the results of the MM analyses.

Table7

First, the wave is always significant: the variance is lower in wave 2 than in wave 1, which means that the non differentiation is higher. This is probably linked to the repetition of the survey, which decreases the motivation of respondents the second time.

Second, for the two grids the variance is higher when answering through a smartphone using an optimized version rather than a non-optimized version. This means that optimizing the design (splitting the grids into separate questions) leads to lower non-differentiation, as hypothesized in **H1c**. However, the coefficient for SO is not significant for the three sets of sensitive questions. But since the questions are presented in a question per question format in both the optimized and non-optimized versions, this is not surprising.

Finally, the coefficient for PC is significantly different compared to the reference category SNO in grid 2 but not in grid 1, nor in the three sets. In grid 2, the variance is higher for PC, which means that the non-differentiation is lower when answering through PC, as hypothesized in **H2c**. Once more, we see that grids 1 and 2 lead to different results, so there are probably other characteristics that are interacting here. Further research would be needed to study them.

5) Conclusions

In this paper, we studied the impact of the device (PC versus smartphone) and of the questions' layout (grid versus item-by-item) on the interitem correlations (measured by the Cronbach's alpha), on the completion time for a given set of questions, and on the non-differentiation (measured by the variance of a subject's answers). In order to do so, we implemented a two-wave experiment using the Netquest online panel for Spain. Our target population includes members who have access to both a PC and a smartphone. The selected panellists were randomly assigned to three groups in both waves 1 and 2: a PC group, a smartphone optimized group, or a smartphone non-optimized group.

By comparing first the three experimental groups in each wave, and second the answers of the individuals across waves using Mixed-Models with observations nested in individuals or simple OLS regressions, we found the following support for our hypotheses:

- **H1a**: "splitting grids into separate questions in surveys completed through smartphones decreases the interitem correlation when all items go in the same direction, and increases this correlation when there are both positively and negatively formulated items": hypotesis ***not supported***.

- **H1b**: "splitting grids into separate questions in surveys completed through smartphones increases the completion time": hypotesis ***partially supported*** *(support mainly for grid 1).*
- **H1c**: "splitting grids into separate questions in surveys completed through smartphones reduces the non-differentiation": hypotesis ***supported***.
- **H2a**: "answering through smartphones instead of PCs increases the interitem correlation in grids": hypotesis ***not supported***.
- **H2b**: "answering through smartphones instead of PCs increases the completion times for grids": hypotesis ***supported***.
- **H2c**: "answering through smartphones instead of PCs increases the non-differentiation in grids": hypotesis ***partially supported*** *(support mainly for Grid 2).*

Overall, even if some hypotheses are not supported, the results indicate that there are some significant differences between answers if the participation in the survey is made through a PC rather than by means of a smartphone, and if the questions are presented as a grid rather than using an item-by-item format. Nevertheless, further research is needed in order to study why significant differences are found for one grid and not for the other.

Therefore, our findings suggest that all web surveys' users should be careful in analysing their data and take into account that the device used can affect the results. In the case of the smartphone-optimized mobile surveys (where grids are split into separate questions), the risk of losing comparability is even higher. One possible way to improve comparability would be to avoid the use of grids in both PCs and smartphones surveys, and instead always use item-by-item format. Our results show that when this format is used (i.e. for all the three sets of questions), less or no significant differences are found across devices for the indicators studied. Thus, we would recommend using preferably an item-by-item layout in web surveys when there is a high probability to have respondents accessing by means of both PCs and mobile devices.

However, further research is still needed, to test the robustness of these results in different countries, for different topics, scales, and for different target populations (our findings are currently limited to individuals that have access to both PC and smartphones).

Besides, since our goal was to maximize the participation in both waves, we asked people to commit to participate in both waves. This may have led to selection bias. Thus, further study should deal with the external validity of the results. On the contrary, we expect a high internal validity, since the same persons answer twice.

In some cases, our analyses led to different findings for grid 1 and grid 2; this suggests that more research is needed in order to investigate which properties of the grids (topic, scale, position in the questionnaire, etc.) can affect the differences between devices. Finally, it would be interesting to develop a similar analysis for tablets.

