Universiteit Utrecht

Conducting a WhatsApp style smartphone survey

The influence of an instant research messaging survey style on the quality of open narrative question answers compared to a traditional web survey

Name: *Puck van Tussenbroek*

Students number*: 5968135*

Course*: Bachelor's project in Methodology and Statistics*

Supervisors*: Peter Lugtig and Vera Toepel*

Second assessor*: Lakshmi Balachandran Naiar*

Abstract

The smartphone as an optional survey device is one of the recent challenges survey research faces. An increasing number of people are filling in computer surveys on smartphones, sometimes intended and often unintended by the researchers. While smartphones have smaller screens, smaller fonts and different input options, it would lead to difficulties filling in traditional web surveys. Several studies find shorter open question answers, lower response rates and suboptimal data for instance. Smartphone-optimized surveys are being created and investigated as a possible solution. These optimized surveys usually entail larger fonts, larger input options and content that fits the smaller screen. No consensus has yet been established on whether these optimized surveys solve all existing problems. In this study, a new way of smartphone survey optimization is investigated. Not only do we look at the smartphone as a device, but more importantly as a communicational tool. We lay emphasis on open narrative questions and instant messaging in smartphone surveys. In an experimental design, we investigated whether an instant research messaging survey style positively influenced open narrative answer length and quality and the survey evaluation in comparison to a traditional web survey. Although no significant effect was found, this study contributes to the continued investigation of smartphone survey optimization.

Keywords

Smartphone surveys, smartphone optimized survey, open narrative questions, data quality, instant messenger, traditional web survey

Introduction

Smartphones have become a part of everyday life for many people. New technologies play an increasingly important role in survey research. With rapid changes in the world of digital communication and the widespread usage of new devices like smartphones and tablets, recent research has focussed on the exploration of survey designs that are compatible with these changes (see, e.g., Link et al., 2014). Early studies compared computer interviews to paper and pencil interviews (see, e.g., Wrights, Aquilino, and Supple, 1998) and cell-phone surveys were compared to landline surveys (see, e.g., Link et al., 2007). Today, the smartphone as an optional survey device is one of the main challenges methodological survey research faces.

An increasing number of people are filling in computer surveys on smartphones, sometimes intended and often unintended by the researchers (Peterson, 2013; de Bruijne & Wijnant, 2014). This development raised various methodological questions. Smartphones have smaller screen sizes, smaller keyboard sizes and different input options (Raento et al., 2009; Sweeney & Crestani, 2006). This might lead to inconvenience while filling in the computer-designed survey. Eventually, it could result in various outcome disadvantages such as lower response rates, longer completion times and shorter answers on open narrative questions (Mavletova et al., 2013).

Smartphones and computers do not only differ in design, but even more so in usage. Smartphones tend to be used in shorter sessions than computers (Budiu & Nielsen, 2015), and in a wider variety of environments. Therefore, smartphone surveys could be filled in with a lack of attention. This might lead to higher breakoff and suboptimal data (Couper & Peterson, 2017; Mavletova & Couper, 2015a). Although research on this topic has shown different results (Borger & Funke, 2015)., traditional web-surveys designed for computers are increasingly seen as inadequate for smartphones.

As a possible solution to these problems, smartphone optimized surveys are being created and investigated (see, e.g., Revilla, Toninelli & Ochoa, 2017). These optimized surveys are specifically adjusted to fit smaller screens, with larger fonts and a vertical orientation for readability. Several studies underlined the advantages of these smartphone optimized surveys (see, Horwitz, 2016). However, some studies find almost no difference between optimized and non-optimized smartphone surveys on various research dimensions (see, Revilla & Ochoa, 2016). Consensus has not been established on all aspects concerning whether smartphone optimized surveys work better than smartphone non-optimized surveys.

In the establishment of smartphone optimized surveys, a lot of emphasis has been laid on the design. The survey style has not changed. It has merely been altered to fit the

smartphone as a device, and that has shown to be quit an improvement in many areas. However, in survey research, the smartphone is not yet reckoned for the important communicational tool it has become.

Smartphones not only differ from computers by size and looks, but even more so by purpose. WhatsApp is the most commonly used communication application on smartphones (Montag et al., 2015). Online conversations via this app are every day's business. Based on the monthly active users, it is the most popular messaging app worldwide. Unlike e-mail for instance, instant messaging is a much more direct and conversational way of communicating (Brewer, 2000). Survey research could respond to this phenomenon, adapting even better not only to the smartphone as a device but as a communicational tool as well.

Regarding the popularity of instant messaging apps on smartphones, instant messaging interviews could possibly be an even better way to conduct a smartphone optimized survey. Some research has already been done on instant message interviewing, creating a conversational interview designed for sensitive topics. This has shown to be very useful in health-related research (Pearce, Thogerson-Ntoumani & Duda, 2017). Research has implicated that texting could lead to higher data quality, while they found fewer rounded numerical answers and more disclosure of sensitive information compared to voice interviews (Schober et al., 2015). Instant messaging entails a combination of the humanizing and conversational aspects with behold of anonymity. These humanizing and conversational aspects could be a successful combination to obtain high quality data.

In this study, we experimentally test whether an instant research messaging survey leads to higher data quality and user satisfaction compared to traditional web surveys. We are especially interested in open narrative question answers concerning the conversational interviewing style. Two open questions will be compared on length and quality between four conditions; instant research messenger on smartphone, traditional web survey on smartphone, instant research messenger on computer and traditional web survey on computer. We will focus on the first mentioned condition; instant research messenger on smartphone.

In the next section, I will elaborate on the main results from previous research on smartphone surveys. The method section on the experimental design will follow. The results will be presented after and finally, a conclusion and discussion will be provided.

Background

In this section, the existing literature on smartphone survey research will be outlined to show the recent context in which smartphone survey research is conducted. Then, open narrative questions and instant message survey research will be elaborated on since this will be the main focus of this study. In the end our research question and hypotheses will be provided.

Literature on smartphones in survey research

In recent smartphone survey research, broadly two ways can be distinguished in which the effects of smartphones on survey data have been studied. First, smartphone non-optimized survey research was conducted, whereby smartphone survey data was compared to computer survey data using the same survey. In some studies, the traditional computer survey was adjusted slightly to be compatible with the mobile device, but no substantial differences were made. This would be called semi-optimized. In both non- and semi-optimized smartphone survey research, the assumption was that the device differences will lead to different data (see, Couper & Peterson, 2017). A lot of smartphone problems were identified. While various research showed such smartphone disadvantages, focus shifted to the improvement of smartphone surveys. Therefore, multiple researchers started creating and investigating smartphone optimized surveys to find a solution to the existing problems. They investigated whether smartphone optimized surveys lead to better data compared to non-optimized traditional surveys filled in on a smartphone (see, Revilla, Toninelli & Ochoa, 2017). Some studies compared smartphone optimized and non-optimized surveys not only to each other, but also to traditional web survey computer data (Revilla and Ochoa, 2016). Antoun et al. (2018) analysed various studies done in the field of smartphone survey optimization and established five design heuristics for smartphone optimized surveys. However, not all problems seem yet to be solved. In the next paragraphs, some important studies in the field of smartphone survey research will be elaborated on, divided into smartphone non-optimized survey research and smartphone optimized survey research.

Smartphone non-optimized survey research: identifying the problem

Multiple disadvantages of the smartphone have been pointed out in studies comparing non-optimized smartphone surveys to traditional computer surveys. Mavletova (2013) compared the data quality between PC and non-optimized mobile surveys on five indicators; completion rates, response order effects, social desirability, non-substantive responses and length of open answers. 513 respondents were randomly assigned to either the mobile or

computer online survey. Results showed that mobile web data was associated with a lower completion rate and a shorter length of open answers. No strong differences were found on the other indicators.

It must be taken into account that this study was conducted in 2013, which is quite dated when it comes to this topic. Technology rapidly develops and smartphone use is becoming more widespread every year. In the United States for instance, the percentage of people who own a smartphone has increased from 35% to 68% between 2011 and 2015 (Anderson, 2015). This could have impact on smartphone survey research, while familiarity with the device could positively influence the outcome data (Couper & Peterson, 2017).

A recent study by Couper and Peterson (2017) took a different research approach. They conducted a multiyear survey study three years from 2012 on. In this case, the conducted surveys were 'mobile friendly', or 'partially optimized', which entailed bigger fonts and input controls on smartphones. However, all other features were similar to the PC version. It was therefore not identified as fully smartphone optimized. The researchers were especially interested in breakoff rates and completion time. Results showed that breakoff rates and completion times were significantly higher for smartphones.

Several studies mentioned important reasons for the smartphone disadvantages in surveys. A slower internet connection on the cell phone was pointed out as one of the main reason by Mavletova (2013). The fast technological development and increasing speed of internet connection resulted in negligible differences between the devices on this matter. Second, participation via mobile phone was identified as more difficult due to the smaller screen size and lack of mouse and keyboard. While a lot of questions aren't fully visible on mobile screens, users need to scroll in order to read the questions; resulting in longer completion times for instance (Couper & Peterson, 2017). The smaller font size could also lower the reading speed, causing longer completion time again. Typing on the small smartphone touch screen was identified as more difficult compared to typing on a computer keyboard. Moreover, significantly more mobile respondents indicated to have had difficulty filling in the mobile web questionnaire (Mavletova, 2013). Third, multitasking and distraction were mentioned as important influencing factors. Mavletova (2013) found that a higher rate of mobile respondents completed the survey outside of the home. This could lead to distraction, causing lower response rates. Couper and Peterson (2017) also stressed the fact that multitasking, which is done frequently on smartphones, can possibly contribute to the suboptimal results found in smartphones.

