methods, data, analyses | Vol. 9(2), 2015, pp. 185-212 DOI: 10.12758/mda.2015.011

Evaluation of an Adapted Design in a Multi-device Online Panel: A DemoSCOPE Case Study

Birgit Arn, Stefan Klug & Janusz Kołodziejski
DemoSCOPE

Abstract

In this paper, we look at the challenge of optimizing survey layout in online research to enable multi-device use. Several studies provide useful advice on target-oriented implementation of web design for CAWI surveys. This paper presents results of the implementation of a new adapted design at the panel of DemoSCOPE that allows the participants to take part in a survey on multiple (especially mobile) devices. To evaluate this adapted design, we compare interview data and question timing of panellists who participated in an insurance study before and after the design transition. Central key figures concerning the completion rate, item non-response, open questions, straightlining, timing of single questions and the length of the total interview are presented. In addition, we have presented examples of both old and new design to the community and invited them to assess these examples concerning orientation, color, design and usability. We evaluate the differences in these assessments before and after the design transition for smartphone and desktop users. We end with suggestions for best practice for online studies on different devices.

Keywords: CAWI surveys, design guidelines, simplicity, tile design

© The Author(s) 2015. This is an Open Access article distributed under the terms of the Creative Commons Attribution 3.0 License. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

1 Introduction

The visual design of CAWI surveys has become a pivotal topic within the area of market research. With the internet as the main form of communication and the extensive dissemination of mobile devices such as smartphones and tablets computers, market researchers need to adapt more than ever (Revilla et al., 2014; Brujine & Wijnant, 2014). The current technological and cultural conditions suggest a trend towards self-administration (Stern et al., 2014). If the use of self-administered surveys increases, so will the importance of a convenient and convincing visual design of those allowing for a multi-device mobile use.

It has long been recognized that because of the absence of an interviewer in self-administered surveys, respondents search for guidance within the questionnaire itself (Schwarz et al., 1991; Schwarz, 1995). Therefore, design elements such as symbols and graphical elements (spacing, font size, location, color and so forth) are crucial in guiding respondents through a questionnaire the way we want them to. During the 1990s the industry's focus was on a question's wording and how that affects the response process (e.g. Tourangeau et al., 2000). However, several studies had already indicated that visual changes of a survey questionnaire produce different outcomes. The importance of design features for the resulting data quality has been documented long since (e.g. Wright & Barnard, 1975; Wright & Barnard, 1978; Rothwell, 1985; Sanchez, 1992; Jenkins & Dillman, 1997).

With the wide distribution of the Internet during the 2000s and the subsequent proliferation of online research, the visual design of self-administered surveys and its consequences on different stages of a survey process has led to further studies on this matter. These studies support the notion that different design elements affect how people answer questions in self-administered surveys. There is much evidence that certain design choices, such as layout of a question (Christian & Dillman, 2004; Christian et al., 2007) or question-order effects (Krosnick & Alwin, 1987; Couper et al., 2001) are as important as the wording of a question. Furthermore, there is a great variety of issue-specific studies on survey design. For example, Dillman et al. (1993) tested the correlation between response rates and questionnaire design and found that shortening the questionnaire and utilizing a user-friendly design improved response rates of the U.S. decennial census (for an overview on response rates and questionnaire design see Vicente & Reis, 2010). A major part of these studies analysed different effects certain design choices had on surveys. For example: the placement, spacing, and sequence of answer options (Tourangeau et al., 2004), the use of images (Couper et al., 2004; Couper et al., 2007; Deutskens et al., 2004; Shropshire et al., 2009) or the question layout (Dillman & Christian, 2002;

Direct correspondence to
Birgit Arn, DemoSCOPE, Klusenstrasse 17, 6043 Adligenswil, Schweiz
E-mail: birgit.arn@demoscope.ch

Christian & Dillman, 2004) There are many more specific topics being looked at, like grid questions and web surveys (e.g. Couper et al., 2013), questionnaire design and nonresponse bias (Vicente & Reis, 2010) or invitation design (e.g. Whitcomb & Porter, 2004; Kaplowitz et al., 2012). Another practiced approach is to evaluate whether design effects differ with respondents' socio-demographic characteristics (e.g., Krosnick & Alwin, 1987; Knäuper et al., 2004; Fuchs, 2005; Stern et al., 2007; Tourangeau et al., 2007). The past years would suggest that the visual design of web-based surveys is as influential to a respondent's answers as any documented interviewer or wording effect (cf. Stern et al., 2014, p. 294).

Thus, the importance of a good web survey design seems to be evident. But, what is a good web survey design? In a fast-paced multi-device environment and changing user habits, surveyors need to be up to speed and recognize the importance of a state-of-the-art survey design. Besides the question of a good design, there are also technological constraints and nuances to take into account. In the next section, we discuss some specific, more technical challenges when it comes to using mobile devices.

In Section 3 we outline our attempt to offer an optimized web design to our online community. DemoSCOPE (www.demoscope.ch) is the third-largest market research company of Switzerland. To fulfil the high standard of the requirements of our clients we have built up a large online panel that we call the DemoSCOPE community. This community consists of about 30,000 active panellists which come from very diverse socio-demographical strata. The panellists are asked about twice a month to take part in an online survey. To keep the community members at it, we want to offer an optimal web design and the possibility to communicate with each other and directly with the community support at DemoSCOPE. To fulfil these aims we formulated the design guidelines which are presented in Section 3. Note that already 41% of our community participate in the surveys using a mobile device (27% smartphone, 14% tablet). Hence, our specific attention is on users of mobile devices.

In Section 4 we propose two ways to evaluate the adapted online design. First, we propose methods to compare the response behaviour of panellists which participated in an insurance study before and after the adaption of the new design based on the design guidelines. As a second idea we invited the community members to take part in a design evaluation, where we showed examples of the old and the adapted design. The task of the participants was to evaluate the shown screen using 4 different criteria: orientation, color, design and usability. In Section 5 and 6 we present the results of this evaluation. Section 7 contains our conclusions.

2 Specific Challenges when Using Mobile Devices

It is expected that in the near future internet traffic among mobile devices will exceed that of desktop computers (Buskirk & Andrus, 2012a). Smartphones represent a convenient tool for survey data collection, as they are a multimode device accessible through voice, text or web, including synchronous multimedia messaging (SMS) and an ever-increasing variety of apps. Not to mention the possibility to take a survey on the spot. However, the very same opportunities smartphones give also imply great variability with their different devices, operating systems and browser capabilities. As a result, the complication level for the implementation of online surveys for mobile versus desktop computers increases (Buskirk and Andrus 2012b). As the spread of smartphone usage is a relatively recent phenomenon, there is still only little literature on mobile surveys using smartphones and other devices (e.g. Raento et al., 2009; Fuchs & Busse, 2009; Buskirk & Andrus, 2012a; Buskirk & Andrus, 2012b; Mavletova, 2013; de Bruijne & Wijnant, 2013; Wells et al., 2014; Buskirk & Andrus, 2014).

With respect to online questionnaires, researchers nowadays must anticipate the diversity among the end user's device. Designing questionnaires for usage across such a variety of devices is not a matter of can-do attitude but rather already a must-do, as the end-user is also the one deciding on which device an online survey will be taken on.

Obviously, the main constraint is the screen size of the respective device used for survey participation. Screen sizes range from 14-40 inches for computers and laptops, 6-13 inches for tablets and 4-6 inches for smartphones, with the boundary values beginning to overlap across these categories. A web survey should therefore keep its functionality and desired look from the smallest smartphones to the widest TV-like PC screens. Web designers solved the multi-screen problem by following the rules of two main schools, namely adaptive web design (AWD) (Gustafson, 2012) as well as responsive web design (RWD) (Marcotte, 2010). With AWD a server sends the same data packages to each device and the browser of the corresponding device decides which of the upfront designed layouts to choose. Unlike the predefined device specific layouts AWD relies on, RWD uses fluid layout grids, flexible images and media queries to treat every viewport (device) the same way and adapts the layout according to the device's features. Without going into details, in both scenarios the layout of a webpage or, like in our case, an online survey is adapted to the screen used. The main difference is in how this adaption takes place – if it's using predefined solutions to exactly corresponding devices (AWD) or if it responds to any device thanks to a more fluid (flexible) way of defining one layout only (RWD). Despite the promise of an easy sounding solution, as the designer of our online survey we face a multitude of challenges when it comes to putting theory into practice. First, do we want to design a web survey device-specifically or do

we want to design a survey that adapts automatically to every viewport (device)? Furthermore, if we opt for the RWD solution, we still need to consider most of the imperatives on web questionnaires in general, irrespective of the nuances mobile research poses.

