

## ARTICLES

# Adapting surveys to the modern world: Comparing a research messenger design to a regular responsive design for online surveys

Vera Toepoel<sup>1</sup> <sup>a</sup>, Peter Lugtig<sup>1</sup>, Bella Struminskaya<sup>1</sup>, Anne Elevelt<sup>1</sup>, Marieke Haan<sup>2</sup>

<sup>1</sup> Utrecht University, <sup>2</sup> University of Groningen

Keywords: respondent burden, data quality, research messenger, mobile survey, mixed-device survey

<https://doi.org/10.29115/SP-2020-0010>

---

## Survey Practice

Vol. 13, Issue 1, 2020

---

Online surveys are increasingly completed on smartphones. There are several ways to structure online surveys so as to create an optimal experience for any screen size. For example, communicating through applications (apps) such as WhatsApp and Snapchat closely resembles natural turn-by-turn conversations between individuals. Web surveys currently mimic the design of paper questionnaires mostly, leading to a survey experience that may not be optimal when completed on smartphones. In this paper, we compare a research messenger design, which mimics a messenger app type of communication, to a responsive survey design. We investigate whether response quality is similar between the two designs and whether respondents' satisfaction with the survey is higher for either version. Our results show no differences for primacy effects, number of nonsubstantive answers, and dropout rate. The length of open-ended answers was shorter for the research messenger survey compared to the responsive design, and the overall time of completion was longer in the research messenger survey. The evaluation at the end of the survey showed no clear indication that respondents liked the research messenger survey more than the responsive design. Future research should focus on how to optimally design online mixed-device surveys in order to increase respondent satisfaction and data quality.

## 1. Introduction

Online surveys are increasingly completed on smartphones (Tourangeau et al. 2017). However, smaller screen sizes and touchscreen functionalities on mobile phones decrease ease of responding and appear to have a negative effect on response quality compared to regular desktop PCs. For example, several researchers found that responses to open-ended questions are shorter for mobile surveys compared to regular desktop PC surveys, in terms of words and characters (Mavletova 2013; Revilla and Ochoa 2016). Furthermore, mobile phones are typically used for short messaging and can be used at any time and place. Antoun, Couper and Conrad (2017) showed that respondents multitask more frequently while completing a survey on mobile phones compared to regular desktop PCs. Sendelbah et al. (2016) showed that respondents who engaged in multitasking produced higher item-nonresponse than those who

---

<sup>a</sup> v.toepoel@uu.nl

did not multitask. Despite differences in size and functionality, cognitive response processing seems to be similar between respondents on different devices (Peytchev and Hill 2010).

There are several ways to structure online surveys to create an optimal experience for any screen size. First, surveys can be designed specifically for small screens. While most survey software uses a responsive design resembling a mobile-friendly design, most surveys are still not designed “mobile-first.” Second, surveys can be designed to be more conversational or gamified. For example, communicating through applications (apps) such as WhatsApp and Snapchat closely resembles natural turn-by-turn conversations between individuals. Service chatbots mimic that communication style as well, turning a survey into a WhatsApp type of survey. We conjecture that these opportunities might improve the survey experience and lower respondent burden, and hence generate data of higher quality.

An innovative way to administer questions is via a research messenger, a WhatsApp-like survey software that communicates as one does via WhatsApp (see <http://surveyfriendly.com/demos/chatbot-store/> for an example of a survey research messenger design). Since we increasingly live in a culture of texting, using a research messenger to administer questions might be a way to increase respondent motivation compared to a responsive design. In this study, we compare a research messenger layout to a responsive design in order to investigate data quality (measured by dropout, time of completion, use of “back” button, number of nonsubstantive answers, primacy effects, and length of open-ended answers) and respondent satisfaction.