Various other studies emphasised the influence of distraction and multitasking in smartphone surveys. Mavletova and Couper (2013) found that mobile users were more likely to report completing the survey away from home or office, and more likely to report the presence of others than PC users. Similarly, Lorch and Mitchell (2014) reported that more mobile users admitted doing other things while completing the survey. Moreover, Antoun (2017) found that smartphone users were substantially more likely to report multitasking and being away from home or work while completing the survey. Because of these distracting factors, smartphone surveys would be filled in with lack of attention resulting in suboptimal data (Couper & Peterson, 2017; Mavletova & Couper, 2015a).

Multitasking and other distracting factors could especially influence open questions. As Mavletova (2013) and various other studies found (see, Revilla and Ochoa, 2016), smartphone respondents tend to fill in shorter open question answers than computer respondents. As argued by Kaikkonen (2009), mobile web users will only answer to urgent emails from their phones as briefly as possible, and prefer to use a PC with a traditional keyboard for writing longer texts. Along with the notion that mobile phones are used in various places and contexts, people would be less likely to direct their full attention to fill in an elaborated answer as they would on a PC. While respondents experience more difficulty due to the smaller screens and keyboards, it is often found that they will avoid open questions (Peytchev & Hill, 2010) or use less characters (Mavletova, 2013). However, several studies found no difference in character number between PC and smartphone respondents (Toepel & Lugtig, 2014; Buskirk & Andrus, 2014).

Various studies underlined the negative influence of smartphones on survey data quality. Some of the most important identified problems and causes are depicted in Table 1. Despite these concerns, few discouraged the use of smartphones in survey research. Therefore, research increasingly started to focus on the optimization of smartphone surveys.

Table 1

Several identified problems and its causes in smartphone survey research

Causes	Problems
Smaller screen size	Longer completion times
Smaller font size	Lower response rates
Different input options (no keyboard)	Higher breakoff rates
Distraction	Shorter open question answers
Multitasking	Suboptimal data

Smartphone optimized survey research: towards a solution

In studies comparing smartphone-optimized to non-optimized surveys, research strives to establish an ideal design heuristic. Hereby, the ultimate goal is to solve the existing smartphone survey problems. Antoun et al. (2018) conducted a review study in order to identify general principles for creating effective smartphone questionnaires. They reviewed 55 presented and published studies between 2009 and 2016 on smartphone questionnaire design features. They rated studies as relevant if they reported on a smartphone survey and empirically examined the effect of questionnaire design features on indicators of response quality. They settled on five heuristics for smartphone surveys; readability, ease of selection, visibility across the page, simplicity of design features and predictability across devices. In line with this, the optimized versions mostly entail larger fonts, content fit to width of screen, vertical orientation and usually one question per page. The heuristics are cited in Table 2 (Antoun et al., 2018, p.569) along with the description. While some of these descriptions are still quite abstract, some clear guidelines have been given.

Table 2

Five Design Heuristics for Smartphone Questionnaires

Heuristics	Description
Readability	Text should be large enough to promote easy reading.
Ease of selection	Touch targets should be large enough to tap accurately.
Visibility across the page	All content should fit the width of the screen so that it is visible without horizontal scrolling.
Simplicity of design features	Design features should be simple both for researchers to deploy and for respondents to use.
Predictability across devices	Questionnaires should function in a predictable way across different devices.

Note. Reprinted from 'Design heuristics for effective smartphone questionnaires' by Antoun, C., Katz, J., Argueta, J., & Wang, L. (2018), 36(5), 569.

Readability was concretized by determining a minimum readable font size according to smartphone industry guidelines, which is 17-18 digital pixels or a height of 4,6 mm (Androids Developer's Guide, 2016; iOS Human Interface Guidelines, 2016). When it comes to ease of selection, these same guidelines suggest that targets should be 6.7-8 mm in length and width to promote accurate touching. Visibility is guaranteed when all content fits the width of the

screen and no scrolling is needed. When it comes to design features, it should be simple for both researchers and respondents. Adding complicated features can increase technical problems and decrease understanding under respondents (Funke 2016; Mendelson, Gibson & Romano-Bergstrom, 2016). Some enhanced features may be beneficial in the long term, however, they can be problematic for one-time smartphone surveys. Concerning predictability across devices, no clear guidelines have been established. However, Antoun et al. (2018) do advice researchers to test their surveys on different devices and browsers in order to identify problems early during the establishment of their survey.

In the next paragraphs, studies will be referred to described in Table 1 and Table 2 from the article of Antoun et al. (2018). The most consistent finding is that smartphone optimized surveys improve respondent's satisfaction compared to smartphone non-optimized surveys. User satisfaction was especially high when it comes to visual appearance and user friendliness. Respondents were more likely to rate the smartphone optimized surveys as "visually appealing" (Borger & Funke, 2015) and "easy to complete" (Tharp, 2015). The findings were less consistent when it comes to completion time and breakoff. Optimization has shown to reduce completion time in five of the eight studies that reported on this variable. Three studies reported fewer breakoffs in the smartphone optimized survey but four studies reported no significant effects.

Besides satisfaction, breakoff and completion rates; effectiveness of different types of questions were studied. In this case, smartphone surveys were compared to computer surveys. Six studies found that smartphone respondents typed fewer characters than computer respondents. In contrast, five of the reviewed studies report that respondents type at least as many characters in smartphone and computer condition. On breakoff, completion rates and open question answers, no consensus has been established.

Several reasons could underlie these research discrepancies (Antoun et al., 2018). First, variations in the respondent's familiarity with the smartphone could lead to different data across studies. Second, every study held a slightly different optimization approach. This could lead to different data according to the optimization style used. Across all research studies however, three consistent features are found. Larger fonts responsive to screen size, larger response options (such as wide buttons for instance), and content that is fit to the width of screen (Antoun et al., 2018). While Antoun et al. (2018) has shown that smartphone optimized surveys can be an improvement compared to non-optimized surveys on various research dimensions, there is still no consensus on all aspects.

While the established design heuristics (Antoun et al., 2018) do not seem to solve all existing smartphone survey problems, another approach might be needed. While these design heuristics focus on the smartphone as a device, with different sizes and input options, we will examine the smartphone as a communicational tool. This is what makes the smartphone stand out from other devices, not only the size and looks but more importantly the purpose. In the following section, we will turn to new techniques that could be useful for making smartphone surveys truly optimized concerning the important communicational tool it has become: open narrative questions and instant message interviewing.

Open narrative questions

In this study, we are interested in the difference in open narrative question answers between PC and smartphones, both optimized and non-optimized for mobile completion. Although closed questions have the advantage of being easier to answer by the respondent and easier for the researcher to analyse (Revilla & Ochoa, 2015), open narrative questions can be very valuable. A distinction has been made in open questions that call for short answers, and open narrative questions that call for elaborated, rich and developed answers (Couper et al., 2011). Open narrative questions are used when the researchers want to go deeper into the respondent's thoughts. This information can be valuable to researchers, while respondents have the possibility to give their opinion in a more detailed, personal and nuanced way in comparison to closed questions.

Several open (narrative) question advantages have been pointed out by Neumann (2016). One of the advantages is the fact that they permit an unlimited number of possible answers. While no predetermined answer has to be chosen, unanticipated findings can be discovered. Open questions permit creativity, self-expression and richness of detail to the respondents. They can reveal a respondent's thinking process and frame of reference. Although open (narrative) questions are valuable, there are some downsides. While an unlimited number of possible answers is possible, researchers must deal with different degrees of detail in answers. Some answers may be irrelevant or way too exhaustive. The coding of the answers can be very difficult, as well as the statistical analysis. Besides, respondent differences in education or literacy can influence the data. Highly literate respondents may be better able to give an elaborated and well-written answer than less literate respondents. Another downside for both researchers and respondents is that open (narrative) questions take up a greater amount of respondents' time, thought and effort. Respondents might be more reluctant to answer, and

lower quality data may be found. Despite of the disadvantages, open question answers can be very valuable, also in smartphone survey research.

Revilla and Ochoa (2016) compared the answers to open narrative questions on a PC survey, a smartphone non-optimized survey and in a smartphone-optimized survey. They looked at the number of total characters and the use of abbreviations. Remarkable was that no differences were observed between optimized and non-optimized versions for smartphones. Only the speed of answer was found to be slightly longer in the smartphone non-optimized version. When comparing PC data with smartphone data, the results showed various smartphone disadvantages both in non-optimized and optimized conditions. Speed of answer was longer in smartphones, precision of answers lower and use of abbreviations higher. However, various studies as mentioned before found no differences on open narrative question length (Antoun et al., 2018)

The ambivalence between the various study results could lie in the different research approaches, and slightly different ways of survey optimization. Remarkable is the fact that answer length is often measured as indicator for quality, while quality and length might not be inseparable. In this paper, both quality and length will be taken into account in order to establish a more comprehensive understanding of the quality of open narrative questions in our different device and survey conditions. In the method section the way in which we have measured these indicators will be explained.