3 The Design Guidelines

In the beginning of 2014, DemoSCOPE changed its web questionnaire layout. On the one hand, the aim was to provide respondents with an enjoyable, convenient and mobile optimized design; on the other hand, it was as important to ensure functionality, feasibility and good data quality. Before we have a look at the new survey design and its properties, we firstly present the DemoSCOPE design guidelines.

We consciously decided to use an RWD approach where you develop one questionnaire design that then automatically adapts to the different devices and their parameters. In order to provide a mobile-optimized survey design, different constraints regarding the relatively small screen of smartphones had to be considered.

Firstly, to enable a reasonable legibility for smartphone users, we turned away from using a fixed font size. We changed the pixel-based size definition, which means from an absolute and rigid unit of measurement, to "em" – a relative unit equal to the currently specified point size (in any device or browser). The name used to refer to the width of the capital "M" in the typeface and size being used (the same as the point size). This enables to choose a reasonable ratio where the font size adapts to the actual screen size in use.

Secondly, given that only vertical scrolling is acceptable for smartphones, the use of grid questions should be avoided. There is no technical reason for the preference of vertical scrolling over horizontal scrolling, but it has emerged as the preferred usage and almost all apps and mobile friendly web pages are designed for vertical scrolling. Additional to the omission of horizontal scrolling, we decided to use a one-screen-per-page approach, where normally only one question per screen is displayed. This assures that respondents experience a stable and convenient survey flow. Apart from the no-scrolling advantage of a one-screen-per-page-approach, Couper et al. (2001) and Tourangeau et al. (2004) found that the intercorrelations between items presented on the same page are higher than when items are displayed sequentially on one screen per page. These authors also state that, although the effect as such does not seem to be severe, there is evidence that respondents use proximity among the items as a hint to their meaning, which results in a faster advancement within the survey. However, Couper et al. (2001) found that the one-item-per-screen-approach takes respondents more time to complete the survey than a multiple-item-per-screen approach.

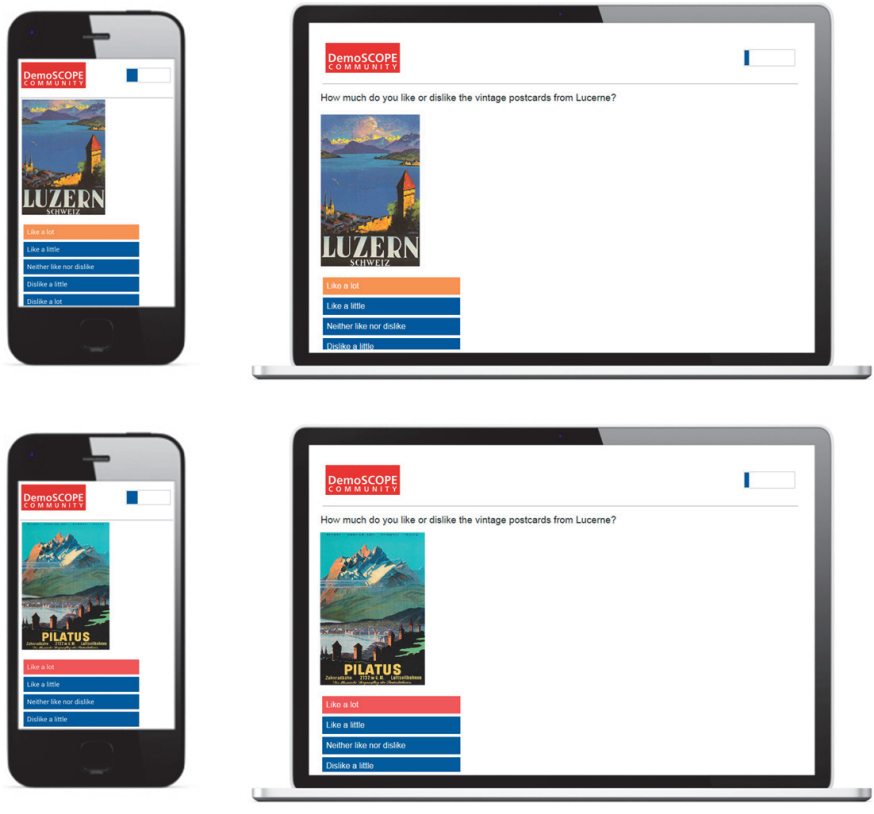

Figure 1 The HSM format for 2 examples: Vintage postcards of Luzern and Pilatus (smartphone and desktop version)

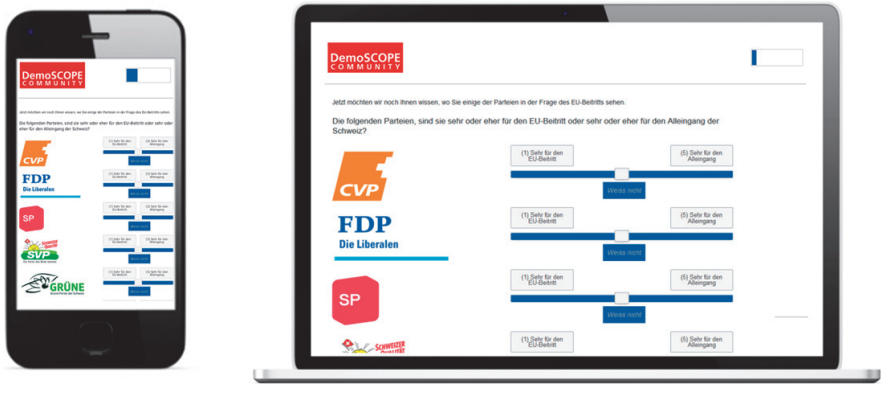

Figure 2 The visual scale sliders for an example with Swiss parties (smartphone and desktop version)

The usage of grid-questions imposes a problem not only to mobile-devices. It is a remainder of research with paper-and-pencil questionnaires, where print-outs were costly. Visually grid-questions make the questionnaire appear shorter, but have the disadvantage of non-careful reading and other negative effects such as straight-lining (Schaeffer & Presser, 2003). Klausch et al. (2012) tested a format where the answer-scale stays on one screen and the question is replaced by horizontally replacing one question with the next one (HSM: horizontal scrolling matrix format, not to be confused with "horizontal scrolling" by the respondent). These authors proof positive effects on data quality when using HSM formatted questions instead of grid-questions. As an example see Figure 1: The response scale stays the same for both examples ("Luzern" and "Pilatus"), but the shown vintage postcard is different. Sometimes a visual comparison between answers given is desired. For such cases, alternatively to the HSM format, we propose visual scale sliders that reduce the scaling-dimension such that it can be displayed on one screen together with the line of statements (see an Example in Figure 2).

Obeying the one-item-per-screen with limited scrolling policy, we introduced an auto-submit function for single-choice items. This enables the respondent to proceed to the next question as soon as he or she selects the answer. However, the use of the auto-submit function carries certain risks, especially when applied on small screens, since some respondents may not notice that they have already progressed to the next question and mix up answers.

Further, we quit using Flash-based elements, complex headers, and website-like tabular depictions.

Altogether, these rules and features form the rules of simplicity which will be the basis for our design guidelines described further down:

- Simple design with as few visual distractions as possible
- One-item-per-screen
- No horizontal scrolling
- No Adobe Flash

The rules of simplicity should enable a quick orientation and easy navigation in an online survey irrespective of the device used.