References

Baker-Prewitt, J. (2013). Mobile Research Risk: What Happens to Data Quality When Respondents Use a Mobile Device for a Survey Designed for a PC?. CASRO Online Research Conference, San Francisco, US, March 2013. Available at:

http://c.ymcdn.com/sites/www.casro.org/resource/collection/0A81BA94-3332-4135-97F6-6BE6F6CEF475/Paper_-_Jamie_Baker-Prewitt_-_Burke.pdf

Bell, D.S., Mangione, C. M., and C.E. Kahn (2001). “Randomized testing of alternative survey formats using anonymous volunteers on the world wide web”. *Journal of the American Medical Informatics Association*, 8:616–620.

Boreham, R., and A. Wijnant (2013). Developing a web-smartphone-telephone questionnaire. IBUC 2013 15th International Blaise Users Conference. Available at: www.blaiseusers.org/2013/papers/4c.pdf.

Callegaro, M., Shand-Lubbers, J., and J.M. Dennis (2009). Presentation of a single item versus a grid: Effects on the vitality and mental health subscales of the sf-36v2 health survey. Paper presented at the annual meeting of the American Association for Public Opinion Research, Hollywood, FL, May.

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

Couper, M.P. (2008). *Designing effective web surveys*. New York: Cambridge University Press.

Couper, M.P. (2013). “Surveys on Mobile Devices: Opportunities and Challenges”. Presented in the NCRM Conference, London, February 2013. Available at: http://www.natcenweb.co.uk/genpopweb/documents/Mick-Couper.pdf

Couper, M.P., R. Tourangeau, F.G. Conrad and C. Zhang (2013). “The Design of Grids in Web Surveys”. *Social Science Computer Review* 2013 31(3):322-345. Originally published online January 2013. DOI: 10.1177/0894439312469865

Couper, M.P., Traugott, M., and M. Lamias (2001). “Web survey design and administration”. *Public Opinion Quarterly*, 65:230–253.

Cronbach, L.J. (1951). “Coefficient alpha and the internal structure of tests”. *Psychometrika* 16(3):297–334. DOI:10.1007/bf02310555

de Bruijne, M., and A. Wijnant (2013a). Comparing survey results obtained via mobile devices and computers: An experiment with a mobile web survey on a heterogeneous group of mobile devices versus a computer assisted web survey. *Social Science Computer Review*, 31(4):482-504. DOI: 10.1177/0894439313483976

de Bruijne, M. and A. Wijnant (2013b). “Can Mobile Web Surveys Be Taken on Computers? A Discussion on a Multi-Device Survey Design”. *Survey Practice*, 6(4): 1-8.

de Bruijne, M., and A. Wijnant (2014). “Mobile Response in Web Panels”. *Social Science Computer Review*, March 2014, 1-15. Available at: http://ssc.sagepub.com/content/early/2014/03/26/0894439314525918

de Leeuw, E.D., J.J. Hox and D.A. Dillman (eds.) (2008) *International Handbook of Survey Methodology*, Lawrence Erlbaum Associates, London.

Dillman, D.A., Smyth, J.D., and L.M. Christian (2009). *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. New York: Wiley.

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

Grandmont, J., Graff, B., Goetzinger, L., and K. Dorbecker (2010). “Grappling with grids: How does question format affect data quality and respondent engagement?” Annual meeting of the American Association for Public Opinion Research, Chicago, IL, May 2010.

Hardy, N. (2009). “Love for Sale: 10 ways to keep your panel respondents happy”. *Quirks*, January, 2009:40-43.

Iglesias, C.P., Birks, Y.F., and D.J. Torgerson (2001). “Improving the measurement of quality of life in older people: The york SF-12”. *Quarterly Journal of Medicine*, 94:695–698.

Jeavons, A. (1998). “Ethology and the web: Observing respondent behaviour in web surveys”. Proceedings of the Worldwide Internet Conference, London, ESOMAR.