Instant message interviewing

Since the rise of the internet, web based synchronous (*in which question and answer follow up immediately, real-time*) and asynchronous interviews (*in which question and answer do not follow up immediately*) have been conducted in qualitative research (James & Busher, 2006). Approaches such as focus groups, conferences and chat rooms were used to create an online experience synchronous with real time (Eichorn, 2001; O'Connor and Madge, 2001). James and Busher (2006) explored the possibilities of qualitative research via asynchronous email interviewing, specifically questioning the respondent's user experience. Based on Henson's earlier study (2000), they expected a high quality interview while email exchange encourages participants to revisit their insights and to reconsider their responses. In this way, they can be as precise as possible in their answer. Both advantages and disadvantages have been pointed out in this email based study.

Email interviewing held various benefits for both researchers and participants. One of the methodological benefits pointed out was the low costs of the study compared to face-to-

face studies for instance. No travelling is needed and email is free. Another important benefit of emailing is that both the researcher and the participant can independently decide at which moment it is suitable for them to answer the email. When conducting a telephone interview, it used to be difficult for both researcher and participant to agree a on convenient time of conversation. This problem was solved by email interviewing. However, the disadvantages seemed to outweigh the advantages. The discontinuous way of interviewing posed a lot of problems. Various participants addressed forgetting what has been said previously resulting in loss of coherence and flow of thought. Some experienced difficulty in getting clarification on meaning of questions. Others mentioned to have been distracted while answering an email. Consent, confidentiality and collaboration seemed to be other contentious issues pointed out by the researchers. Besides, the lack of real life social cues could possible result in uncertainty. Overall, asynchronous email interviewing held more disadvantages than might have been expected. However, these studies were valuable in methodological research as they paved the way for investigating more online interview options.

More recently, instant messaging services, such as SMS, MSN Messenger and Skype have been recommended as online interviewing tools while it enables a synchronous method of exchange between the interviewer and the respondent (Kazmer & Xie, 2008, Opdenakker, 2006). Instant message interviewing is a conversational form of dialogue, which increases the validity of the method (Brewer, 2000). Several advantages have been found in instant messaging. In MSN interviews, emoticons can be added to an answer. Herby, the lack of real life social cues is partly overcome (Opdenakker, 2006). Besides, instant messaging solves a lot of problems that were found in discontinuous interviewing. No loss of coherence and flow of thought will occur, while it is an ongoing conversation. Questions on clarification of questions can be posed instantly and answered quickly. Hereby, synchronous and continuous interviewing has shown to be much more effective compared to discontinuous interviews.

When it comes to smartphone surveys, instant messaging is very suitable considering the way people use their smartphone on a daily basis. Not only does it fit the device, it might also provide for valuable qualitative data. Pearce, Thøgersen-Ntoumai and Duda (2014) investigated a synchronous text-based online interviewing tool with a continuity of private discussion. The study was conducted especially in the light of medical research, while anonymity and self-disclosure is very important in this field. Topics that may cause embarrassment were discussed easier via instant messaging than face-to-face or via telephone. This is because of the combination of anonymity and humanizing aspects together. On the one hand, the participant feels anonymous behind their texts. On the other hand, they do feel like

they are in contact with a human being who is interested and sincere. This combination results in more self-disclosure than face-to-face or via telephone. Compared to email, it is closer to a real conversation, while also allowing the participants to look back and reflect upon the dialogue. Not only in medical studies, but in various other fields could this be valuable for research.

Instant messaging and open narrative questions could possibly be useful in smartphone surveys. We therefore hypothesize the following. *H1: the instant research messaging survey style, when filled in on the smartphone device, will positively influence the length and quality of the open narrative question asnwers.* The second hypothesis concerns the evaluation. *H2: the instant research messaging survey style, when filled in on the smartphone device, will positively influence the evaluation.* We will now describe how a WhatsApp-style method of interviewing was used and compared to different smartphone and computer conditions.

Methods

The Method

We investigated whether an instant research messaging survey style leads to longer and higher quality open narrative question answers compared to a traditional web survey style. We also investigate whether the respondents evaluate the instant research messaging survey style with higher satisfaction. We differentiate between the two survey styles and the devices on which the survey can be filled in; smartphone and computer.

One survey was created in two different formats. The instant research messaging survey was established with https://researchmessenger.com/. The traditional survey was established in a traditional web survey format. The most important difference between the surveys is that the research messenger format uses a 'WhatsApp' style of interviewing (see, Figure 1). To give an impression, more screenshots of both survey styles are included in Appendix A.

In the research messenger, the question and answer style is similar to WhatsApp and therefore the interview looks like a WhatsApp-conversation. However, when answering a question, the next question automatically appears. This is called auto-forward (see., Arn et al., 2015). Although no real-time conversation was present, the humanizing aspects of the WhatsApp-style survey should simulate a real digital conversation (see, Appendix B). Despite of the differences between the surveys, they are kept rather similar in terms of layout. When it comes to aspects such as colour and presentation of the answers, the survey appearances are very alike. This is to make sure we measure the right differences between the survey styles.

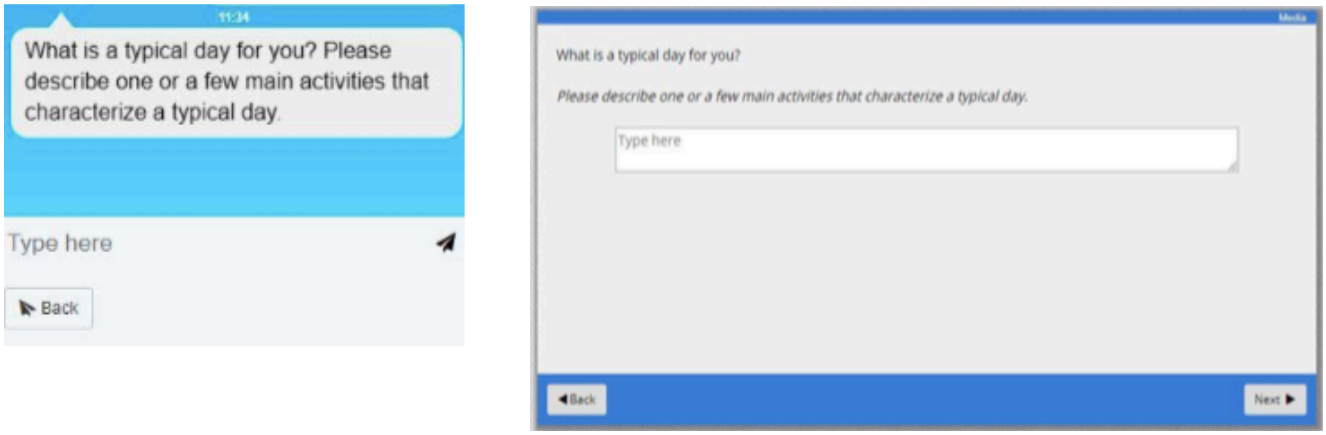


Figure 1. In these screenshots, the same question is shown in both research conditions. The instant research messenger survey design is shown on the left. On the right, the traditional web survey design is shown.

The sample

The total sample consists of 2078 American respondents, reached via Amazon Mturk (https://www.mturk.com/). On this platform, respondents receive a monetary contribution for filling in surveys. The contribution consists of 3$ per survey. The researchers did not have a lot of influence on the composition of the sample. Therefore, various demographic traits were questioned in order find out more about the background of the people who filled in the survey.

The sample consisted of 67.4% women and 32.6% men. The age of the respondents ranged between 18 and 84 years old, with an average of 35 years. When it comes to education, 46.6% of the respondents had a college degree, 40.8% went to college but did not have a 4-year degree, 12.0% finished high school and 0.6% did not finish high school. We divided education into 'high school or lower' and 'college bound or degree', ending up with 12.6% 'high school or lower' and 73.7% 'college bound or degree'. When it comes to smartphone familiarity, 98.3 percent of the respondents owned a smartphone. 47.3 percent of the respondents rated themselves 'advanced' in smartphone skills, which is the highest option on the 5-point scale. Only 0.8 percent called themselves a beginner; option 1 on the 5-point scale. Respondents self-selected their device used, ending up with 62.6% desktop, 31% smartphone and 6.3% tablet responses. We decided to exclude the tablet respondents while it did not fit in well enough with either mobile or PC, which are the devices of main focus in our study. Having excluding tablet respondents, the sample consisted of 66.9% PC and 33.1% smartphone respondents.

After selecting a device, the respondents were randomly assigned to either the traditional web survey layout or the instant research messenger. Four conditions can be

distinguished; research messenger survey on smartphone, research messenger survey on computer, traditional web survey on smartphone or traditional web survey on computer. In Table 3, this distribution is shown. Hereby, RM stands for instant research messaging survey style and TRAD stands for traditional, web-based survey style. 'PC' refers to personal computer and 'Mobile' refers to smartphone. The total number in the table differs from the total sample number mentioned before. This is because some respondents did not fill in their device used. Almost all the respondents who did not fill in their device used did not fill in the survey at all. Therefore, they were identified as 'missing', do not appear on the table and were not used for eventual analysis.

Table 3

Distribution of the research conditions.