The following paragraphs conclude the core of what we call the 7 DemoSCOPE design guidelines:

1. The signature feature of our new survey design is tile-like buttons (tiles), which superseded the allegedly immortal radio-buttons. Over the past decade, tiles have emerged more and more in software of various companies all over the world. Just think of the tiles for apps on iPhones and smartphones based on Android OS. Furthermore, Microsoft has changed its layout to tiles in the latest versions of the Windows software. The tiles we use in our online surveys offer a large area to click on, which is particularly important for small mobile screens. The tile design is

the central, most crucial improvement when comparing the new to the previous survey design. The flat tile design is combined with a modest and steady color concept, which is based on the DemoSCOPE colors red and blue. See Figure 1 for an example; note that the screen is shown for smartphone and desktop users in order to demonstrate the usability and appearance of the tile design on different devices. Also for the following Figures 3-5, the images are shown in both smartphone and desktop modes.

2. Response scales are even, aligned and logical. We follow the considerations of Tourangeau, et al. (2004) that the leftmost or the top item in a scale is seen as the "first", meaning it is expected to represent an endpoint (e.g. "Like a lot"). Further, the listed options are expected to follow some logical order where the final answer option represents the opposite endpoint (e.g. "Dislike a lot"). It was noted by Christian and Dillman (2004) that respondents would answer more quickly and accurately with the scales visually and conceptually kept in logical order.

3. "Don't know" (DK) answer options are visually separated from the substantive answer options, as there is evidence that respondents are misled about the midpoint of a scale when there is no visual distinction. Survey takers tend to be guided by the visual rather than conceptual midpoint of a scale (Tourangeau, Couper, and Conrad, 2004). In our example in Figure 9 this is achieved by a different typography of the "Don't know" text.

4. We are confident that giving the respondent the ability to track his progress within a survey is an absolute must. In that respect online market research is not any different from any web-based endeavour, where it is simply expected to be transparent about any processes people are engaging in while they stay connected. For that reason we use a rather prominent progress bar in the top right of every screen shown. In literature, this issue still causes controversies. Couper et al. (2001) argue that the presence of a progress bar increases the motivation for completing a survey as you get less frustrated by long surveys. However, they also found no significant evidence for this hypothesis. Furthermore, Conrad et al. (2010) find that a progress bar increases the respondents' overall satisfaction with the survey. However, in Villar et al. (2013) a meta-analysis is conducted and the authors find that a permanent progress bar does not actually decrease the drop-off rate. Leaving the discussion aside, we think that it is the researcher's responsibility to offer transparency also on this front. An example of the progress bar is shown in Figure 3.

5. To ensure an engaging and brisk survey experience, we use pictograms for answer options as visual relief from the mere completion of a survey. Figure 4 shows the pictograms that can be used to obtain the most favourite activity for a day in Lucerne.

6. We intentionally deny the use of Flash for any animated or otherwise dynamic questions. The reason for it is that it can be no longer be assumed that Flash is

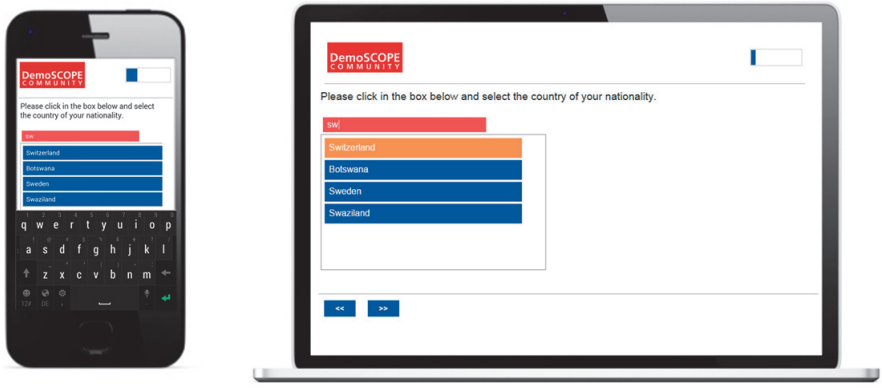

Figure 3 A text search single-choice list with progress bar in page header (smartphone and desktop version)

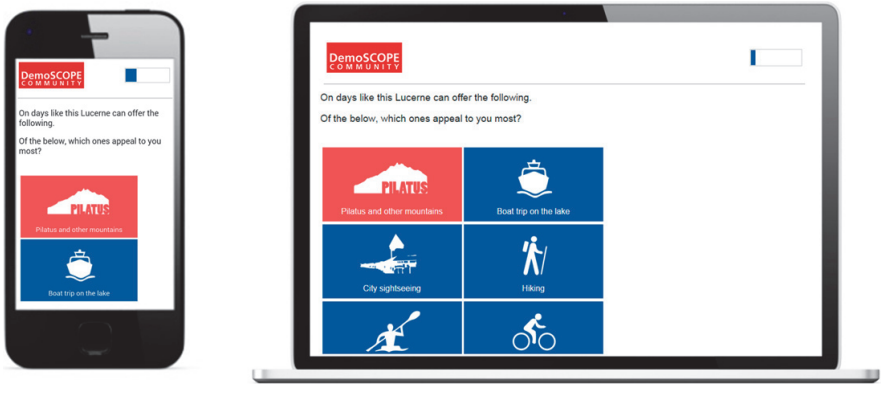

Figure 4 The use of pictograms (smartphone and desktop version)

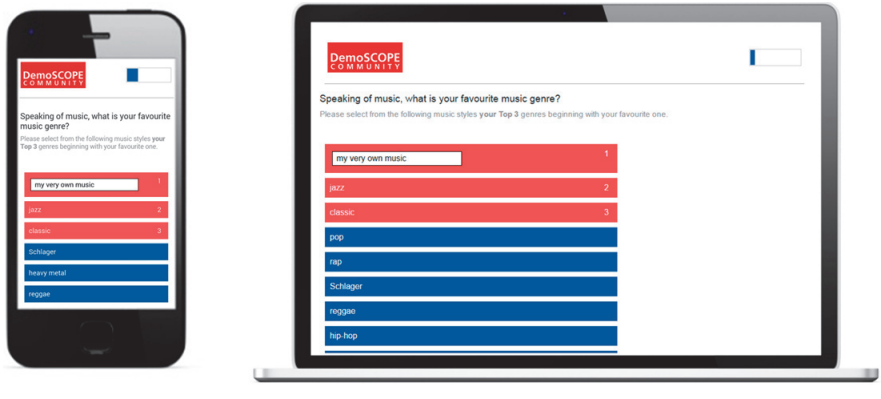

Figure 5 An interactive, yet Flash-free ranking question with built-in text fields (smartphone and desktop version)

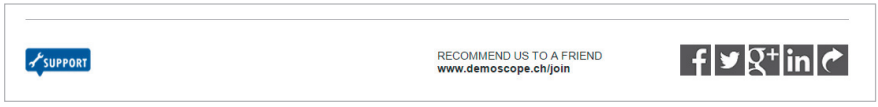

Figure 6 Options to connect and share and to contact the support team

installed on people's devices. Hence, we decided to introduce a zero tolerance policy for any Flash-animated elements in our questionnaires. See Figure 5 for a Flash-free ranking question which has animated elements but doesn't require Flash. Hence, we offer interactive questions without the necessity of Flash.

7. Furthermore, we provide our respondents possibilities of linking themselves to different social networks as well as contacting our support staff directly via a prominent support button at any stage of the questionnaire. See Figure 6 for those links.

In summary, these considerations result in the following mainly visual design guidelines:

1. Consistent flat tile design
2. Even, aligned and logical response scales
3. Visually separated "Don't know" and "No answer" options
4. Transparent progress bar
5. Pictograms as answer options or visual relief
6. No use of Flash
7. Direct opportunity at any stage to contact support team

These design guidelines were implemented at DemoSCOPE in spring 2014. Since then, almost all online studies are implemented based on the design guidelines.

4 Methods to Evaluate the Adapted Design

In the following sections we propose two ways to evaluate the design transition. Section 4.1 deals with a comparison of interview data and question timing for the old and the new design and shows differences in respondent behaviour. In Section 4.2 we present the results of a feedback study among community members concerning the old and the new design.

4.1 Analysing Interview Data and Question Timing of Panellists who Participated in an Insurance Study Using the Old and the Adapted Design

Basically, we can use two sources of data. There is the interview data itself, which can give us answers concerning a modified respondent behaviour related to the actual questions in the questionnaire. The second source of conclusions is a question timing file which contains the time needed by the respondent for each screen. Both sources can be used to check if there are any differences in the respondents' behaviour related to the adapted web design.