## 2. Methods

The experiment has been carried out using panel members from Amazon Mechanical Turk in the United States. In this self-selected sample, we administered four survey batches between June and August 2018. Respondents could self-select into a particular device. For batch 3 and 4, we selected panel members who indicated that their main device for accessing the Internet is a mobile phone. We use all four batches in the analyses. The estimated time to complete the survey was 5-10 minutes. Respondents were randomly assigned to the research messenger survey or the responsive design (control condition). In addition, we randomly varied the type of questions (long answer scale, short answer scale, open-ended) to investigate if the type of answer scale is related to the type of survey and respondents’ evaluation of the survey. We used four blocks of questions regarding politics, news, sports, and health, which were ordered randomly. The appendix shows screenshots of the survey in the research messenger and control condition, taken on a smartphone or regular desktop PC. Apart from the difference in layout between the research messenger and control condition, there was also a difference in how respondents navigated through the open questions in the surveys. For open questions, “next” buttons were used in the control condition, while

Table 1. Breakdown of completions by device

% (n)	Research messenger	Responsive design
Regular desktop PC	60.8 (528)	64.3 (543)
Tablet	6.3 (55)	5.8 (49)
Mobile phone	32.8 (285)	29.9 (253)
Total	100 (868)	100 (845)
Pearson chi-square	2.15	$p=.34$

Note: for 15 respondents, the device used to complete the survey could not be determined.

Table 2. Number of incomplete questionnaires (research messenger and responsive design survey)

% (n)	Research messenger	Responsive design
Complete	91.3 (871)	92.8 (857)
Incomplete	8.7 (83)	7.2 (66)
Total	100 (954)	100 (923)
Pearson chi-square	1.54	$p=.21$

respondents in the research messenger condition had to press “enter.” Autoforwarding was used for closed questions in both survey types. Overall, 1728 respondents completed the survey. We investigate dropout, time of completion, use of “back” button, number of nonsubstantive answers, primacy effects, length of open-ended answers and respondent satisfaction with the survey.

### 3. Results

As can be seen in Table 1, most respondents completed the survey on a regular desktop PC (60.8% in the research messenger condition, 64.3% in the responsive design condition). Despite our selection of panel members who indicated that their main device for accessing the Internet is a mobile phone, for two out of four batches (about 65% of our sample), the percentage of respondents that used a mobile phone was 32.8% for the research messenger condition and 29.9% for the responsive design condition. A small group of respondents completed the survey on a tablet (6.3% in the research messenger condition and 5.8% in the responsive design condition). For some respondents, the device used to access the survey could not be derived (.3% in the research messenger condition and 1.4% in the responsive design condition). There was no significant difference in the type of device used to complete the survey between the research messenger design or the responsive design.

#### 3.1 Dropout

Table 2 shows the dropout rate for both types of surveys. The dropout rate for the research messenger was 8.7% and 7.2% for the responsive design survey. The difference was not statistically significant.

Table 3. Time of completion in seconds for completes (research messenger and responsive design survey)

Seconds	Mean (standard deviation)	N
Research messenger	788 (397)	871
Responsive design	732 (443)	857
Total	760 (422)	1728
ANOVA F (1, 1727)=7.8	$p=.005$	Eta-squared=.004

Table 4. Number of times respondents went back in the survey (research messenger and responsive design survey)

	M (st.d.)	N
Research messenger	1.54 (1.72)	871
Responsive design	2.62 (3.01)	857
Total	2.08(2.50)	1728
ANOVA F (1)=85.75	$p<.001$	Eta-squared=.047

Note: we truncated one outlier to the maximum of 15.

### 3.2 Time of completion

Table 3 shows the mean time of completion for both survey types. The responsive design survey was on average 56 seconds shorter than the research messenger survey. This difference was significant ( $p = .005$ ). There was no interaction effect between survey condition and device. The total time of completion was 760 seconds or about 13 minutes.

### 3.3 Number of back actions

In both survey types, respondents could go back in order to change an answer to the previous question. The mean number of back actions in the research messenger design was 1.54 compared to 2.62 in the responsive design, as can be seen in Table 4. This difference was significant at an alpha level of .001. Since there is no function to go back in a conversation in a ‘real’ WhatsApp setting, respondents might not have been aware of this function. A Tukey post-hoc test revealed three homogeneous subsets. Respondents who completed the survey in the research messenger on a PC/tablet ( $M=1.49$ ,  $n=583$ ) and respondents who completed the survey in the research messenger on a mobile phone ( $M=1.64$ ,  $n=285$ ) had fewer back actions than respondents who completed the survey in the responsive design on a PC/tablet ( $M=2.48$ ,  $n=592$ ), who in turn had fewer back actions than respondents who completed the survey in the responsive design on a mobile phone ( $M=2.96$ ,  $n=253$ ).