Device	**PC**	**Mobile**	**Total**
TRAD	583	271	854
RM	583	307	890
Total	1166	578	1744

The measurements

The survey contained of 16 main questions, often supplemented with probes. It was split into 5 sections, of which the topics are media use, most important problem in the country, politics, sports and evaluation. For the purpose of this study, four questions are examined. Two open narrative question answers will be analysed on length and quality in the different conditions mentioned before. We will explain why we chose each question shortly.

Question 1: 'What is a typical day for you? Please describe one or a few main activities that characterize a typical day'. This question was chosen as it is a very useful open narrative question. On the one hand, it demands an elaborated answer and asks to describe at least one or a few activities that characterize a typical day. In this way, respondents need to think about what their day generally looks like before writing it down. On the other hand, it is not the most difficult question to answer. Generally, people would know what they do on a regular day. Therefore, we expected that almost anyone could answer this question.

The second question we looked at was a follow-up question on *'to what extent do you agree or disagree with the following statement? People like me don't have any say about what the government does'.* This was a closed question with several answer options, ranging from 'strongly agree' to 'strongly disagree', including a 'can't choose' option. The follow-up

question we looked at was; *Question 2: 'could you explain why you chose to answer [chosen answer category on question 2a]?'.* With this question, a more elaborated answer was requested. It is a valuable open narrative question, while it goes deeper into the respondent's thoughts on why he or she agrees or disagrees with the statement. Not only do we know whether the respondent agrees or disagrees, with this follow-up question we can get to know why. However, a possible disadvantage is that this question can be more difficult to answer and might require some deliberate political knowledge. We look at both question 1 and question 2, assuming Q1 being a bit less and Q2 being a bit more difficult. Hereby we try to establish a comprehensive impression of the effects of our research conditions on open narrative questions.

Lastly, two evaluative questions will be examined to see in which condition respondents had the most difficulty to answer the questions, and in which condition they were most satisfied. *Question 3: 'Was it difficult to answer the questions?'* is examined to find out in which condition the respondents experienced most difficulty. *Question 4: 'did you enjoy answering the questions?'* is examined to find out in which conditions the respondents enjoyed answering the questions most. On these questions, a five-point answer scale was given. Both answer scales ranged from (1) 'certainly not' to (5) 'certainly yes'. The questions were examined not only for the respondents who filled in the open narrative questions mentioned before, but for all who filled the evaluation. Table 4 is presented below in which is shown how many answers were given on the questions in the different research conditions.

Table 4

Distributions of the research conditions per answered question.

	TRAD		**RM**		
	PC	**Mobile**	**PC**	**Mobile**	**Total**
Q1 *'typical day'*	554	263	553	293	1663
Q2 *'explain'*	558	263	555	295	1671
Q3 *'difficult'*	547	256	531	286	1620
Q4 *'enjoy'*	547	256	531	286	1620

Operationalisation

The answer length of the two open narrative questions was defined as number of characters used. We also looked at number of themes. The number of themes was absolute, and quality was measured on a three-point scale. The value 1 was given when an answer was almost completely unclear and/or nonsense, and the value 3 was given when an answer was almost completely clear and/or very detailed. I will elaborate on this three-point scale later in this section. With four researchers, a coding scheme was established to guarantee conformity on number of themes and quality. To test the conformity, we coded the first 150 answers on question 1 and question 2 individually. We calculated how much agreement was found under the researchers using Fleiss' kappa (Fleiss & Cohen, 1973). This statistical measure calculates the degree of agreement in classification over that which would be expected by chance. When it comes to the themes for the first question, a kappa of .715 was calculated. A kappa of .531 was calculated on the quality. On question 2, a kappa of .76 was calculated for themes and a .46 kappa for quality. A sufficient kappa is .75 or higher. Only themes on question 2 measured up to that, therefore we had to improve the coding scheme to ensure more conformity, especially when it comes to measuring quality. After discussing on which aspects clarification was needed, we improved the coding scheme by making it even more precise. We conducted a new test with two researchers this time. After coding following the new coding scheme, we ran another test using Fleiss' kappa. The results showed a higher agreement rate. On the first question, a kappa of .86 was calculated for the themes. Looking at quality, a kappa of .77 was calculated. When it comes to the second question, a kappa of .91 was calculated for the themes and .75 for the quality. All of these kappa's are valid and higher than in the first test round. While the new coding scheme was found to improve conformity, we decided to use the this as the final coding scheme for all of the coding. The rest of the coding was divided between the researchers.

We will now provide with a short overview of the most important criteria; it must be noted that the full scheme is much more detailed and provides with examples to ensure as much conformity as possible. The two most important aspects of counting the number of themes are the following. (1) Every mentioned aspect should be counted. (2) When different examples are given on the same theme, the theme and all examples should be counted. Quality needed more criteria. An important question to be kept in mind wile coding was; is it clear, or would I have asked for further explanation would it have been a face-to-face interview? The tree-point value scale was defined as following.

(1) The answer is completely unclear or even nonsense, no details are elaborated on and the researcher can barely make out what the respondent meant with the given answer.
(2) The answer is partly clear and/or some details are elaborated on and the researcher can form quite an idea of what the respondent meant with the given answer.
(3) The answer is almost completely clear and/or some parts are elaborated on in a lot of detail and it is almost completely clear to the researcher what the respondent meant.

Analysis and controls

In order to answer our research question, we will analyse the influence of the survey style and device on the open narrative question answer length, quality, number of themes and on the overall evaluation of the survey. When it comes to answer length and number of themes, we looked at the effect on both questions together.

We will control for age, gender, educational level and smartphone skills. These control variables are used for several reasons. While young people are the most profound users of new technologies (van Deursen et al., 2015), age might influence the ease with which one handles a smartphone or computer device in general. Therefore, age might be a relevant control variable. When it comes to gender, several studies concluded that women used their smartphones more for social purposes and conversations than men (see, e.g., van Deursen et al., 2015). Therefore, the influence of the instant message survey style might differ in both genders, perhaps being stronger for women. Educational level might also be of influence. In the study of Revilla and Ochoa (2015), it was found that more educated people typed longer answers for instance. Smartphone skills are used as control variables, while higher smartphone skills might predict more ease using the smartphone device in general. Therefore, smartphone skills are also taken into account.

In the next section, the influence of the survey style and the device used on answer length, number of themes and overall evaluation of the survey will be analysed using multiple regression. The combination of the research messaging style and the smartphone device will be highlighted, while this is the main focus of the study. While quality was approached on a 3-point scale, the effect of the surveys and devices on the quality of the open narrative questions will be analysed using Chi2.

Results

Answer length

To measure the effect of the research messenger (RM) and the smartphone (Mobile) on the number of characters used (CharAC), a multiple regression analysis was employed. First, we combined the two open narrative questions when it comes to character number, so that one dependent variable was used. Then, we explored the data further. In figure 2 below, the distribution of the number of characters typed in the four research conditions is shown.

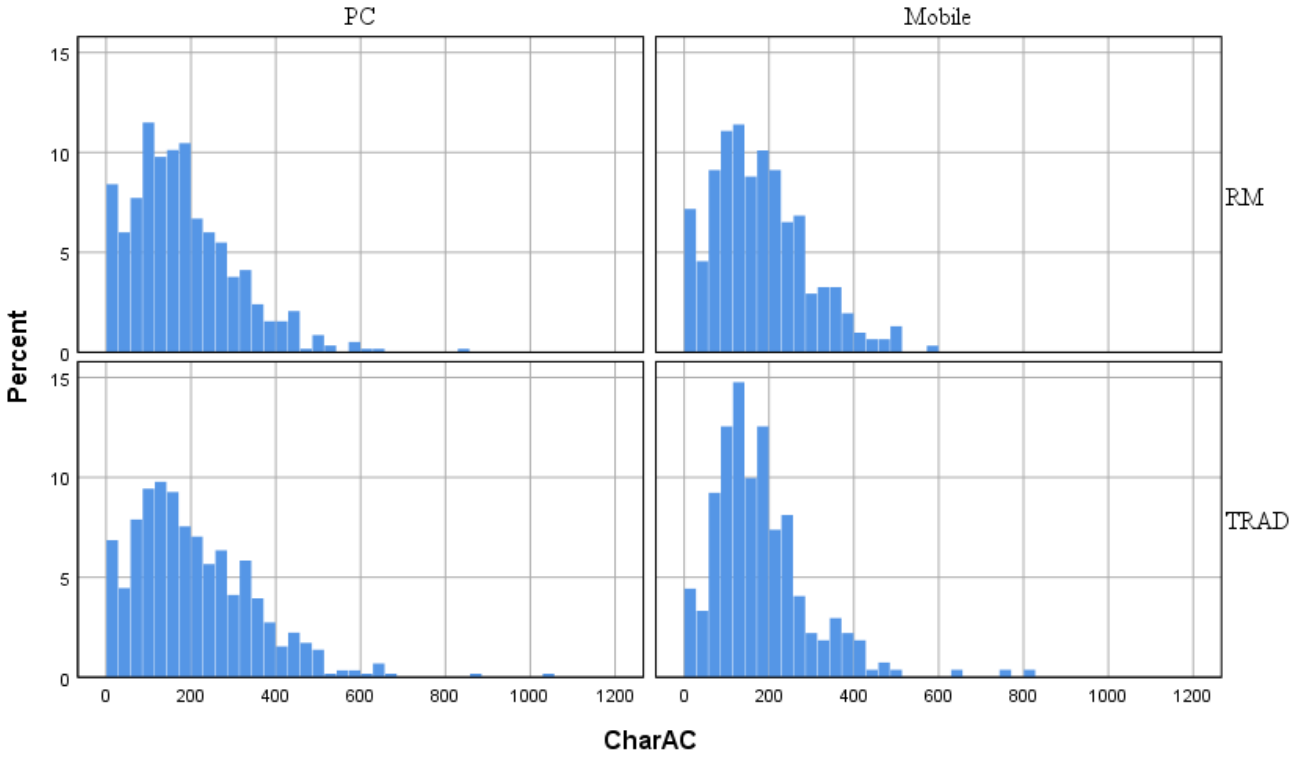


Figure 2. The percentage distribution of the number of characters (CharAC) typed in the four different research conditions, RM/TRAD and PC/Mobile.

