

ARTICLES

Using Buttons as Response Options in Mobile Web Surveys

Christopher Antoun¹, Elizabeth Nichols², Erica Olmsted-Hawala², Lin Wang²

¹ University of Maryland, ² U.S. Census Bureau

Keywords: checkbox, radio button, response options, measurement error, web survey instrument, web survey design, questionnaire design, mobile web survey

<https://doi.org/10.29115/SP-2020-0002>

Survey Practice

Vol. 13, Issue 1, 2020

Survey designers are starting to optimize online questionnaires for smartphones by enlarging the control elements used to accept answers (radio buttons and checkboxes) or by replacing these controls with alternatives such as stacked groups of wide buttons, but these approaches have not yet been carefully evaluated. In the lab experiment reported here, 61 older adults (age 59+) were randomly assigned to complete a survey on an iPhone with response options displayed either with conventional controls (radio buttons and checkboxes) or one of three contemporary formats: larger controls, larger controls enclosed in wide buttons, and wide buttons without any controls. The three optimization approaches were equally effective at improving respondents' tapping accuracy but were not equally rated in post-survey evaluations. Respondents tended to prefer the larger HTML-style controls over wide buttons. The results suggest that one simple way for researchers to effectively optimize their questionnaires for smartphones is to increase the size of conventional input elements without changing their basic shape or design.

Introduction

Survey designers have long used radio buttons and checkboxes to accept answers in online surveys (see e.g., Couper 2008). However, they are starting to question whether these HTML-style control elements are effective for the growing numbers of respondents using smartphones to complete surveys. Traditional form elements raise at least three concerns in mobile surveys: (1) their actual size when rendered on a smartphone screen can be quite small (e.g., 2 mm in diameter/width); (2) they are harder to select when directly touching the screen rather than using a mouse and pointed cursor (e.g., Forlines et al. 2007); and (3) they fail to provide a visual cue that the accompanying text label is also an active area for tapping (when selectable text is used). Figure 1 illustrates these issues.

Designers are addressing these issues in mobile optimized surveys in different ways, and we highlight a select few here. Some designers are increasing the size of the input tools to make them easier to select (Figure 2A). Others are adding a border around a response option or changing its background color to create something akin to a wide button that provides a visual cue that the text inside is a selectable area. This approach turns a set of response options into a stacked group of wide buttons (Figure 2B). Still others are doing away with traditional form elements altogether and using only wide buttons since the buttons presumably afford tapping on their own without the use of a radio button or checkbox (Figure 2C). One potential drawback of this format is

Figure 1. Relatively small radio buttons that provide no cues that their accompanying text is selectable. (Source: Nichols et al. 2015)

Figure 2. Different design formats (from left to right: A. larger control elements; B. larger control elements enclosed in wide buttons; and C. wide buttons without any control elements.)

that “choose-all-that-apply” questions are not visually distinct from “choose-one” questions because respondents are not provided with a visual cue (i.e., checkbox) that they can select more than one option.

Figure 3. Experimental conditions (from left to right: conventional controls, larger controls, larger controls enclosed in wide buttons, and wide buttons without any controls).

To our knowledge, there have been no direct comparisons between these different design features. The current study investigates how versions of these designs affect respondents' experience taking a mobile survey, in particular their tapping efficiency and accuracy for choose-one questions and choose-all-that-apply questions.

Methods

Experiment

Data were collected at senior centers and community centers in and around the Washington DC area in December 2016 and January 2017. Recruited participants first completed a paper questionnaire containing demographic questions. Then they were asked to use a smartphone to complete a series of tasks. For the task reported here, a test administrator handed an iPhone 5S with a preloaded survey app to the participant and instructed them to complete a survey. After completing the survey on the iPhone, the participant completed a paper questionnaire about their experience with the survey.

Participants were sequentially assigned (first participant to *condition 1*, second participant to *condition 2*, and so forth) to one of four conditions (Figure 3):

1. Conventional controls (*CC*);
2. Larger controls (*LC*);
3. Larger controls enclosed in wide buttons (*LC+WB*); and
4. Wide buttons without any controls (*WB*).

Table 1: Participant Characteristics

Characteristic	Percent
Age	
59-69 years old	56
70-80 years old	44
Sex	
Male	30
Female	70
Race	
White	76
Black	10
Asian	14

We refer to conditions 2–4 as the “optimized” formats because they were designed to mirror the formats commonly used in mobile-optimized surveys. The conventional controls were 2 mm in diameter, which was the approximate display size of the controls used in some of the U.S. Census Bureau’s online surveys at the time of the study. The larger controls were 6 mm in diameter, which was the largest size that did not have the unintended effect of reducing the number of response options that could be displayed on the screen at one time. A small area of “padding” above and below each control element was selectable; thus, the active area for touching in vertical space was approximately 4 mm for the smaller controls and 8 mm for both the larger controls and wide buttons. The active area for touching from left-to-right was the same for all the conditions by design and consisted of most of the screen (including the response option text and space to the right of the text on each row). The buttons would change color upon selection.

