Mathematical Population Studies

An International Journal of Mathematical Demography

ISSN: 0889-8480 (Print) 1547-724X (Online) Journal homepage: http://www.tandfonline.com/loi/gmps20

Sliders, visual analogue scales, or buttons: Influence of formats and scales in mobile and desktop surveys

Vera Toepoel & Frederik Funke

To cite this article: Vera Toepoel & Frederik Funke (2018) Sliders, visual analogue scales, or buttons: Influence of formats and scales in mobile and desktop surveys, Mathematical Population Studies, 25:2, 112-122, DOI: 10.1080/08898480.2018.1439245

To link to this article: https://doi.org/10.1080/08898480.2018.1439245

MATHEMATICAL POPULATION STUDIES
2018, VOL. 25, NO. 2, 112–122
https://doi.org/10.1080/08898480.2018.1439245

Routledge
Taylor & Francis Group

Sliders, visual analogue scales, or buttons: Influence of formats and scales in mobile and desktop surveys

Vera Toepoel[a] and Frederik Funke[b]

[a]Department Methods and Statistics, Utrecht University, Utrecht, Netherlands; [b]Independent Scholar, Germany

ABSTRACT
In an experiment dealing with the use of personal computer, tablet, or mobile, scale points (up to 5, 7, or 11) and response formats (bars or buttons) are varied to examine differences in mean scores and nonresponse. The total number of "not applicable" answers does not vary significantly. Personal computer has the lowest item nonresponse, followed by mobile and tablet, and a lower mean score than for mobile. Slider bars showed lower mean scores and more nonresponses than buttons, indicating that they are more prone to bias and difficult in use. Sider bars, which work with a drag-and-drop principle, perform worse than visual analogue scales working with a point-and-click principle and buttons. Five-point scales have more nonresponses than eleven-point scales. Respondents evaluate 11-point scales more positively than shorter scales.

KEYWORDS
Likert scale; mobile surveys; questionnaire design; response formats; slider bars; visual analogue scales

1. Introduction

Responses may depend on the devices used to complete online surveys. Web survey designers mostly gather data with rating scales looking like radio buttons, but can also choose visual analogue scales, which are often used in the medical sector, or slider scales, which are popular in market research (Funke, 2015). Smith (1995), Jenkins and Dillman (1995), Couper et al. (2004), Toepoel et al. (2009), and Toepoel and Dillman (2011) showed that response formats affect responses. Couper et al. (2004) showed that responses depend on the visual presentation of questions. These visibility effects may be increased on mobile phones, which have smaller screens, and by long response formats, including the total number of scale points.

In an experiment conducted along the *Growth from Knowledge Online Panel* in the Netherlands, we asked respondents to complete the survey on either a regular desktop personal computer, a tablet, or a mobile phone. We varied the types of radio buttons and bars, as well as the scales up to 5, 7, or

CONTACT Vera Toepoel v.toepoel@uu.nl Utrecht University, Padualaan 14, 3584CH Utrecht, The Netherlands.

11 points. We assess the frequency of response, nonresponse, and evaluation as dependent on the type of device.

2. Background

2.1. *What device to conduct surveys?*

Online surveys where respondents can complete surveys on a regular personal computer, a tablet, or a mobile phone, increasingly resemble mixed-mode designs where respondents answer either face-to-face, by telephone, or with paper and pencil. Schouten et al. (2013) argue that different modes generate selection and measurement biases. The decision for a mode of data collection depends on response rates, costs, and delays. Respondents also choose the device with which they complete the questions. Do we encounter the same difficulties in mixed-device surveys?

Poggio et al. (2015) report 4% of panel members in the GESIS Pilot Panel (Germany) accessing surveys through their mobile phones in 2011 and 2012. Bruijne and Wijnant (2014) report up to 16% in the CentERpanel (the Netherlands) in 2013. Amin et al. (2016) report 10% of all panel members in the American Life Panel to use their mobile phones to answer questionnaires, and this for all surveys in a five-wave period ending in 2014. Toepoel and Lugtig (2014) used responsive design—where questionnaire layout is automatically adapted to device use—in a pilot study in the MarketResponse panel. They report that 57% of panel members used their mobile device when being told that the questionnaire was suitable for mobile completion.