Some assumptions were tested for multiple regression. While the assumption of normality was not met, we looked at outliers that could possibly influence the outcomes. We ran two analysis, the first one including and the second one excluding the outliers. In figure 2, possible outliers can already be distinguished, such as the character number above 1000 in the computer and traditional survey condition. We will elaborate on the outliers in the next section. The first analysis is shown in the table 5. In the first model, the research messenger survey was found to reduce the number of characters used compared to the traditional survey style (B=-13.046, t=-2.100, p=0.033). The smartphone device was also found to reduce number of characters typed compared to computer device (B=-17.784, t=-2.707, p=.007).

Table 5

	Model 1			Model 2		
	B	**SE**	**ß**	**B**	**SE**	**ß**
Constant	244.089	12.967		192.065	33.177	
RM	-13.046*	6.212	-.053	-5.674	7.572	-.023
Mobile	-17.784**	6.570	-.068	-18.569**	6.651	-.071
RM*Mobile				20.592	13.108	.048
Male				-14.932*	6.635	-.057
Age				.790*	.323	.063
Education				17.718	9.496	.047
Mobile skills				.301	3.908	.002
R^2		.008			.018	

*Dependent variable: CharAC, *=P<.05, **=P<.01, ***=P<.001. Note: N=1577*

In the second model, an interaction term and the control variables gender, age, education and mobile skills were added. The interaction term (RM*Mobile) shows the effect the combination of the research messenger and the mobile device have on characters used. While multicollinearity was an issue, we mean-centered the variables before computing this interaction variable. Model 2 was found to have higher explained variation, R^2=.018 compared to R^2=.008 in the first model. The usage of a mobile device significantly reduced the number of characters used (B=-18.569, t=-2.792, p=0.005). When it comes to gender, males filled in significantly less characters than women (B=-14.932, t=-2.250, p=0.025). A higher age was significantly related to the number of characters used (B=.790, t=2.445, p=.015).

For the second multiple regression analysis, we took a closer look at the outliers. By exploring the variable, 47 extreme cases were identified. These 47 answers had a character number of 497 or higher. Besides, 310 respondents filled in 0 characters. We therefore identified 0 and 497 or higher as outliers. After excluding these outliers, we ran the same analysis again. The explained variation was found to be lower (R^2:0.005, R^2: 0.016). Therefore, the table will not be presented. However, some remarkable differences were found in the results. The interaction term was found to be significant in the second model of the second analysis (B=21.298, t=1.870, p=.031).

Number of themes

To find out whether the survey style and device used influenced the number of themes in the given answers, we ran another multiple regression analysis. We combined the number of themes of both questions to one dependent variable. In figure 3 below, the distribution of the number of themes mentioned in the four research conditions is shown. While number of themes also had several outliers, we ran two analysis. The first one including and the second excluding the outliers. In the first model (see, Table 6), the research messenger survey was found to significantly reduce the number of themes mentioned (B=-.558, t=2.956, p=0.003). In the second model, the interaction term and control variables were added similar as mentioned before. The second model resulted in more explained variation, R^2=.020 in comparison to R^2=.008 in model 1. Males answered significantly fewer themes in comparison to women (B=-.672, t=-3.352, p=0.001). When it comes to education, an educational level significantly increased the number of themes mentioned (B=.587, t=2.095, p=.040). While the number of themes also had several outliers, we ran another multiple regression analysis after excluding these extreme values. The value 0 and 17 or higher were identified as extreme values. No major differences were found on the results; therefore, we will not elaborate on it further.

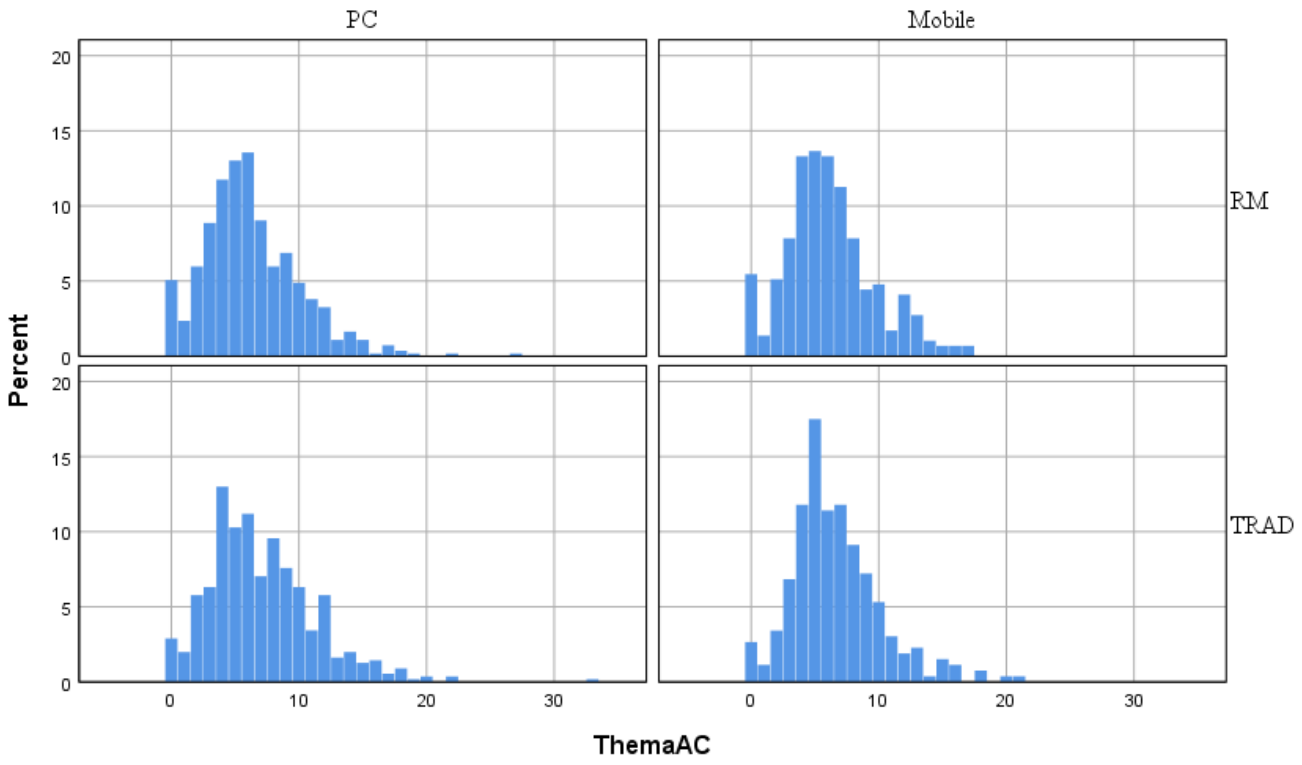


Figure 3. The percentage distribution of the number of themes (ThemaAC) mentioned in the four different research conditions, RM/TRAD and PC/Mobile.

Table 6

	Model 1			Model 2		
	B	**SE**	**ß**	**B**	**SE**	**ß**
Constant	8.203	.394		6.847	1.000	
RM	-.558**	.189	-.076	-.447	.229	-.061
Mobile	-.306	.199	-.040	-.372	.201	-.048
RM*Mobile				.283	.397	.022
Male				-.672**	.200	-.086
Age				.017	.010	.045
Education				.587*	.285	.053
Mobile skills				.115	.118	.026
R^2		.008			.020	

*Dependent variable: ThemaAC, *=P<.05, **=P<.01, ***=P<.001. Note: N=1500*

Answer quality

A Pearson's chi-square test was used to assess whether the different survey styles and devices used were related to the following answer quality. We analysed the two questions separately, in contrast to number of themes and number of characters. Quality was rated on a 3-point scale, and taking together the quality scale of the two questions would not lead to useful interpretations. 'KwaliQ1' refers to the quality values of the first question and 'KwaliQ2' refers to the quality values of the second question. For the first question 'KwaliQ1', chi-square test was not statistically significant, X2 (2, *N*= 1663) = .537, p=.765. As for the second question, 'KwaliQ2', chi-square test was not statistically significant, X2 (2, N=1671) = .106. Therefore, no further analyses will be done on the influence of the research styles and devices used on open narrative answer quality.