The first hypothesis concerning the adapted design with its characteristic tiles is that it fits more into the present state-of-the-art environment of software in use for mobile devices. The distraction of the user by an uncommon or complicated design is minimized and it is easier to keep the interest of the respondent in the actual topic of the study high. Thus, we hypothesize that the completion rate for the new design is higher than for the old design. The completion rate is defined as the rate of respondents starting the survey that fully complete all questions. I.e. the completion rate is a quantitative measure for the persistent interest in the study.

A related idea is to measure item non-response for questions which are not obligatory in order to see if the new design stimulates the respondents more to answer also difficult questions properly. Here, we consider especially the interest in pre-formulated multiple and single choice questions with given answer possibilities.

A further topic is open questions. Open questions can be very tiring for the respondent as they have to come up with own proposals or answers. The question is how the respondent can be motivated to give answers to open questions and not to skip them or even leave the study, as the question is conceived as too hard or too long. We propose a tailor-made idea to guide the respondent through an open question by introducing kinds of "motivating" elements.

Another idea is to estimate design effects related to the step from grids to the one-item-per-screen approach: Consider a grid where the single questions or statements are ordered from the top to the bottom and the answering scale is given from the left to the right. In the adapted design we have designed a one-item-per-screen approach where each question is on a single page. Our hypothesis is that the respondents tend to give the same answers when the questions are shown in a grid, as they just go from the top to the bottom clicking on the same radio button. With the one-item-per-screen approach the respondent might be animated to think of a new answer for each statement and less so-called straightlining can be found. In Lugtig and Toepoel (2015) it is discussed that straightlining can be seen as a measure of measurement error and, therefore, it is an issue to think of strategies to reduce this effect.

Conclusions concerning the respondents' behaviour can also be drawn from the question timing file. First, we can look at the total time used for the questionnaire. An idea might be that because of the clearer and easier structure of the questionnaires, the respondent is able to answer the questionnaire in the new design in less time. However, it is worth to examine the issue more detailed: In the questionnaire we have general elements that – as we claim above – clarify the structure of the questionnaire. For example, we use pictograms wherever possible, which might reduce the interview timing. Another issue is the autosubmit-function that is used whenever there is a single choice question. However, the one-item-per-screen approach may induce that the total time for a former grid increases, because several screens are shown.

We try to find empirical evidence for all these hypotheses by analysing key figures (e.g. medians, means, proportions) for the old and the new design. An integral property to guarantee the comparability of an old and an updated version of the questionnaire is that the number and order of questions in the questionnaire have not changed over several or at least two waves of the study. Furthermore, there should be no changes in the sampling process for the potential respondents.

The example chosen here is a multi-client study in the insurance market. This study is conducted quarterly with about 1,250 complete interviews. The topic is the popularity of specific insurances in Switzerland. Furthermore, questions about the use and attitude towards insurances in general are asked.

The available data we have are 9 quarterly waves in total. Four of these waves were presented completely in the old design (2013-1 to 2013-4) and four of these waves were presented completely in the new design (2014-2 to 2015-1). Wave 2014-1 cannot be used for our comparison purposes as some elements of the new design were implemented and some weren't.

As the DemoSCOPE community is a panel, we use respondents that answered the questionnaire using the old and the adapted design. Hence, we can assess key figures for the old and the new design by paring the respective interview and timing files for the same e-mail addresses. We assume that an identical e-mail address means that the questionnaire was filled by the same person.

To obtain an appropriate dataset, we first joined the interview and the timing data for each interview in the waves 2013-1 to 2015-1 (complete and incomplete interviews, excluding 2014-1). Then we merged the datasets for the old and the new design, respectively. However, it is possible that the same person (identified by the e-mail address) answered the questionnaire for the old or the new design more than once. For these cases we reduce the multiple entries to a single entry. This is done by sorting the datasets by a completion indicator and by wave. For multiple entries, we decided to choose the latest, complete interview. After obtaining datasets with single entries we have 7,666 interviews for the old design and 6,370 interviews for

the new design. In the next step, both dataset are joined by the e-mail address. By doing this, we obtained a dataset with 2,032 matching pairs of interviews.

For the analysis of the completion rate we need the complete and incomplete interviews. For the rest of the analyses we need the complete interviews only. Hence, in a second step we chose e-mail addresses with complete interviews in both designs. This results in 1,188 email addresses with paired interview and timing data for the old and the new design.

In Section 5 we will first show a descriptive analysis of several key figures. For statistical analysis of proportions, means and medians we use significance tests for paired samples.

4.2 Analysing Feedback from the DemoSCOPE Community

Another idea was to involve the community and to obtain their opinion about the adapted and the old design. A design test was implemented, where 5 screens from the old and the respective 5 updated screens from the new design were shown in rotated order. For each screen the community members had to assess the following 4 statements on a scale from 1 to 10:

1. The design enables a quick and easy orientation in the questionnaire. (Orientation)
2. I like the color composition of the questionnaire. (Color)
3. I like the design of the questionnaire in general. (Design)
4. The design of the questionnaire is user-friendly. (Usability)

Smartphone screens were shown to the smartphone users. Desktop screens were shown to the laptop and PC users.

In total, 4 * (5+5) assessments had to be made. This results in 20 pairs of scores that can be compared to each other in an analysis. In a further analysis we can sum up the evaluations for the 4 different statements for each design. This results in a total score for the 4 assessed topics for each design. We obtained answers from 112 smartphone and 200 desktop users. Community members from all socio-demographic strata were invited to conduct the study; no filters were set.

Additionally, the community was asked the following question: “Which factors are especially important for you when taking part in an online survey?” The possible choices were:

- Comprehensibility of the questions
- That a quick orientation in the questionnaire is possible
- Appealing visual design
- Interesting topics

- Varied topics
- Feedback on the results of the study, e.g. within a newsletter
- Rewards
- That smartphone or tablet can be used to take part in the study
- That the surveys are short
- That surveys are as much detailed as possible
- General user-friendliness

Each respondent had to select those 3 factors which are most important for them.

5 Results for the Insurance Study

The first proposed key measure is the completion rate. Looking at the completion rates of the 2,032 matching e-mail addresses we find a completion rate of 69.3% for the old design and a completion rate of 78.1% for the new design. A test of proportion for a paired design shows that these two proportions are significantly different on a 95% confidence level (p-value < 0.001). Hence, the completion rate for the new design is significantly different from the completion rate of the new design. We cannot prove that this difference is caused by the new design, but it is a fact that within very short time the completion rate rose by almost 10%.

Another key figure analysing completion behaviour is item non-response. Unfortunately, almost all questions within our insurance study are obligatory (however, most of them offer a "Don't know/No answer" radio button/tile). There are only two questions where we can measure "real" item non-response, i.e. where it is allowed to tick no radio button/tile. The first question analysed is: "Suppose, you want to contract a property insurance (car, furniture). Which insurance would be your first choice?" The second question analysed is: "Suppose, you want to contract a life insurance. Which insurance would be your first choice?". For both items, the percentage of item nonresponse is very low. For the first question, the item nonresponse proportion is 0.9% (old design) and 2.6% (new design), respectively. For the second question, the item nonresponse proportion is 2.0% (old design) and 3.6% (new design), respectively. If we do a statistical test for paired samples (n=1,188 in this and the following paired tests) the proportion of item nonresponse for the old and the new design is found to be significantly different in both cases (p-value < 0.001 and p-value = 0.026, respectively). Hence, the item nonresponse is very small, but significantly lower for the old design.

The next topic are open questions. The first "insurance question" after the introductory questions (language, sex, age and post code) is to write down all insurances the respondent remembers spontaneously. In the old design there are nine

boxes offered on the screen where the answers can be written. Our feeling is that many respondents are stressed by the feeling that they have to come up with nine answers that they just decide to quit the study. Looking at the interview data it was found that actually 12.6% of the 2,032 respondents left the study just at this first screen. Hence, an aim of the new design was to reduce the number of incompletes resulting from the layout of this question. The idea was to show only three empty boxes at the beginning and offer an additional empty box when a third, fourth, …, eighth insurance was written down. Using this new strategy, the quitting proportion could be reduced by 1.1% to 11.5%. However, the difference in the quitting proportion is not significantly different on a 95%-level (p-value: 0.227). A second issue is the number of insurances remembered and written down by the respondents. For the old design, the mean number of insurances is 5.51 (median: 5.0) and for the new design the mean number is 5.35 (median: 5.0). A t-test for paired samples shows that the mean number of insurances is significantly different on a 95%-level (p-value: 0.003). Hence, the quitting proportion tends to be lower for the new design. However, the mean number of entries is also lower, as less empty boxes are shown from the beginning.