The position of an item on a screen differs per device and is dependent on personal preferences. As mobile Internet use increases, people are expecting that surveys are adapted to mobile devices. This changes the way we should design online surveys. Tourangeau et al. (2013) show that the position of an item on a screen has a systematic but not necessarily large effect on answers. They argue that screen position effects may ruin comparisons if items have different screen positions. Bruijne and Wijnant (2013a) found no significant differences in answer distributions in a regular personal computer layout and a mobile web layout using a responsive design (when the software detects the respondent's browser type). Respondents, however, were less satisfied with the mobile than with the regular web layout.

Apart from positioning effects, response formats may interact with the device used. In market research, large buttons or tiles are used in a touchscreen layout. They are easier to pinpoint than small radio buttons. It is unknown if they provide answers comparable to those obtained with small radio buttons. Bars (slider scales or visual analogue scales) might become more popular response formats because they require less space on a

screen than buttons. With small screen sizes on mobile phones, and because visibility determines response endorsement, assessing the influence of answer formats has become necessary at an epoch when people use many media.

2.2. *Influence of the total number of scale points*

In Likert-type rating scales, for which answers range in a continuum, Beuckelaer (De) et al. (2013) argue that 11-point scales allow a higher level of precision, but at the price of a lack of readability, resulting in errors. A seven-point scale is the best compromise. A five-point scale produces inconsistent answer scores. Preston and Colman (2000) also favor Respondents favored response scales with seven or more points for their reliability, validity, and discriminating power. Test-retest reliability tends to decrease for scales with more than 10 response categories, whereas respondents' preferences are highest for 10-point scales, followed by 7- and 9-point scales. Preston and Colman (2000) recall Miller's (1956) theoretical analysis of human information processing to prefer 7-point scales, which is confirmed in the review by Maitland (2009). Krosnick and Fabrigar (1997) found a curvilinear pattern such that scales of 5 to 7 points are the most reliable.

Weathers et al. (2005) claim that involvement with the scale, motivation to respond accurately, or processing capacity could be related to scale reliability and to the total number of scale points. Smarter people could handle more alternatives. Visibility is affected by the total number of scale points. Couper et al. (2004) show that visibility increases endorsement. The visibility of answer options depends on the size of the screen and on personal settings. A 5-point scale is entirely visible on mobile phones for 99% of respondents, whereas an 11-point scale is visible only for 59% of respondents (Bruijne and Wijnant, 2014). The horizontal or vertical orientation of the scale makes no difference (Funke et al., 2011; Bruijne and Wijnant, 2014).

2.3. *Response formats*

Possible formats are radio buttons, big buttons, slider bars, and visual analogue scales. The choice of a format should take account of measurement error (Smith, 1995; Jenkins and Dillman, 1995; Couper et al., 2004; Toepoel et al., 2009; Toepoel and Dillman, 2011).

Radio buttons are round circles in which the respondent clicks to answer. Radio buttons are coded in HTML (Hypertext Markup Language) and work with every browser, but they take much space on the screen and cannot be reduced in size.

Big buttons, in which cells or tiles are clickable, are easy to handle on small screens, but take much space on the screen.

Slider scales, where respondents drag a "handle" to the desired answer point on a bar representing the range of choices, require less space on the screen. Funke et al. (2011) have shown the dependence of the answer on the initial position of the slider.

Visual analogue scales are also on a bar, most often horizontal, but respondents give a rating by positioning their click on the line. Visual analogue scales save space similar to slider scales: 50 response options take the same space as a 3-point scale with radio buttons (Funke, 2015). They can be implemented either as a continuous (with pixel precision) or as a discrete rating scale. Visual analogue scales require JavaScript or another technology available on the respondent's device.

On mobile devices, Sikkel et al. (2014) find that dragging formats are better evaluated by respondents, but that they lose their ergonomic advantage at the follow-up test. They should then be used sparingly. Funke (2015) shows that a small difference in rating can have a large influence on data quality for slider scales; in contrast, visual analogue scales are as robust as radio buttons. This discrepancy still holds true on touch screen devices such as smartphones.