The mobile questionnaire used a paging design with one question per page. It contained 26 questions, most of which were adapted from the World Values Survey (2012). There were two types of questions: choose-one questions (23 items) and choose-all questions that had text instructions to select all response options that applied (3 items). Each question had seven response options, regardless of the question type.

Participants

The sample consisted of 61 adults with previous experience using smartphones. Older adults were intentionally recruited based on the assumption that an effective design for them would be at least as effective for younger adults (given their increased levels of dexterity and familiarity using smartphones, see e.g., Zhou, Rau, and Salvendy 2012).

Selected demographic characteristics are shown in Table 1.

Data from the survey app were not available for 14 participants due to technical issues; the post-survey ratings from these participants were available and analyzed.

Performance Metrics

We focus on six performance measures. Four measures were captured passively by the survey app, including:

1. *Question-level completion times*: time from page load to selection of “next” button;
2. *Misses*: number of times a participant tapped a location on the screen that was not an active selection area (and did not make a scrolling gesture);
3. *Answer changes*: number of times a participant selected a different response option after their initial selection; and
4. *Number of categories selected*: number of categories selected for each choose-all-that-apply question.

Two measures were self-reported in the post-survey questionnaire:

5. *Respondent ratings* of ease of answer selection; and
6. *Preferences* after seeing all four of the response option designs.

The question-level completion times provide a measure of efficiency, where shorter times reflect better efficiency. *Misses* and *changed answers* provide different measures of tapping accuracy; the former occurred when a selection was not recorded (error of omission) and the latter occurred when a selection was seemingly recorded by mistake (error of commission). In some cases, though, respondents may have purposefully made an initial selection and then changed it.

Data Analysis

For measures captured by the app, there were 1,222 observations in total (47 participants x 26 pages). We fit a linear mixed model (LMM) using the experimental conditions and a random effect of respondents to account for the hierarchical data structure (pages nested within respondents). Page-level timings were truncated at the 95th percentile to remove extreme values (N=1,161). A sensitivity analysis of the log-transformed timings did not lead to changes in our conclusions. *Misses* were aggregated across screens for each individual participant to produce a summary measure of the percentage of pages in which they made an errant tap. Similarly, *answer changes* were aggregated across screens for each individual participant to produce a summary measure of the percentage of pages in which they changed their answer.

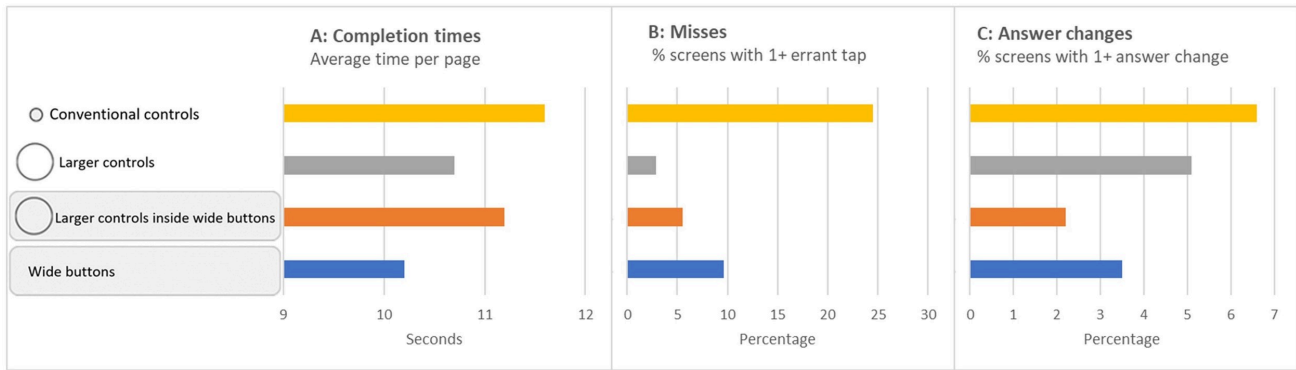


Figure 4. Performance measures across the four experimental conditions: (A) completion times; (B) misses; and (C) answer changes.

Results

Question-level completion time. We found that participants took less time per page using the optimized formats than the conventional controls (see Figure 4A). However, the differences were modest in magnitude (less than 1.5 seconds faster per page) and not statistically significant ($F = .76, p = .52$).