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

Lorch, J. and N. Mitchell (2014). “Why You Need to Make Your Surveys Mobile-Friendly NOW”. Webinar organized by the American Marketing Association, May 2014: https://www.ama.org/multimedia/Webcasts/Pages/why-you-need-to-make-your-surveys-mobile-friendly-now-051514.aspx?tab=home

Macer, T. (2011). “Making it fit: how survey technology providers are responding to the challenges of handling web surveys on mobile devices”. In D. Birks, R. Banks, L. Gerrard, A.J. Johnson, R. Khan, T. Macer, P. Malloy, E. Ross, S. Taylor, and P. Wills (Eds), *Shifting the Boundaries of Research*, Proceedings of the Sixth ASC International Conference, University of Bristol, UK, September 2011, pp.259-281. Available at: http://www.asc.org.uk/publications/proceedings/ASC2011Proceedings.pdf#page=269

Manfreda, K.L, Batagelj, Z. and V. Vehovar (2002). “Design of Web survey questionnaires: Three basic experiments”. *Journal of Computer-Mediated Communication*, 7(3):0, DOI: 10.1111/j.1083-6101.2002.tb00149.x

Mavletova, A. (2013). “Data quality in PC and mobile web surveys”. *Social Science Computer Review*, 31(4):725–743.

Mavletova, A., and M.P. Couper (2013). “Sensitive Topics in PC Web and Mobile Web Surveys: Is There a Difference?” *Survey Research Methods*, 7(3):191-205.

Maxl, E., and T. Baumgartner (2013). Influence of Mobile Devices in Online Surveys. Poster presented at the General Online Research Conference (GOR), Mannheim, March 2013.

Occhipinti, S. (2012). "Mixed Modelling using Stata". Workshop for GSBRC, July 2012. Available at: http://www.griffith.edu.au/__data/assets/pdf_file/0011/439346/Stata_mixed_intro-1.pdf

Oppenheimer, D.M., Meyvis, T. and N. Davidenko (2009). “Instructional manipulation checks: Detecting satisficing to increase statistical power.” *Journal of Experimental Social Psychology*, 45:867–872. DOI:10.1016/j.jesp.2009.03.009

Peterson, G. (2012). Unintended mobile respondents. Conference paper, CASRO Technology Conference, May 2012, New York.

Peytchev, A. (2005). “How questionnaire layout induces measurement error”. Paper presented at the annual meeting of the American Association for Public Opinion Research, Miami Beach, FL.

Peytchev, A., and C.A. Hill (2010). “Experiments in mobile web survey design: Similarities to other modes and unique considerations”. *Social Science Computer Review*, 28(3):319-335. DOI: 10.1177/0894439309353037

Poynter, R. (2001). “A Guide to Best Practice in Online Quantitative Research”. In A. Westlake, W. Sykes, T. Manners, and M. Rigg (Eds.), *The Challenge of the Internet*; Proceedings of the ASC International Conference on Survey Research Methods (pp.3–19). London: Association for Survey Computing.

Puelston, J., and D. Sleep (2008). “The Survey Killer: A look at the impact of boredom on the respondent experience”. *Quirks*, November, 2008:54-58.

Revilla, M., Toninelli, D., Ochoa, C., and G. Loewe (2015). “Who has access to mobile devices in an online opt-in panel? An analysis of potential respondents for mobile surveys”. In D. Toninelli, R. Pinter, and P. de Pedraza (eds), *Mobile Research Methods: Opportunities and challenges of mobile research methodologies*, pp:119-139. London: Ubiquity Press. DOI: http://dx.doi.org/10.5334/bar.h.

Revilla, M., Toninelli, D., Ochoa, C., and G. Loewe (forthcoming). “Do online access panels really need to allow and adapt surveys to mobile devices?” *Internet Research*

Thorndike, F.P., Carlbring, P., Smyth, F.L., Magee, J.C., Gonder-Frederick, L., Ost, L.-G., and L.M. Ritterband (2009). “Web-based measurement: effect of completing single or multiple items per webpage”. *Computers in Human Behavior*, 25:393–401.