Survey evaluation

To find out whether the survey style and device used influenced the survey evaluation, we ran two multiple regression analyses. We looked at the two evaluative questions separately. The two questions have very different meanings and putting together the five-point scale would not lead to useful interpretations. In figure 4 and 5, the distributions of the evaluative question answers in the different research conditions are shown.

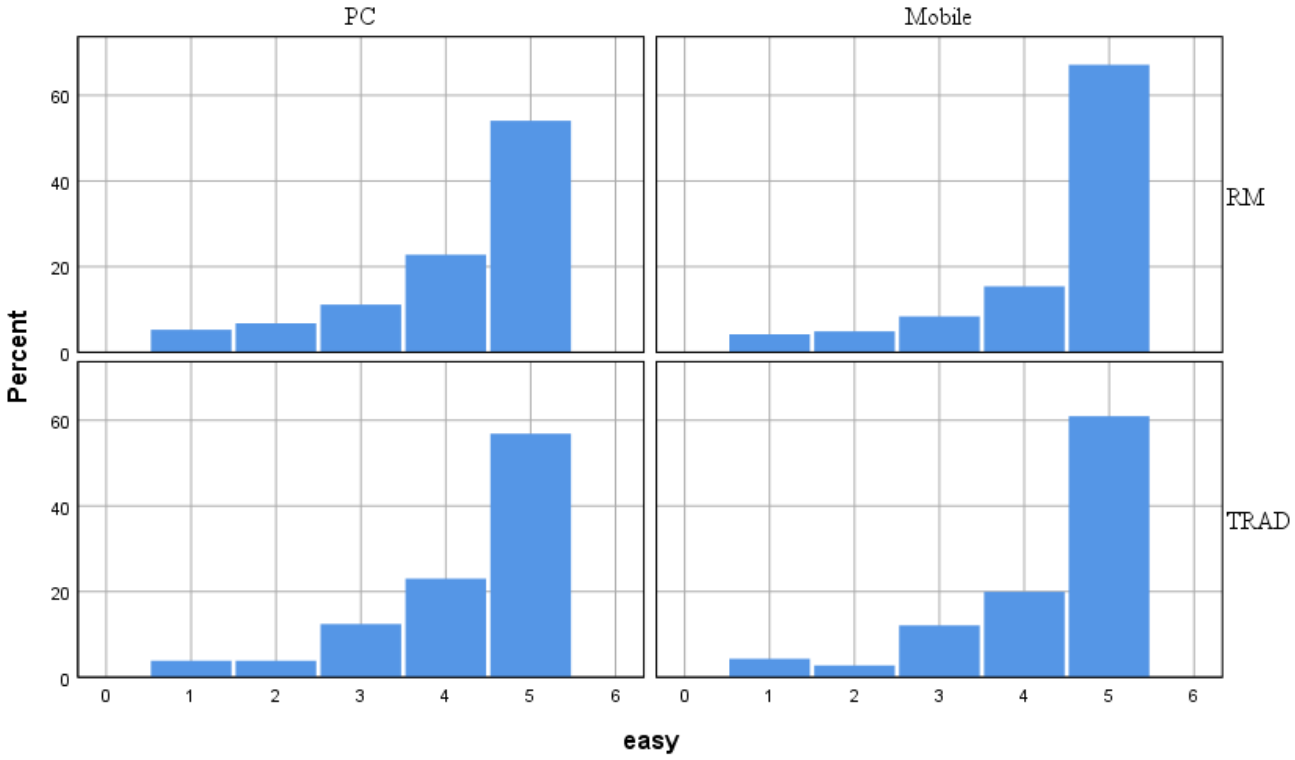


Figure 4. The percentage distribution of the first evaluative question answers (easy) in the four different research conditions, RM/TRAD and PC/Mobile.

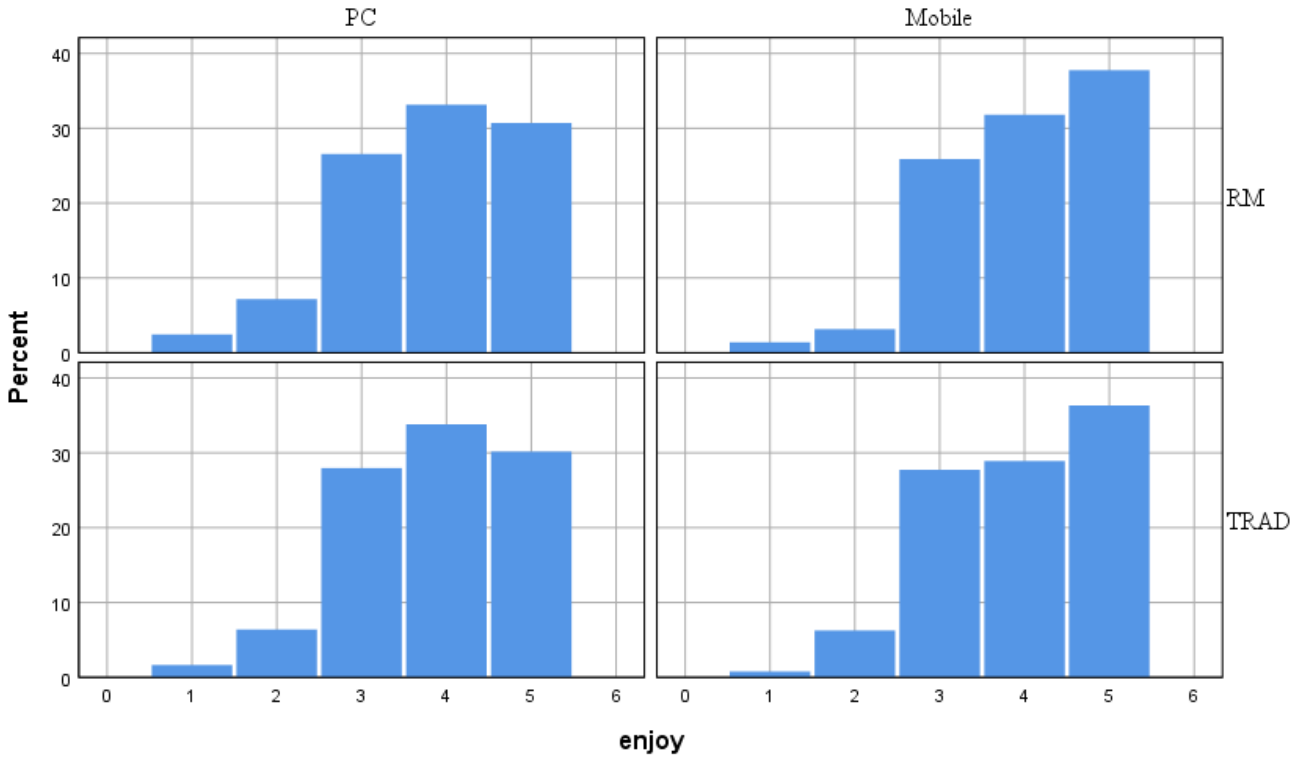


Figure 5. The percentage distribution of the second evaluative question answers (enjoy) in the four different research conditions, RM/TRAD and PC/Mobile.

In table 7, the results of multiple regression analysis on the first evaluative question are shown. We recoded the variable so that a higher score meant lower difficulty. In the second model, in which the interaction term and control variables were added, males rated the survey significantly easier than women (B=-.225, t=-3.804, p=<0.001). A higher age significantly increased the extent to which the survey was experienced as easy (B=0.006, t=2.049, p=0.041). Higher mobile skills also significantly resulted in a lower difficulty rating of the survey (B=0.110, t=3.148, p=0.002). The second model resulted in more explained variation, R^2=.021 in comparison to R^2=.003 in model 1.

Table 7

	Model 1			**Model 2**		
	B	**SE**	**ß**	**B**	**SE**	**ß**
Constant	4.166	.116		3.618	.297	
RM	-.039	.056	-.018	.012	.068	.005
Mobile	.114	.059	.049	.070	.059	.030
RM*Mobile				.144	.117	.038
Male				-.225***	.059	-.096
Age				.006*	.003	.053
Education				.083	.085	.024
Mobile skills				.110**	.035	.082
R^2		.003			.021	

*Dependent variable: easy, *=P<.05, **=P<.01, ***=P<.001. Note: N=1569*

The results of the regression analysis we ran on the second question are shown in table 8. The second question 'enjoy' was on the extent to which respondents enjoyed filling in the survey. In the first model, using a mobile device to fill in the survey was significantly related to a higher enjoyment of the survey (B=.137, t=2.604, p=0.009). In the second model, in which the interaction term and control variables were added, more explained variation was found with R^2=.030 in comparison to R^2=.004 in model 1. Using the mobile device to fill in the survey was still significantly related to a higher enjoyment of the survey (B=0.106, t=2.014, p=0.044). Age significantly increased the enjoyment of the survey (B=0.014, t=5/473, p=<0.001) and higher mobile skills were significantly related to a higher enjoyment of the survey (B=0.113, t=3.653, p=<0.001).