Misses. As shown in Figure 4B, participants made significantly fewer *misses* using the optimized formats than the conventional controls (LC vs. $CC, p < .01$; $LC+WB$ vs. $CC, p < .01$; and WB vs. $CC, p = .03$). These differences were relatively large in magnitude: participants made an errant tap on approximately 25% of pages when using the smaller controls compared to 10% of pages when using the wide buttons, 6% of pages when using the larger controls embedded in wide buttons, and 3% of pages when using the standalone larger controls. There were no significant differences with respect to *misses* among the three optimized conditions.

Answer changes. The other metric of tapping accuracy was *answer changes*. Participants did not change their answers at a high rate in any of the conditions. We found that participants changed their answers slightly less often when using the optimized formats than the conventional controls (see Figure 4C), but this did not reach statistical significance.

Number of categories selected. The use of wide buttons without any checkboxes for choose-all questions raises the possibility that respondents would not realize that they could select more than one option. This did not appear to be the case; we found that participants selected approximately about the same number of categories across the four designs ($CC: 5.1$; $LC: 5.3$; $LC+WB: 5.2$; $WB: 5.6$) ($F = .24, p = .87$).

Respondent ratings and preferences. Since the unit of analysis for the post-survey evaluations was the respondents themselves ($n=61$) rather than individual survey pages, our statistical power to detect differences was diminished compared to the earlier analyses. Still, the pattern of results suggests

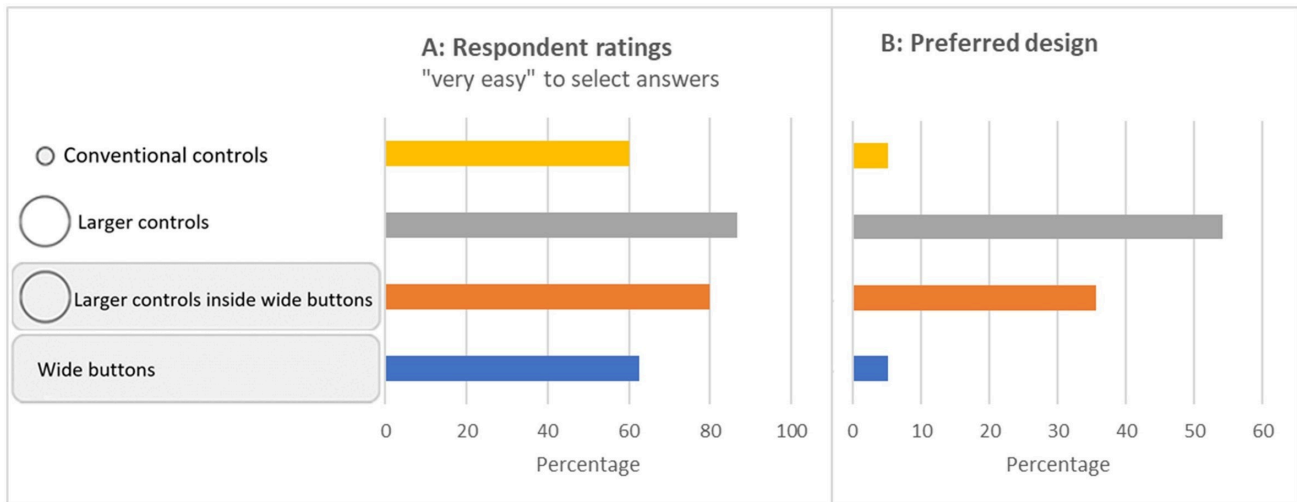


Figure 5. Respondent ratings: (A) ease of selection and (B) preferred design.

a preference for the larger HTML-style controls, either on their own or enclosed in wide buttons (Figure 5). More participants who used one of these formats reported that it was “very easy” to select their answers (87% and 80%, respectively) than those who used the plain buttons (63%) or conventional controls (60%), though the differences were nonsignificant ($\chi^2(3) = 3.9, p = .28$). Similarly, after participants were shown all of the formats, most of them reported a preference for the larger controls (54%) or larger controls embedded in wide buttons (36%) over the plain buttons (5%) or conventional controls (5%). The two most preferred formats did not differ from one another ($p = .13$) and did differ reliably from each of the other two designs ($p < .01$ for each pairwise comparison).

Discussion

This study compared conventionally-sized control elements (radio buttons and checkboxes) to three alternatives for accepting answers in mobile-optimized surveys. One format involved a simple change in target size (larger controls); another involved drawing a border around each response option (larger controls enclosed in wide buttons); and the final format omitted the control element altogether (plain wide buttons).