Toepoel, V., Das, M., and A. van Soest (2009). “Design of web questionnaires: The effects of the number of items per screen”. *Field Methods*, 21:200–213.

Toepoel, V. and P. Lugtig (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. DOI: 10.1177/0894439313510482

Tourangeau, R., Couper, M.P., and F.G. Conrad (2004). “Spacing, position, and order: Interpretive heuristics for visual features of survey questions”. *Public Opinion Quarterly*, 68:368–393

Tourangeau, R., Conrad, F.G. and M.P. Couper (2013). *The Science of Web Surveys*. Oxford University Press, 2013.

Wells, T., Bailey, J.T., and M.W. Link (2013). “Filling the void: Gaining a better understanding of tablet-based surveys”. *Survey Practice*, 6(1)

Wojtowicz, T. (2001). “Designing Lengthy Internet Questionnaires: Suggestions and Solutions”. In A. Westlake, W. Sykes, T. Manners, and M. Rigg (Eds.), *The Challenge of the Internet*; Proceedings of the ASC International Conference on Survey Research Methods (pp.25–32). London: Association for Survey Computing.

End notes

[1] The full list of items used in this paper is shown in Appendix 1. This grid also included an Instructional Manipulation Check, which is a "question embedded within the experimental materials [that] asks participants (…) to provide a confirmation that they have read the instruction." (Oppenheimer, Meyvis and Davidenko 2009, pp.867). However, we do not consider it in this paper, since it does not correspond to a real item.

[2] These questions are similar to the ones used by Mavletova and Couper (2013).

[3] The percentage of respondents that declared they were at least doing one other task than answering the survey is around 73% in both waves. However, we do not know if the multitasking occurred during the completion of the questions studied in this paper. Thus, we will not use this information further.

[4] Instead of using the answers' variance, we also developed the same analyses using the proportions of pure straight-liners as a strongest form of non-differentiation. Appendix 2 presents the results. Conclusions are overall similar, even if there are fewer significant differences.

Tables

Table 1: The nine experimental groups

	Group name	Wave 1 device	Wave 2 device
Control	PC-PC	PC	PC
	SNO-SNO	Smartphone not optimized	Smartphone not optimized
	SO-SO	Smartphone optimized	Smartphone optimized
Treat-ments	PC-SNO	PC	Smartphone not optimized
	SNO-PC	Smartphone not optimized	PC
	PC-SO	PC	Smartphone optimized
	SO-PC	Smartphone optimized	PC
	SO-SNO	Smartphone optimized	Smartphone not optimized
	SNO-SO	Smartphone not optimized	Smartphone optimized

Table 2: Summary of the questions used for our analyses

	Main topic	No. Items	No. Pages	Scale
Grid 1	Attitudes towards immigrants	14	1	From "1. Totally agree" to "5. Totally disagree"
Grid 2	Attitudes towards alcohol consumption	14	1	From "0. Totally bad" to "10. Totally good"
Set 1	Justification of undesirable behaviours	15	4	From "1. Always justified" to "4. Never justified"
Set 2	Undesirable behaviours done	15	2	Yes/No
Set 3	Alcohol consumption	9	2	Yes/No

Table 3: Interitem correlation in waves 1 and 2 by group

	Wave 1			Wave 2		
	PC	SO	SNO	PC	SO	SNO
Grid 1*	.94	.94	.93	.95	.96	.94
Grid 2	.87	.85	.88	.88	.87	.91
Set1	.79	.78	.77	.80	.78	.79
Set2	.59	.59	.59	.62	.62	.66
Set3	.76	.77	.75	.76	.78	.80

Note: * indicates that Grid 1 includes positively and negatively formulated items

Table 4: Completion time for the different sets of questions: Medians (in seconds)