### 3.4 Nonsubstantive answers

For nonsubstantive answers we looked at 10 items in the module on politics that contained the nonsubstantive answer options “don’t know” and “can’t choose”. Six out of 10 items were measured on a five-point Likert scale with a “don’t know” option. Four items were measured on a four-point scale with

Table 5. Nonsubstantive answers in 10 questions (politics module)

% (n)	Research messenger	Responsive design	Total
At least one nonsubstantive answer	9.1 (157)	7.6 (131)	16.7 (288)
No nonsubstantive answer	41.3 (714)	42.0 (726)	83.3 (1440)
Pearson chi-square	2.33	$p=.13$	

a “can’t choose” option. We investigated the number of respondents that gave at least one nonsubstantive answer in the 10 items. As can be seen in Table 5, there were no significant differences in the mean number of nonsubstantive answer options between the two surveys. In the research messenger format, 9.1% had selected at least one nonsubstantive answer in the list of 10 compared to 7.6% in the responsive design survey.

### ***3.5 Primacy effects***

We define primacy effects as selecting (one of the) first answer categories in a list. We randomly selected three questions from each module. From the sport module, we asked respondents what sports they watch on TV. We took any of the first three answers chosen as indicating primacy effects. For the two other variables, we took the first answer only as an indicator of primacy effects (due to the shorter list of answer options). In all three questions, we did not find any significant differences between the two designs, indicating that there was no difference in primacy between the two surveys, as can be seen in Table 6.

### ***3.6 Length of open-ended answers***

The survey included seven open-ended questions. We computed a variable indicating the total length of these open answers by adding the number of characters of these seven questions as a sum score. Table 7 demonstrates that the average open-answer length in the research messenger survey was 117 characters shorter than the responsive design survey.

### ***3.7 Evaluation of the survey***

At the end of the survey, we asked respondents on a Likert five-point scale if they thought the survey was difficult ( $M=1.75$ , std. dev.=1.10), enjoyable ( $M=3.88$ , std. dev.=.99), and interesting ( $M=4.04$ , std. dev.=.97). We used a linear regression analysis to predict the evaluation score from the type of survey, device used, type of answer scale that was used, gender, age, education, and a dummy for privacy (how concerned people feel about their own personal privacy on the Internet). We also tested an ordered logistic regression model to account for the ordinal measurement level of the scale. We found similar results compared to the linear regression and present results of the linear regression for the less complicated nature of the output. Table 8 shows that there are no significant differences in the type of survey.

Table 6. Primacy effects in the research messenger and responsive design surveys

% (n)	Research messenger	Responsive design	Total
How worried are you about your personal privacy when it comes to using the Internet? (list of 4)			
1 <sup>st</sup> answer	37.5 (327)	36.1 (309)	36.6 (636)
Not 1 <sup>st</sup> answer	62.5 (544)	63.9 (548)	63.2 (1092)
Pearson chi-square	.41 (n.s.)		
What sport or physical activity do you take part in most frequently (long list of 27 sports)			
1 <sup>st</sup> –3 <sup>rd</sup> answer	6.4 (56)	8.5 (73)	7.5 (129)
Later answer	93.6 (815)	91.5 (784)	92.5 (1599)
Pearson chi-square	2.73 (n.s.)		
Which of the things from the list below do you think is the most important problem facing the United States? (list of 8 most important problems)			
1 <sup>st</sup> answer	6.7 (58)	6.0 (51)	6.3 (109)
Not 1 <sup>st</sup> answer	93.3 (813)	94.0 (806)	93.7 (1619)
Pearson chi-square	.37 (n.s.)		

Table 7. Length of open answers to seven questions in the research messenger and responsive design surveys

No. of characters	Mean (standard deviation)	N
Research messenger	453 (298)	797
Responsive design	570 (362)	782
Total		1579
ANOVA F (1577,1)=49.63	$p < .001$	Eta-squared=.031

Whether respondents received the research messenger survey or the responsive design survey did not affect their evaluation of the survey. Respondents who used a mobile phone enjoyed the survey more than their regular desktop PC counterparts. An open or closed format did not have an effect on the evaluation questions. Female respondents evaluated the survey as less difficult compared to men. Young individuals (30 and younger) evaluated the survey as less enjoyable and less interesting. College graduates also thought the survey to be less enjoyable. There was no effect of respondent's level of privacy concerns on the evaluation of the survey.