The next topic considered is the so-called straightlining. The following example is chosen from the insurance study: There are 20 insurances where the respondents have to indicate, if they know the insurance

- well and have personal experience.
- well and have no personal experience.
- know only the name.
- don't know it at all.
- don't know/No answer.

In the old design, the insurances are shown in a grid, see Figure 11 (left); in the new design this is solved by a one-item-per-screen approach with so-called sliding statements (HSM format), see Figure 11 (right). There are several ideas, how straightlining in the grid/sliding statements can be evaluated: If the statements are always shown in the same order, it can be counted how often the same answer is given for two successive statements. However, in the insurance study the insurances were always shown in a different order for the old and the new design. Hence, we measure the variance within the answers for the whole grid/for all sliding statements. The higher the mean/median of the variance, the less straightlining is present. For the old design, the mean of variances is 0.99 (median: 0.99); for the new design, the mean of variances is 0.96 (median: 0.94). A paired t-test for the variances shows that they are significantly different on a 95%-level (p-value: 0.009). Hence, the straightlining cannot be reduced by the new design.

The last issues to be analysed are the timing questions. Let's first look at the introduction of the auto-submit function for single choice questions. The question

is, if the respondent can navigate quicker through the questionnaire, if the auto-submit functionality is used. We check this for a simple single choice question: "How likely is it that you will check alternative offers to your actual property insurance (car, furniture) or will even look for a new offer within the next 12 months". The mean response time for the old design is 22.4 seconds (median: 20.0 seconds) and the mean response time for the adapted design is 24.9 seconds (median: 20.0 seconds). A paired t-test shows no significant difference of the response times on 95%-level (p-value: 0.45). This means that the auto-submit button doesn't decrease the question timing substantially. However, there is still one click less the respondent has to make and this might be more comfortable for the respondent.

Another topic is the timing for questions presenting pictograms. Our examples are the questions about language (German, French) and sex (male, female). The design of the used pictograms is similar to Figure 4. The mean question timing for the old design for the language question is 27.0 seconds (median: 7.0 seconds). The mean and the median are very different in this case. A look at the data vector shows that there are high outliers. The reason might be that the language question is the first question in the questionnaire and people possibly leave it open for a while before they start the survey. Hence, for a comparison in this instance we use the median. For the new design the values are 13.3 (mean) and 5.0 (median). A Wilcoxon-test for medians in paired samples shows that the medians are significantly different (p-value < 0.001). The second pictogram we look at is the one for sex. For the old design, the mean question timing is 10.8 seconds (median: 6.0 seconds); the mean question timing for the new design is 6.0 seconds (median: 5.0 seconds). A t-test for paired samples shows no significant difference between the means (p-value: 0.138). Hence, although the finding is not significant for sex, there is a tendency that the orientation for the respondent is easier when pictograms are used.

The last issue on question timing is the grid for the evaluation of 20 insurances, we already discussed concerning straightlining. We want to know, which influence the one-item-per-screen design has on the question timing. It has to be noted that in this example the auto-submit function is implemented for each screen in the new design as the assessment of the insurances is based on a single choice selection. Hence, although we have a one-item-per-screen approach, only 1 click per page is needed. The mean question timing for the grid in the old design is 90.4 seconds (median: 73.0 seconds); the mean question timing for the new design is 91.7 seconds (median: 81.0 seconds). Again, the mean and the median are rather different which means that there are some high outliers and the question timing might not be normally distributed. Hence, we prefer to test the median rather than the mean. A Wilcoxon test for paired samples shows that the medians are significantly different on a 95% level (p-value < 0.001). Hence, the question timing for the sliding statements (new design) is significantly higher than for the grid (old design).

Table 1 Means and results of paired t-tests for the 4 scores in the old and the new design

Statement	Mean		Difference (New-Old)
	Old Design	New Design	
Orientation	31.3	39.8	8.4*
Color	27.5	36.4	8.8*
Design	28.9	37.6	8.7*
Usability	31.5	39.9	8.4*

Note. * = p< 0.001

As a last issue we look at the interview timing in total. Note however, as discussed above, that the question timing is influenced by the mix of the interview components. While pictograms in the new design need less time, the one-item-per-screen approach requires longer question timing than the old grids. The mean interview timing for the old design is 1528.3 seconds (median: 1092.0 seconds); for the new design it is 1566.0 seconds (median: 1110.0). A Wilcoxon test for paired samples shows that the total median interview length of the old and the new design are not significantly different. So, looking at the questionnaire as a whole there is no reduction in the interview time by the new design.

6 Design Evaluation Questionnaire

In the following figures 7-11 you can see the screens from the old and the new design that had to be evaluated in the design evaluation questionnaire. In a first analysis, we look at the means for each statement and screen. The mean rating for the new design is always higher than the mean rating for the old design. Looking at the differences between the old and the new design (not shown), they are especially high for screen 2 (larger than 2). For the other screens, the differences in the rating are between 1 and 2. Using a t-test for paired values, all mean differences are significant on a 95%-level (p-value < 0.001). In order to aggregate the data a little, the idea was to sum up the ratings for all screens for the 4 topics. The resulting means are shown in Table 1. For the old and the new design the mean of the Color score is lowest. In general, Color and Design are rated lower than Orientation and Usability. The differences between the two designs are between 8 and 9 for all statements (p-value < 0.001). To find out, if the used device plays a role in the rating of the 4 statements, we conducted an ANOVA for repeated measurements with

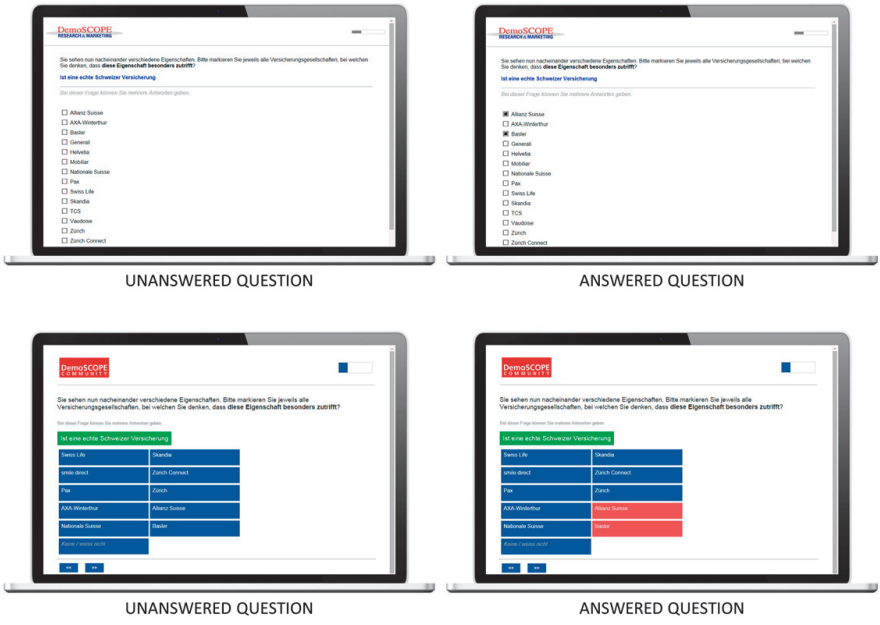

Figure 7 Screen 1 for the design test in old and new design (desktop version)