Table 8

	Model 1			Model 2		
	B	**SE**	**ß**	**B**	**SE**	**ß**
Constant	3.683	.103	.010	2.928	.264	
RM	.019	.050	.066	.055	.060	.028
Mobile	.137**	.053		.106*	.053	.051
RM*Mobile				.100	.104	.029
Male				.002	.053	.001
Age				.014***	.003	.140
Education				-.123	.075	-0.41
Mobile skills				.113***	.031	.095
R^2		.004			.030	

*Dependent variable: enjoy, *=P<.05, **=P<.01, ***=P<.001. Note: N= 1569*

Conclusion and discussion

In this section, the main findings will be discussed in relation to the literature and our expectations. In the end the limitations of this study will be elaborated on, along with implications for further research.

No support was found for the first hypothesis. No significant effect was found of the combination of the instant messaging research survey style and smartphone device usage on open narrative answer length or quality. In the first analysis, the mobile device was found to significantly reduce the number of characters used. This was also the case in the second model to which the controls were added. In line with the literature, this effect could still be a consequence of the greater inconvenience the smartphone device brings with filling in open answer questions (Raento et al., 2009; Sweeney & Crestani, 2006). Hardware such as the smaller screens and keyboards cannot be altered by the survey style. However, the combination of the instant messaging research survey style and smartphone device did not significantly decrease answer length or quality. Hence, no evidence is found for the opposite either. Therefore, it can still be valuable to keep on investigating this field of instant messaging smartphone survey research.

In contrast to the expectations, more characters were typed with a higher age. The effect however was small. Males filled in significantly shorter answers in comparison to females. A remarkable pattern was found while in characters used, number of themes

mentioned and the first evaluation question, men had a significant negative effect. When it comes to characters and number of themes, this effect is not unexpected. As found in other studies, women tend to use their smartphones more for social purposes and conversations than men (see, e.g., van Deursen et al., 2015). Therefore, they might be more inclined than men to type elaborate answers with various themes on questions just like they would do so in a What's-app conversation. However, further research is needed to determine whether this is the case.

The second hypothesis concerning evaluation was not supported by the results. No significant effect was found of the combination of the instant messaging research survey style and smartphone device usage on both evaluation questions. This was not as expected. In the literature, many smartphone-optimized surveys were found to increase positive survey evaluation (Antoun et al., 2018). However, no negative effect was found either. Hence, is still valuable to keep on investigating the effects of the instant research messenger on the way people evaluate the survey.

On the first question, higher mobile skills were related to a lower perceived difficulty of the survey. This is in line with the expectations. When one is more skilled using the smartphone device, one might experience less difficulty in technological matters in general, as well as in the survey. A higher age significantly increased the extent to which filling in the survey was experienced as easy. When it comes to the survey styles and devices, this might be unexpected. In most research, young people are identified as the most pronoun users of new technologies, which makes for the assumption that they will be skilled users (see, e.g., van Deursen et al., 2015). It must be taken into account that the whole survey was evaluated, including the content of the questions. Younger people could have had more difficulty filling in questions on politics for instance, which would lead to a higher perceived difficulty. This could be one of the reasons age is related to higher rating of difficulty on this evaluative question, relating to content more than to survey and device usage.

When it comes to the second evaluative question, usage of the mobile device was related to a higher rating on enjoyment of the survey. A possible explanation might be that the usage of a smartphone device in a survey is quite a new phenomenon, which could make for respondents to be interested and enjoying it more than a on a PC, while they are already used to it. Even more so because the people filling in these surveys are probably used to filling in surveys, while they are users of Amazon Mturk. When it comes to age, a higher age was significantly related to higher enjoyment of the survey. Higher mobile skills were significantly related to higher enjoyment as well. A possible explanation for this effect could

be the following. The easier the survey can be filled in; the more enjoyment can be experienced. When one is not skilled with new technology, one might experience more difficulty and therefore also experience less enjoyment while they are struggling to properly fill in the survey. However, more research need to be done in order to find out if this really is the case.

While no support was found for our first and second hypothesis, we can not conclude that the instant research messaging survey style had a positive effect on open narrative answer length, data quality and the evaluation of the survey. However, no negative effects were found either. Therefore, instant messaging smartphone surveys as a new way of looking at smartphone survey optimization still entails a lot of future research opportunities.

Limitations

There are several limitations to this study. When it comes to the sample, it is not widely representative. All respondents were approached via the same American website. Moreover, while the respondents were members of Amazon Mturk, it is highly possible that the data is influenced by their experience with filling in various surveys. Besides, they had interest in filling in the survey while they received money for it. When it comes to the device used, a higher number of PC respondents participated compared to smartphone respondents. More efforts could be made to reach even more smartphone participants. Despite these limitations, this study contained an experimental design which is valuable for measuring the difference between the two survey styles. In future research, efforts could be made to examine the questions with a more representative sample and more neutral respondents.

When it comes to the effect of an instant research messaging survey style on answer length and quality, only two questions were included in this study. In further research, more questions could be examined while two questions might be too narrow to answer our research question. Besides, not only could the instant research messaging survey have influence on open narrative questions, but also on closed questions. Future research could investigate more on the influence of the instant messaging survey on different kinds of questions.

The operationalisation of the quality in this study leaves some room for discussion. Although a lot of effort was put into the establishment of a solid and reliable coding scheme, most of the final coding was done by one researcher at a time. Reliability can therefore not be guaranteed completely. More solid coding could be done when several researchers code the same data.

Concerning the analysis, we had some problems with the assumptions of normality. While these were not met in the first analysis on number of characters for example, this might be of influence on the outcome data. This is important to keep in mind when interpreting the results of this study.

Another possible implication for further research is to look at the effects of a real research messenger in which no auto-forward is used. To strengthen the conversational and humanizing aspects of the survey, a real-time conversation between the researcher and the respondent could take place via instant message. Although this might be time-consuming and costly, it might be an interesting addition to current research in this field.

In this study, a new way of smartphone survey optimization has been investigated; the instant research messaging survey and its effect on open narrative question answers and survey evaluation. Although this study did not find the expected results, it adds a new perspective to existing smartphone survey research. It possibly paves the way for continued investigation of instant research messaging in smartphone surveys and the influence on the data quality of different types of survey questions.

References

Anderson, M. (2015). Technology Device Ownership: 2015. Pew Research Center, Washington, D.C. (October 29th). Retrieved from http://pewinternet.org

Android Developer's Guide. (2016). Material design components. Retrieved from https://material.google.com/ components/bottom-navigation.html

Antoun, C., Couper, M. P., & Conrad, F. G. (2017). Effects of mobile versus PC web on survey response quality: A crossover experiment in a probability web panel. *Public Opinion Quarterly*, *81*(S1), 280-306.

Antoun, C., Katz, J., Argueta, J., & Wang, L. (2018). Design heuristics for effective smartphone questionnaires. *Social Science Computer Review*, *36*(5), 557-574.

Arn, B., Klug, S., & Kołodziejski, J. (2015). Evaluation of an adapted design in a multi-device online panel: A DemoSCOPE case study. *Methods, Data, Analyses, 9*, 185–212.

Borger, C., & Funke, F. (2015, March). Responsive questionnaire design for higher data quality in mobile surveys. Paper presented at the GOR annual conference, Cologne, Germany.

Brewer, M. B., & Crano, W. D. (2000). Research design and issues of validity. *Handbook of research methods in social and personality psychology,* 3-16.

de Bruijne, M., & Wijnant, A. (2013a). Can mobile web surveys be taken on computers? A discussion on a multi-device survey design. *Survey Practice, 6*(4), 1–8.

Budiu, R., & Nielsen, J. (2015). User experience for mobile applications and websites. 3rd edition. Nielsen Norman Group. Retrieved from http://www.nngroup.com/reports/mobile/.

Couper, M. P., Kennedy, C., Conrad, F. G., & Tourangeau, R. (2011). Designing input fields for non-narrative open-ended responses in web surveys. *Journal of official statistics*, *27*(1), 65.

Van Deursen, A. J., Bolle, C. L., Hegner, S. M., & Kommers, P. A. (2015). Modeling habitual and addictive smartphone behavior: The role of smartphone usage types, emotional intelligence, social stress, self-regulation, age, and gender. *Computers in human behavior*, *45*, 411-420.

Eichhorn, K. (2001). Sites unseen: Ethnographic research in a textual community. *International Journal of Qualitative Studies in Education*, *14*(4), 565-578.

Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass

correlation coefficient as measures of reliability. *Educational and psychological measurement, 33*(3), 613-619.

Funke, F. (2016). A web experiment showing negative effects of slider scales compared to visual analogue scales and radio button scales. *Social Science Computer Review, 34,* 244–254.

James, N & Busher, H (2006). Credibility, authenticity and voice: Dilemmas in online interviewing. *Qualitative Research 6,* 403-420. doi: 10.1177/1468794106065010

O'Connor, H., & Madge, C. (2000). Cyber-parents and cyber-research: Exploring the Internet as a medium for research. *Centre for Labour Market Studies, University of Leicester, UK*.

Couper, M. P., & Peterson, G. J. (2017). Why do web surveys take longer on smartphones? *Social Science Computer Review, 35*, 357–377.

Herzing, J. M. E. (2019). Mobile web surveys. FORS Guide No. 01, Version 1.0. Lausanne: Swiss Centre of Expertise in the Social Sciences FORS. doi:10.24449/FG-2019-00001

Horwitz, R. (2016, May). The differential effect of mobile-friendly surveys on data quality. In *AAPOR annual conference, Austin, TX*.

iOS Human Interface Guidelines. (2016). Retrieved from https://developer.apple.com/ios/human-interfaceguidelines/overview/design-principles/

Kazmer, M & Xie, B (2008). Qualitative interviewing in internet studies: playing with the media, playing with the method. *Information, Communication & Society 11*(2), 257-278.