The research messenger had no significant effect on the evaluation of the survey; however, manual coding of the open evaluation question showed 50 positive comments on the user interface of the survey and 7 negative

Table 8. Evaluation of the survey

	Difficult Beta	Enjoyable Beta	Interesting Beta
Messenger	.02	.01	.01
Regular desktop PC (ref.)			
Mobile phone	-.05	.06**	.04
Tablet	-.02	-.02	-.04
Open	.03	-.01	-.04
Closed short (ref.)			
Closed long	.02	.02	.01
Female	-.12***	-.00	.03
Young (<30 years)	.04	-.06**	-.05**
Education: College	.02	-.08***	-.03
Concerned privacy	.04	.03	.04
Anova	F(1727,9)=4.19***	F(1727,9)=2.89***	F(1727,9)=2.11**
R-square	.02	.02	.01

\*p<.1, \*\*p<.05, \*\*\*p<.01

Note: questions used were: 1. Was it difficult to answer these questions? 2. Did you enjoy answering the questions? 3. Was the subject interesting?

comments. All these comments came from respondents in the research messenger condition; no comments were given about the user interface in the responsive design condition.

## Conclusion

In this paper we used a research messenger design that mimics a messenger app type of communication, and compared it to a responsive design survey. We investigated whether the responses were similar in data quality and if respondents were more positive in evaluation questions. Our results show that there were no differences in relation to primacy effects, number of nonsubstantive answer options such as “don’t know” and “can’t choose,” nor dropout rate. The length of answers to open-ended questions was shorter for the research messenger survey compared to the responsive design survey. The time of completion was longer in the research messenger survey: completion time was about a minute longer. The evaluation questions at the end of the survey showed no significant differences between the research messenger survey and the responsive design survey, although comments on the open questions at the end of the survey showed many positive remarks about the style of the research messenger survey. We conclude that a messenger type of survey yields similar results compared to a responsive design survey, but it takes longer to complete and respondents provide less text in an open-ended question format. Future research should investigate the use of messenger-type surveys in order to shed light on which types of surveys can profit from this type of survey; for example, in surveys in which it is warranted to have a more conversational type of communication with survey respondents, e.g., customer-relation surveys. In addition, future research should focus on finding the right balance between the amount of interaction and data quality in research messenger surveys, since

more interaction results in longer completion times. Research should shed more light on how research messenger surveys affect respondent burden in a positive (higher motivation) or negative (longer time of completion) manner.



## REFERENCES

- Antoun, Christopher, Mick P. Couper, and Frederick G. Conrad. 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. <https://doi.org/10.1093/poq/nfw088>.
- Mavletova, Aigul. 2013. "Data Quality in PC and Mobile Web Surveys." *Social Science Computer Review* 31 (6): 725–43. <https://doi.org/10.1177/0894439313485201>.
- Peytchev, Andy, and Craig A. Hill. 2010. "Experiments in Mobile Web Survey Design: Similarities to Other Modes and Unique Considerations." *Social Science Computer Review* 28 (3): 319–35. <https://doi.org/10.1177/0894439309353037>.
- Revilla, Melanie, and Carlos Ochoa. 2016. "Open Narrative Questions in PC and Smartphones: Is the Device Playing a Role?" *Quality & Quantity* 50 (6): 2495–2513. <https://doi.org/10.1007/s11135-015-0273-2>.
- Sendelbah, Anže, Vasja Vehovar, Ana Slavec, and Andraž Petrovčič. 2016. "Investigating Respondent Multitasking in Web Surveys Using Paradata." *Computers in Human Behavior* 55 (February): 777–87. <https://doi.org/10.1016/j.chb.2015.10.028>.
- Tourangeau, Roger, Aaron Maitland, Gonzalo Rivero, Hanyu Sun, Douglas Williams, and Ting Yan. 2017. "Web Surveys by Smartphone and Tablets: Effects on Survey Responses." *Public Opinion Quarterly* 81 (4): 896–929. <https://doi.org/10.1093/poq/nfx035>.

## SUPPLEMENTARY MATERIALS

### Appendix

Download: <https://www.surveypractice.org/article/14188-adapting-surveys-to-the-modern-world-comparing-a-research-messenger-design-to-a-regular-responsive-design-for-online-surveys/attachment/39479.docx>

---