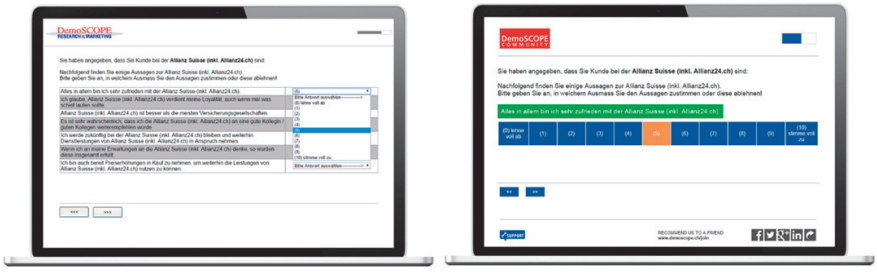

Figure 8 Screen 2 for the design test in old and new design (desktop version)

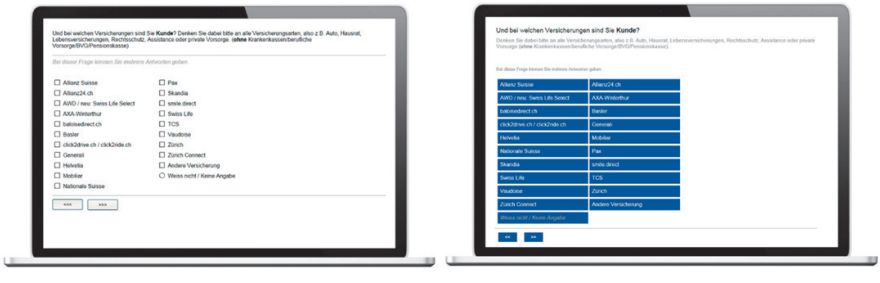

Figure 9 Screen 3 for the design test in old and new design (desktop version)

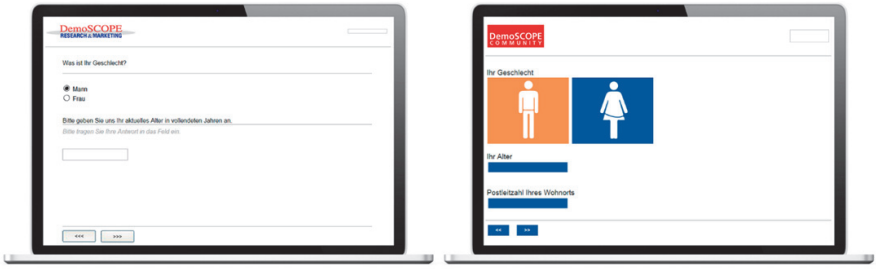

Figure 10 Screen 4 for the design test in old and new design (desktop version)

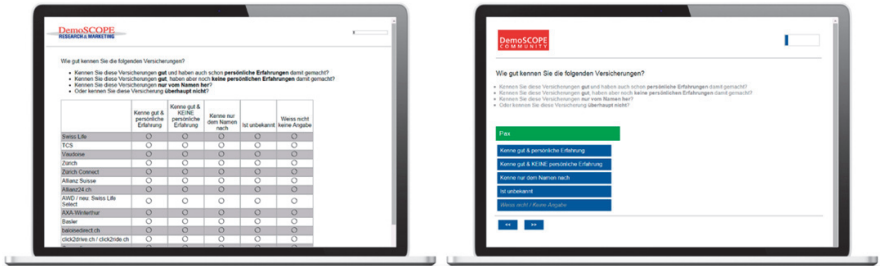

Figure 11 Screen 5 for the design test in old and new design (desktop version)

the used device as an additional factor. For the saturated model for Orientation we obtain a significant design factor ($p < 0.001$). The main effect for device is not significant (p-value 0.06). However, the interaction between design and device is significant (p-value < 0.001). This means that the increase of the rating between the old and the new design is significantly different for smartphone and desktop users. To evaluate this finding a little more detailed, you can see the estimated means for the old and the new design split by device in Figure 12. There is a higher increase in the rating for the smartphone users than for the desktop users. I.e. the benefits of the change from the old to the new design with regard to Orientation are a bit higher for smartphone users. As a result, the Orientation rating for smartphone and desktop users for the new design is almost equal, although the rating of the smartphone users was lower for the old design.

An identical analysis was made for Color. Here the design effect is found significant (p-value < 0.001), but the interaction effect for design * device and the main effect for device are not significant (p-value: 0.44 and 0.28). I.e., when it comes to judging the colors, the new design is rated better than the old one, but this preference is independent of desktop or smartphone usage. The estimated means are also shown in Figure 12.

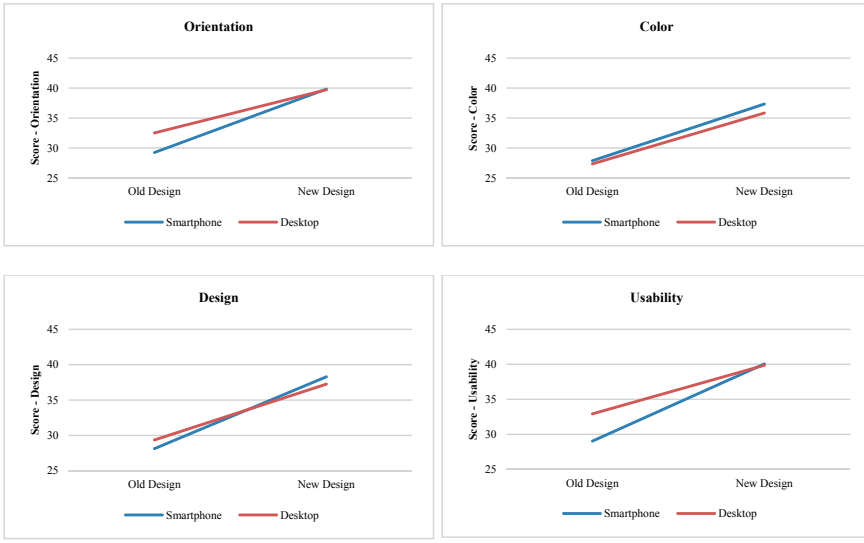

Figure 12 Estimated means for smartphone and desktop users for the old and the new design from the saturated models (two-way ANOVA with repeated measurements) for the 4 statement scores

Regarding the general evaluation of the Design, the new design is rated significantly better than the old design (factor design, p-value < 0.001). The interaction effect is also significant (p-value: 0.05), but the main effect of device is not significant (p-value: 0.92). It can be seen from Figure 12 that the estimated curves for smartphone and desktop users cross each other. Hence, the old Design is liked less by smartphone than desktop users, but the new design is rated better by smartphone users.

For the Usability the design factor is significant (p-value < 0.001). Furthermore, the interaction effect (design * device, p-value < 0.001) and the main effect for device are significant (p-value: 0.03), see Figure 12. This means again, that Usability of the old design is rated worse by smartphone than desktop users, but the new design is rated almost equal by both user groups. There is a very high increase in the Usability rating by the smartphone users. This is the desired effect, because the new design has to be equally well accepted among desktop and smartphone users and it also has to be accepted significantly better than the old design for all user-groups.

In our last analysis we asked the community about their 3 most important components of an online study. In Table 2 you see the proportions of "Yes". It can be seen, that the importance of the different components varies between smartphone and desktop users.

Table 2 Proportions of "Yes" for the importance items for the smartphone and desktop users and results of a significance test

Component		Smartphone	Desktop
1	Comprehensibility of the questions	43.8%	55.5%*
2	Quick orientation in the questionnaire	30.4%	44.5%*
3	Appealing visual design	17.9%	12.5%
4	Interesting topics	41.1%	54.0%*
5	Varied topics	18.8%	21.0%
6	Feedback on the results of the study, e.g. within a newsletter	2.7%	8.0%
7	Rewards	23.2%	15.0%
8	That smartphones and tablets can be used to take part in the study	43.8%*	6.5%
9	That the surveys are short	32.1%	27.5%
10	That surveys are as much detailed as possible	4.5%	10.0%
11	General user-friendliness	36.6%	44.5%

Note. * = $p < 0.001$

However, the Top 1 property is the same for smartphone and desktop users (Comprehensibility of the questions). For smartphone users the further most important components are that smartphone or tablets can be used and interesting topics. For the desktop users the further most important issues are interesting topics, quick orientation in the questionnaire and the general user-friendliness. A significant difference in the absolute proportions can be found for the comprehensibility of the questions, the quick orientation in the questionnaire, interesting topics and that smartphones and tablets can be used to take part in the study, see Table 2.