Link, M. W., Battaglia, M. P., Frankel, M. R., Osborn, L., & Mokdad, A. H. (2007). Reaching the US cell phone generation: comparison of cell phone survey results with an ongoing landline telephone survey. *Public Opinion Quarterly*, *71*(5), 814-839.

Link, M. W., Murphy, J., Schober, M. F., Buskirk, T. D., Hunter Childs, J., & Langer Tesfaye, C. (2014). Mobile technologies for conducting, augmenting and potentially replacing surveys: Executive summary of the AAPOR task force on emerging technologies in public opinion research. *Public Opinion Quarterly, 78*, 779–787.

Mavletova, A. (2013). Data quality in PC and mobile web surveys. *Social Science Computer Revier, 31*(6), 725-743.

Mavletova, A., & Couper, M. P. (2013). Sensitive topics in PC web and mobile web surveys: Is there a difference? *Survey Research Methods, 7*, 191–205.

Mavletova, A., & Couper, M. P. (2014). Mobile web survey design: Scrolling versus paging,

SMS versus E-mail invitations. *Journal of Survey Statistics and Methodology, 2,* 498–518.

Mendelson, J., Gibson, J. L., & Romano-Bergstrom, J. (2017). Displaying videos in web surveys: Implications for complete viewing and survey responses. *Social Science Computer Review*, *35*(5), 654-665.

Montag, C., Błaszkiewicz, K., Sariyska, R., Lachmann, B., Andone, I., Trendafilov, B., ... & Markowetz, A. (2015). Smartphone usage in the 21st century: who is active on WhatsApp?. *BMC research notes*, *8*(1), 331.

Neuman, W. L. (2016). Understanding research. Pearson.

Opdenakker, R. (2006, September). Advantages and disadvantages of four interview techniques in qualitative research. In *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* (Vol. 7, No. 4).

Pearce, G., Thøgersen-Ntoumani, C., & Duda, J. L. (2014). The development of synchronous text-based instant messaging as an online interviewing tool. *International Journal of Social Research Methodology, 17*(6), 677-692.

Peterson, G., Mechling, J., LaFrance, J., Swinehart, J., & Ham, G. (2013, March). Solving the unintentional mobile challenge. In *CASRO Online Research Conference, March*(pp. 6-8).

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

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. (2017). An experiment comparing grids and item-by-item formats in web surveys completed through PCs and smartphones. *Telematics and Informatics, 34*, 30–42.

Revilla, M., & Ochoa, C. (2016). Open narrative questions in PC and smartphones: is the device playing a role? *Quality & Quantity*, *50*(6), 2495-2513.

Sarraf, S., Brooks, J., Cole, J., & Wang, X. (2015, May). What is the impact of smartphone optimization on long surveys? Paper presented at the AAPOR annual conference, Hollywood, FL.

Sommer, J., Diedenhofen, B., & Musch, J. (2017). Not to be considered harmful: Mobile-device users do not spoil data quality in web surveys. *Social Science Computer Review*, *35*(3), 378-387

Sweeney, S., & Crestani, F. (2006). Effective search results summary size and device screen size: Is there a relationship? *Information Processing & Management, 42*, 1056–1074.

Toepoel, V., & Lugtig, P. (2014). What happens if you offer a mobile option to your web panel? Evidence from a probability-based panel of Internet users. *Social Science Computer Review*, *32*(4), 544-560.

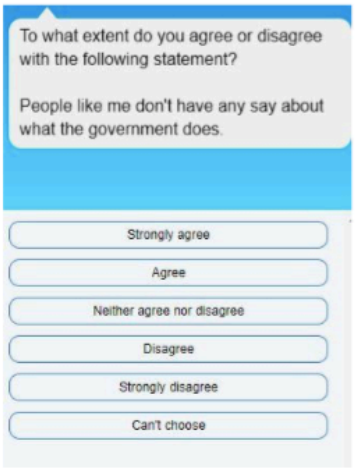
Wells, T., Bailey, J. T., & Link, M. W. (2014). Comparison of smartphone and online computer survey administration. *Social Science Computer Review*, *32*(2), 238-255.

Wright, D. L., Aquilino, W. S., & Supple, A. J. (1998). A comparison of computer-assisted and paper-and-pencil self-administered questionnaires in a survey on smoking, alcohol, and drug use. *Public Opinion Quarterly*, 331-353.

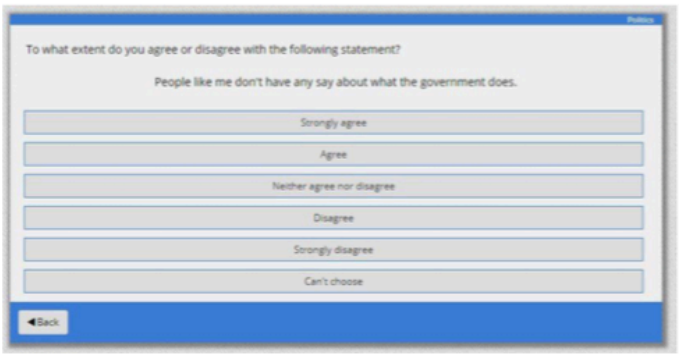
Appendix A

1

RM version



Traditional version



This screenshot shows how the question that preceded our second open narrative question looked in both survey styles.

2

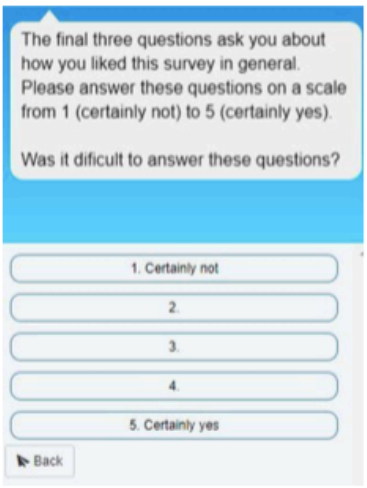
RM version

Traditional version

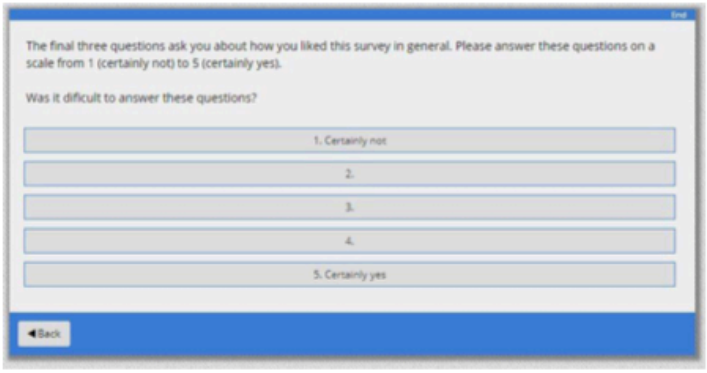
This screenshot shows how the second open narrative question we used looked in both survey styles.

3

RM version



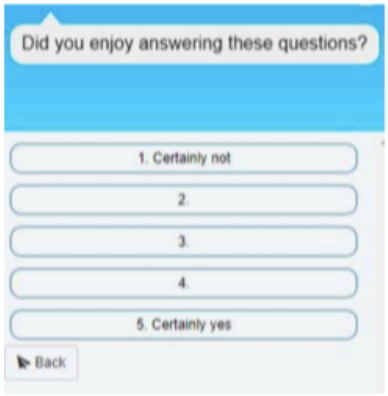
Traditional version



This screenshot shows how the first evaluative question we used looked in both survey styles.

4

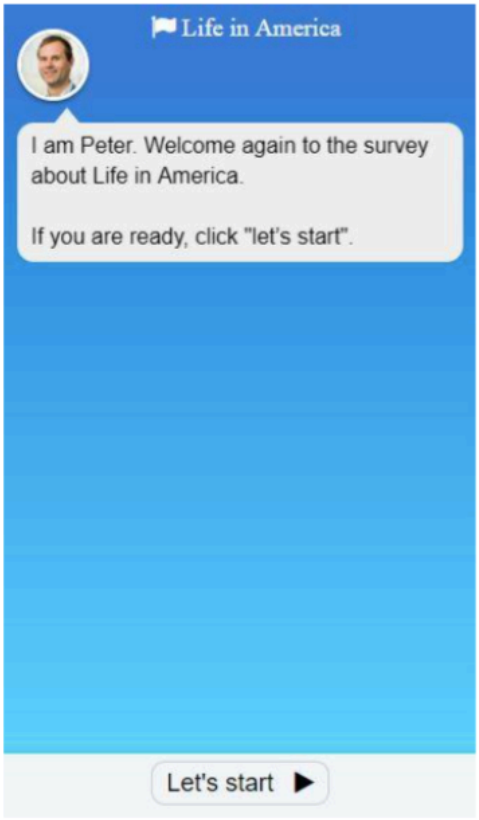
RM version



Traditional version

This screenshot shows how the second evaluative question we used looked in both survey styles.

Appendix B



This screenshot shows one of the humanizing aspects of the instant research messaging survey style.