Our results suggest that all three of these approaches are effective at improving respondents’ tapping accuracy compared to smaller radio buttons and checkboxes. We found that the optimized formats substantially reduced the number of pages on which respondents made an errant tap (from 25% of pages when using conventional controls to 3–10% of pages when using the optimized conditions). This result is consistent with the proposition that the level of accuracy in which respondents can select a response option is a function of its size, other things being equal. The larger the touch target (up to a certain threshold), the easier it will be to select (see e.g., Wang et al. 2018). This axiom is important because tapping errors are not innocuous—they likely

result in increased respondent burden if the errors are noticed and corrected by respondents and measurement error if the errors are not corrected. Input tools that are large enough to promote easy selection are thus a key feature of effective mobile-optimized surveys.

Although wide buttons seem to be increasingly used in mobile surveys (either because buttons are perceived as stylish or because they are meant to provide visual cues that the response option text is selectable), we do not see evidence that wide buttons outperform relatively large HTML-style controls with respect to tapping accuracy or respondent ratings. Our interpretation of this result is that buttons are unnecessary because the text labels themselves afford touching. Indeed, during the experiment, some respondents could be seen selecting text labels even when they were not enclosed in a button (perhaps because selectable text is a ubiquitous feature in online forms). But these observations were not recorded in a systematic way that could be used to formally test our interpretation.

Our sample was comprised of older adults; thus, our results may not necessarily generalize to samples of younger adults. Our sample size was also not sufficiently large to detect small differences between experimental conditions. We also tested only a subset of possible mobile-friendly response formats, and each one reflected one of many possible implementations. Future research would be needed to implement these formats in different ways (e.g., buttons with different shapes, sizes, spacing, and feedback upon selection of an answer) to determine the impact of particular features or combinations of features.

Nonetheless, our findings lead to a clear interpretation. Mobile-optimized input tools can have positive effects on survey usability, and some level of optimization can be achieved by simply increasing the size of conventional input elements without changing their basic shape or design.

Acknowledgements

This study was supported by the U.S. Census Bureau Innovation and Operational Efficiency Program. The authors thank Russell Sanders for managing the project; MetroStar for developing the research software; Sabin Lakhe and Lawrence Malakhoff for technical support; Kevin Younes for recruiting participants; and Brian Falcone and Ivonne Figueroa for informative comments on the study design.

Disclaimer

This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed on statistical issues are those of the authors and not necessarily those of the U.S. Census Bureau. The Census Bureau's Disclosure Review Board and Disclosure

Avoidance Officers have reviewed this product for unauthorized disclosure of confidential information and have approved the disclosure avoidance practices applied to this release. CBDRB-FY19-CED001-B0017

Lead Author's Contact Information

1218 Lefrak Hall, 7251 Preinkert Dr., College Park, MD 20742

Phone: 301-405-0932

E-mail: antoun@umd.edu

REFERENCES

- Couper, Mick P. 2008. *Designing Effective Web Surveys*. New York: Cambridge University Press. <https://doi.org/10.1017/cbo9780511499371>.
- Forlines, Clifton, Daniel Wigdor, Chia Shen, and Ravin Balakrishnan. 2007. "Direct-Touch vs. Mouse Input for Tabletop Displays." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '07*. New York: ACM Press. <https://doi.org/10.1145/1240624.1240726>.
- Nichols, E.M., E. Olmsted-Hawala, R. Horwitz, and M. Bentley. 2015. "Optimizing the Decennial Census for Mobile - A Case Study." In *Paper Presented at the Federal Committee on Statistical Research Conference (FCSM)*. Washington, DC. <https://bit.ly/2J3dSXZ>.
- Wang, L., C. Antoun, R. Sanders, E. Nichols, E. Olmsted-Hawala, B. Falcone, I. Figueroa, J. Katz, and A. Rivas. 2018. "Evidence-Based Standards and Guidelines for Mobile Survey Instrument Design." In *Paper Presented at the Federal Committee on Statistical Research Conference (FCSM)*. Washington, DC, March 7–9. <https://bit.ly/2XjhYPT>.
- World Values Survey. 2012. *Wave 6 (2010-2014)*. <http://www.worldvaluessurvey.org/WVSDocumentationWV6.jsp>.
- Zhou, Jia, Pei-Luen Patrick Rau, and Gavriel Salvendy. 2012. "Use and Design of Handheld Computers for Older Adults: A Review and Appraisal." *International Journal of Human-Computer Interaction* 28 (12): 799–826. <https://doi.org/10.1080/10447318.2012.668129>.