It would be interesting to see, if the response behaviour for the online survey components can predict, if somebody is a smartphone or desktop user. As an instrument for such an analysis we use a logistic regression model. The response variable is if somebody is a smartphone user or not (0 = desktop, 1 = smartphone). The online survey components act as independent variables. Additionally, we can add sex and age as socio-demographic, explanatory variables. To find the optimal model, we used forward model selection based on the Likelihood Ratio statistic. The resulting model contains 3 significant variables: Two of the survey components and age. You can find a summary of the optimal model in Table 3.

The estimated coefficients for components 3 and 8 are positive, which means that the odds for being a smartphone user increases if one of them is ticked as one of the 3 most important components. The most dominant item is "that smartphones

Table 3 Results for the logistic regression model for end device usage based on the online survey components, sex and age

Component	est. coeff.	std. error	Wald	*P*-value	exp (est. coeff.)
3 Appealing visual design	.849	.363	5.48	.019	2.34
8 That smartphones and tablets can be used to take part in the study	2.343	.359	42.71	<.001	10.41
Age (in years)	-.037	.010	14.53	<.001	.96

and tablets can be used to take part in the study". If this component is ticked among the 3 most important the odds that a person is a smartphone user is increases by a factor of 10.41. If somebody rates an appealing visual design as important, the respective odds increased by a factor of 2.34. Note that the older the person is, the lower is the odds for being a smartphone user. The predictive probability of this model is 76.9%. I.e. 76.9% of the respondents are categorized correctly by the model as smartphone or desktop user.

7 Conclusions

Today, a vast majority of Internet users happen to be smartphone users, too. The estimated figure is around 80%, according to GlobalWebIndex. To ignore this fact or to underestimate its importance would be a huge mistake of researchers which try to retrieve information by online surveys. We tried to face this challenge by creating our own design guidelines based on rules of simplicity and wanted to achieve some empirical evidence to evaluate our approach.

To meet this target, we used two approaches: The first approach is to evaluate paired data from members of the DemoSCOPE online panel who have participated in a specific survey before and after a design transition. The second evaluation tool is a study in which panel members were invited to rate examples of the old and the new design.

The analysis of the paired panel data shows that for the new design the completion rate is increased by almost 10%. We see this as a strong hint that a design based on the proposed design guidelines moves a little step towards an optimized online layout. However, we have to consider that this effect could also be caused, e.g., by a novelty effect based on the new setup of the DemoSCOPE online community or a changed general interest in the study topic.

Concerning the topic of question timing, an important finding is that pictograms significantly decrease the question timing. Pictograms reduce time for reading or they increase motivation due to the play-like nature of the pictograms. On the other hand, the one-item-per-screen-approach significantly increases the time needed when compared to the former grid approach. Hence, we can reinforce the findings of Couper et al. (2001) concerning the same issue. Based on our analysis, the introduction of the auto-submit function does not substantially affect the question timing. In total, there is no significant difference in the mean length of interview for both designs for our case study. Thus, the length of the entire interview is influenced by different, often contrary effects of particular interview elements. Therefore, the plain analysis of question timing might not be a useful measure: On the one hand we want participants to carefully read and answer questions, on the other hand we want to support quick and easy navigation through the questionnaire. To get a better understanding of these two conflicting demands, experiments have to be designed where time used for "thinking", and "navigating" is separated from each other.

The question timing for the new design might also be influenced by the introduction of new devices which have a quicker response time. However, the presented insurance study is stripped off from imagery and other media content and as such could not have caused longer loading times even on older devices. Furthermore, throughout the analysed surveying period we have seen no feedback from any respondent to purport this possibility.

Concerning the results from the interview data, we could not show that the new design reduces item nonresponse and straightlining. For an example of an open question in our case study, we showed that tailor-made adaptions can increase the willingness to answer open questions.

When examples of the new and the old design are shown to the DemoSCOPE community as in our rating study, the new design is rated significantly higher when it comes to Orientation, Color, Design and Usability. For Orientation and Design we see that the increase in rating before and after the design transition is significantly different for smartphone and desktop users. Based on this positive feedback from our community we think that we have proposed reasonable guidelines in the direction of an optimized online survey design.

A drawback in interpreting the results of the proposed approaches is that we cannot quantify selection effects which might nuise the result. A selection effect can take place at several stages of the analysis: First, our sample is an online panel which might not represent the true structure of the population. General population's participation in online panels is low and also probability panels are prone to selection bias with potentially large impact on results and decisions: there might be a large group of people who do not like online-research in general or mobile-device adjusted design in particular and are not part of the online panel. This could be

problematic, for example, for the rating study (Sections 4.2 and 6) as people who do not like tile design in general cannot even be invited to participate in the study. On the stage of participation, a self-selection effect might take place: Only people who like the new design participate, those who don't (and liked the old design better) do not participate.

Furthermore, for our paired panel data analysis (Sections 4.1 and 5) we use only the interviews of persons who took part in the insurance study before and after the design transition. This could be a problematic aspect when assessing the high increase in completion rate for the new design. It could be that only the supporters of the new tile design started the survey after the design transition, which results in a higher completion rate as a group of people who refuse the design transition didn't even access the study anymore and are, therefore, not part of our analysis. Furthermore, a lot of results are deduced from interviews of people that completed the insurance study before and after the design transition. This is another stage where selection bias is likely.

However, besides all the possible sources of selection bias we believe that the results of our analyses are valid and can give reasonable hints concerning the setup of an optimized survey layout in online research for multi-device usage. However, real proof for individual aspects of fluid responsive web design has to come from more controlled experiments and true random samples. From within our panel and commercial studies we cannot create controlled experiments with groups that never see the new design or with a random assignment of old and new design to sub-populations, neither can we systematically vary all the design elements mentioned above. But, what can be done is to assess benchmark measures such as the completion rate over time and continuously integrate new and research-based elements into our online design.

In this paper, we offer a handful of ideas how to go about designing online surveys in a new way. We believe it is a constant, ongoing task. This said, some parameters that we consider key in this process, will remain monitored in the day-to-day business. In a playful manner we allow to comment on the topic, user friendliness as well as the length of the survey (as shown in Figure 13). This allows us to ensure that adequate measures are taken in order to maintain a high satisfaction rate amongst our respondents with a positive impact on data quality and response rates.

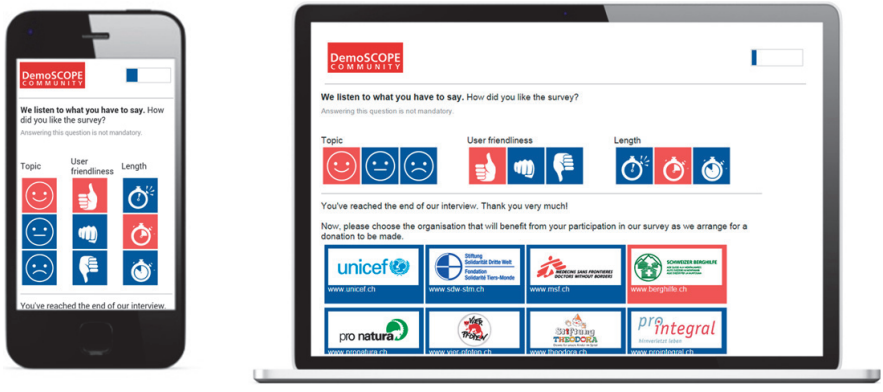

Figure 13 3 Feedback questions asked at the end of each online survey at DemoSCOPE

References

Brujine, M, & Wijnant, M. (2014). Improving Response Rates and Questionnaire Design for Mobile Web Surveys. *Public Opinion Quarterly*, 78 (4), 951-962.

Buskirk, T. D., & Andrus, C. (2012a). Smart Surveys for Smart Phones. Exploring Various Approaches for Conducting Online Mobile Surveys via Smartphones. *Survey Practice*, 5(1). http://www.surveypractice.org/index.php/SurveyPractice

Buskirk, T. D., & Andrus, C. (2012b). Online Surveys Aren't Just for Computers Anymore! Exploring Potential Mode Effects between Smartphone and Computer-Based Online Surveys. Paper presented at the annual meeting of the American Association for Public Opinion Research, Orlando, FL. http://www.amstat.org/sections/SRMS/Proceedings/y2012/Files/400244_500700.pdf

Buskirk, T. D., & Andrus, C. (2014). Making Mobile Browser Surveys Smarter: Results from a Randomized Experiment comparing Online Surveys Completed via Computer or Smartphone. *Field Methods*, 26(4), 322-342.