	Wave 1						Wave 2					
	PC	SO	SNO	$p_{pc\text{-}so}$	$p_{pc\text{-}sno}$	$p_{so\text{-}sno}$	PC	SO	SNO	$p_{pc\text{-}so}$	$p_{pc\text{-}sno}$	$p_{so\text{-}sno}$
Grid 1	102	118	133	*.00*	*.00*	*.01*	85	104	108	*.00*	*.00*	*.26*
Grid 2	54	77	74	*.00*	*.00*	*.13*	50	71	67	*.00*	*.00*	*.07*
Set1	82	99	100	*.00*	*.00*	*.97*	68	78	80	*.00*	*.00*	*.73*
Set2	54	63	67	*.00*	*.00*	*.01*	48	52	60	*.00*	*.00*	*.00*
Set3	32	38	43	*.00*	*.00*	*.00*	28	32	37	*.00*	*.00*	*.00*

Note: $p_{pc\text{-}so}$ is the p-value of the two-sample Wilcoxon rank-sum (Mann-Whitney) test, the two samples being the PC and the SO groups in a given wave. Idem for $p_{pc\text{-}sno}$ and $p_{so\text{-}sno}$ but for respectively the two groups PC and SNO, and SO and SNO.

Table 5: Explaining the completion time using Mixed-Models (MM) or OLS regression

		Coefficient	*P-Value*
Grid1 (MM)	Wave 2	-20.4	*.000*
	PC	-24.8	*.000*
	SO	-9.5	*.019*
	Constant	165.9	*.000*
Grid2 (MM)	Wave 2	-844621	*.345*
	PC	-874853	*.538*
	SO	2261624	*.111*
	Constant	2145847	*.247*
Set1 (OLS)	Wave 2	-2623520	*.188*
	PC	-1282540	*.598*
	SO	1352896	*.580*
	Constant	5253307	*.127*
Set2 (OLS)	Wave 2	-1769831	*.160*
	PC	-2627405	*.087*
	SO	-2668858	*.084*
	Constant	5306885	*.015*
Set3 (OLS)	Wave 2	2687939	*.081*
	PC	-4024355	*.032*
	SO	-3958791	*.036*
	Constant	-43770	*.987*

Table 6: Non-differentiation: Average of the variances and p-values (t-test)

		Wave 1			Wave 2		
		PC	SO	SNO	PC	SO	SNO
Average variance	Grid 1	1.26	1.39	1.27	1.28	1.28	1.22
	Grid 2	6.43	6.98	5.94	5.97	6.63	5.49
	Set1	.76	.72	.72	.70	.64	.68
	Set2	.15	.15	.16	.15	.15	.15
	Set3	.14	.14	.14	.13	.13	.13
		$p_{pc\text{-}so}$	$p_{pc\text{-}sno}$	$p_{so\text{-}sno}$	$p_{pc\text{-}so}$	$p_{pc\text{-}sno}$	$p_{so\text{-}sno}$
p-value	*Grid 1*	*.03*	*.79*	*.06*	*.97*	*.32*	*.32*
	Grid 2	*.01*	*.03*	*.00*	*.00*	*.03*	*.00*
	Set1	*.18*	*.15*	*.90*	*.05*	*.56*	*.19*
	Set2	*.98*	*.27*	*.29*	*.67*	*.67*	*.99*
	Set3	*.80*	*.68*	*.88*	*.23*	*.27*	*.92*

Table 7: Explaining the variance of answers (non-differentiation)

		Coefficient	*p-value*
Grid1 (MM)	Wave 2	-.051	*.003*
	PC	.008	*.770*
	SO	.057	*.045*
	Constant	1.340	*.000*
Grid2 (MM)	Wave 2	-.410	*.000*
	PC	.531	*.000*
	SO	.930	*.000*
	Constant	6.360	*.000*
Set1 (MM)	Wave 2	-.063	*.000*
	PC	.016	*.243*
	SO	-.005	*.696*
	Constant	.800	*.000*
Set2 (MM)	Wave 2	-.003	*.003*
	PC	.000	*.972*
	SO	-.002	*.348*
	Constant	.160	*.000*
Set3 (MM)	Wave 2	-.009	*.000*
	PC	-.001	*.815*
	SO	-.003	*.189*
	Constant	.150	*.000*