Christian, L. M., & Dillman, D. A. (2004). The Influence of Graphical and Symbolic Language Manipulations on Response to Self-Administered Questions. *Public Opinion Quarterly*, 68(1), 57-80.

Christian, L. M., & Dillman, D. A. (2004). The Influence of Graphical and Symbolic Language Manipulations on Response to Self-Administered Questions. Public Opinion Quarterly, 68(1), 57-80.

Christian, L. M., Dillman, D. A., & Smyth, J. D. (2007). Helping Respondents Get it Right the First Time: The Influence of Words, Symbols, and Graphics in Web Surveys. Public Opinion Quarterly, 71(1), 113-125.

Conrad, F.G., Couper, M.P., Tourangeau, R, & Peytchev, A. (2010). The Impact Of Progress Indicators on Task Completion. *Interacting with Computers*, 22, 417-427.

Couper, M. P., Traugott, M., & Lamias, M. (2001). Web survey design and administration. *Public Opinion Quarterly*, 65(2), 230-253.

Couper, M. P., Tourangeau, R., & Kenyon, K. (2004). Picture This! Exploring Visual Effects in Web Surveys. *Public Opinion Quarterly*, 68(2), 255-266.

Couper, M. P., Conrad, F. G., & Tourangeau, R. (2007). Visual Context Effects in Web Surveys. *Public Opinion Quarterly*, 71(4), 623-634.

Couper, M. P., Tourangeau, R., Conrad, F. G., & Zhang, C. (2013). The Design of Grids in Web Surveys. *Social Science Computer Review*, 31(3), 322-345.

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

Deutskens, E., De Ruyter, K., Wetzels, M., & Oosterveld, P. (2004). Response Rate and Response Quality of Internet-Based Surveys. An Experimental Study. *Marketing Letters*, 15(1), 21-36.

Dillman, D. A., Michael D. Sinclair, & Jon R. Clark. (1993). Effects of Questionnaire Length, Respondent-Friendly Design, and a Difficult Question on Response Rates for Occupant-Addressed Census Mail Surveys. *Public Opinion Quarterly*, 57(3), 289-304.

Dillman, D. A., & Christian, L. M. (2002). The Influence of Words, Symbols, Numbers, and Graphics on Answers to Self-Administered Questionnaires: Results from 18 Experimental Comparisons. Retrieved February 23, 2015 on http://www.websm.org/db/12/642/rec/.

Fuchs, M. (2005). Children and Adolescents as Respondents. Experiments on Question Order, Response Order, Scale Effects and the Effect of Numeric Values Associated with Response Options. *Journal of Official Statistics*, 21(4), 701-725.

Fuchs, M., & Busse, B. (2009). The Coverage Bias of Mobile Web Surveys Across European Countries. *International Journal of Internet Science*, 4(1), 21-33.

Gustafson, A. (2011). *Adaptive Web Design. Crafting Rich Experiences with Progressive Enhancement*. Chattanooga, Tennessee: Easy Readers.

Jenkins, C. R., & Dillman, D. A. (1997). Towards a Theory of Self-Administered Questionnaire Design. In Lyberg, L., Biemer, P., Collins, M., de Leeuw, E., Dippo, C., Schwarz, N., Trewin, D. (Eds), *Survey Measurement and Process Quality* (pp. 165-196). Wiley: New York.

Lugtig, B. & Toepoel, V (2015). The use of PCs, Smartphones, and Tablets in a Probability-Based Panel Survey: Effects on Survey Measurement Error. *Social Science Computer Review*, 1-17.

Kaplowitz, M. D., Lupi, F., Couper, M. P., & Thorp, L. (2012). The Effect of Invitation Design on Web Survey Response Rates. *Social Science Computer Review*, 30(3), 339-349.

Klausch, T., De Leeuw, E. D., Hox, J., Roberts, A., & De Jongh, A. (2012). Matrix vs. Single Question Formats in Web Surveys: Results From a Large Scale Experiment. Paper presented at the General Online Research (GOR), March 5-7 2012, Mannheim, Germany.

Knäuper, B., Schwarz, N., & Park, D. (2004). Frequency Reports Across Age Groups. *Journal of Official Statistics*, 20(1), 91-96.

Krosnick, J. A., & Alwin, D. F. (1987). An Evaluation of a Cognitive Theory of Response Order Effects in Survey Measurement. *Public Opinion Quarterly*, 51(2), 201–19.

Marcotte, E. (2010). Responsive Web Design. Retrieved March 2, 2015, from: http://alistapart.com/article/responsive-web-design/.

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

Raento, M., Oulasvirta, A., & Eagle, N. (2009). Smartphones. An Emerging Tool for Social Scientists. *Sociological Methods & Research*, 37(3), 426-454.

Revilla, M., Toninelli, D., Ochoa, C., & Loewe, G. (2014). WP 150 - Who has access to mobile devices in an online commercial panel? From: http://www.uva-aias.net/publications/show/1977

Rothwell, N. D. (1985). Laboratory and Field Response Research Studies for the 1980 Census of Population in the United States. *Journal of Official Statistics*, 1(2), 137-157.

Sanchez, M. E. (1992). Effect of Questionnaire Design on the Quality of Survey Data. *Public Opinion Quarterly*, 56, 206-217.

Schaeffer, N. C., & Presser, S. (2003). The science of asking questions. *Annual Review of Sociology*, 29, 65-88.

Schwarz, N., Fritz, S., & Mai, H. P. (1991). Assimilation and Contrast Effects in Part-Whole Question Sequences: A Conversational Logic Analysis. *Public Opinion Quarterly*, 55(1), 3-23.

Schwarz, N. (1995). What Respondents Learn from Questionnaires: The Survey Interview and the Logic of Conversation. *International Statistical Review*, 63(2), 153-68.

Shropshire, K. O., Hawdon, J. E., & Witte, J. C. (2009). Web Survey Design: Balancing Measurement, Response, and Topical Interest. *Sociological Methods & Research*, 37(3), 344-370.

Stern, M. J., Dillman, D. D., & Smyth, J. D. (2007). Visual Design, Order Effects, and Respondent Characteristics in a Self-Administered Survey. *Survey Research Methods*, 1(3), 121-138.

Stern, M. J., Bilgen, I.,& Dillman, D. D. (2014). The State of Survey Methodology: Challenges, Dilemmas, and New Frontiers in the Era of the Tailored Design. *Field Methods*, 26(3), 284-301.

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

Tourangeau, R., Couper, M. P., & Conrad, F. (2004). Spacing, Position, and Order: Interpretative Heuristics for Visual Features of Survey Questions. *Public Opinion Quarterly*, 68(3), 368-393.

Tourangeau, R., Couper, M. P., & Conrad, F. (2007). Color, Labels, and Interpretive Heuristics for Response Scales. *Public Opinion Quarterly*, 71(1), 91-112.

Vicente, P., & Reis, E. (2010). Using Questionnaire Design to Fight Nonresponse Bias in Web Surveys. *Social Science Computer Review*, 28(2), 251-266.

Villar, A., Callegaro, M., & Yang, Y. (2013). Where Am I? A Meta-Analysis of Experiments on the Effects of Progress Indicators for Web Surveys. *Social Science Computer Review*, 31(6), 744-762.

Wells, T., Bailey, J. T., & Link, M. W. (2014). Comparison of Smartphone and Online Computer Survey Administration. *Social Science Computer Review,* 32(2), 238-255.

Whitcomb, M. E., & Porter, S. R. (2004). E-Mail Contacts. A Test of Graphical Designs in Survey Research. *Social Science Computer Review*, 22(3), 370-376.

Wright, P., & Barnard, P. (1975). 'Just fill in this form' – a review for designers. *Applied Ergonomics*, 6(4), 213-220.

Wright, P., & Barnard, P. (1978). Asking multiple questions about several items: the design of matrix structures on application forms. *Applied Ergonomics*, 9(1), 7-14.