3. Method

Data are those of the Growth from Knowledge Online Panel. They are representative in age over 15, sex, and education of the Dutch population. This nonprobability online access panel is certified by the International Organization for Standardization. We selected respondents owning a personal computer, a tablet, and a mobile phone. Filling the questionnaire lasted 5 minutes. The questionnaire contains three sections and starts with three questions about attitudes towards surveys. Then 16 questions are formulated in the experimental section, and the questionnaire is terminated with seven evaluation questions. We use a fully crossed $3 \times 5 \times 3$ factorial between-subjects experimental design, in which we assign respondents at random:

- Factor 1: either desktop personal computer, tablet, or mobile phone;
- Factor 2: either radio buttons, big buttons, slider, visual analogue scales, or a mix of slider and visual analogue scales;
- Factor 3: either 5-, 7-, or 11-point scales.

Formats are identical in length and color but are either point-and-click or drag-and-drop, and are equipped with either bar or buttons, in order to prevent respondents from confusing the length of the scale with the total number of scale points. In order to distinguish nonresponse from response, the handle in the slider scales is positioned outside the bar, with valid ratings

on the left-hand side. In the mix of sliders and visual analogue scales, respondents have the choice between drag-and-drop and point-and-click. The survey was conducted in April 2014. The response rate on tablets and mobiles is lower than on desktops. We sent a reminder to tablet and mobile users. Table 1 shows the total numbers of invitations and responses. The response percentage is 30%; the nonresponse 34%. We dropped 32% of respondents, because our quota was already fulfilled. The dropout rate is 4%.

4. Results

Many people completed the survey on another device than the one requested (Table 2): among those assigned to a desktop, 15% completed the survey on a tablet and 9% on a mobile phone. Half of the respondents assigned to a tablet or a mobile phone did not comply and used another device.

4.1. *Data quality: response distributions*

We proceed to a preliminary analysis of covariance to characterize self-selection with respect to a particular device. The coefficient of this variable is not significant, which allows us to discard any selection effect with respect to the choice of the device used by the respondent. We then proceed to a one-way between-groups analysis of variance of scores as functions of device, response formats, and total number of scale points. Table 3 shows that, at the 5% confidence level, both device and format matter. Neither the total number of scale points nor the cross-effects are significant. The explained variance is small ($R^2 = 0.01$), the Tukey test indicates that the standardized mean scores for personal computer are significantly lower than for smartphones (mean difference = −0.01, standard error = 0.002). Scores for slider bars as response format are significantly lower than those for radio buttons (mean

Table 1. Frequencies of invitations and responses by device.

	Desktop personal computer	Tablet	Mobile	Total
First invitation	2,694	2,695	2,696	8,085
Second invitation		2,618	6,228	8,846
Total number of invitations	2,694	5,313	8,924	16,931
Total number of responses	1,709	1,702	1,666	5,077

Table 2. Difference between the assigned device and the device used for filling out the questionnaire.

		Assigned			
Used device		Desktop personal computer	Tablet	Mobile	Total number
	Desktop personal computer	76%	28%	19%	1,709
	Tablet	15%	54%	28%	1,702
	Mobile	9%	18%	53%	1,666

Table 3. One-way between-groups analysis of variance of device, format, and total number of scale points on the standardized average score across 16 questions.

	Fisher *F*	Degrees of freedom	Significance
Model	1.19	44	0.18
Intercept	91.07	1	0.00
Device	6.57	2	0.00
Format	3.53	4	0.01
Total number of scale points	0.21	2	0.81
Device × formats	0.96	8	0.47
Device × total number of scale points	0.60	4	0.66
Format × total number of scale points	0.18	8	0.99
Device × format × total number of scale points	0.82	16	0.67
Error		4,435	
Total		4,480	
Corrected total		4,479	

difference = −0.01, standard error = 0.003). The handle of the slider bars was positioned on the left-hand side of the bar, requiring an additional eye movement from the respondent wishing to increase his or her score.

Respondents are randomly distributed among devices (smartphone: 1470, tablet: 1502, personal computer: 1508), response formats (radio buttons: 894, big buttons: 897, sliders: 896, visual analogue scales: 896, mix of slider and visual analogue scale: 897), and total number of scale points (5-point: 1494, 7-point: 1492, 11-point: 1494).

4.2. *Data quality: item nonresponse*

Table 4 presents a one-way between-groups analysis of variance of the total number of missing responses, over 16 questions, as a function of device, response format, and total number of scale points. At the 5% confidence level, the scores change significantly for different devices, response format, and total number of scale points. The response format

Table 4. One-way between-groups analysis of variance as a function of device, format, and total number of scale points of nonresponses across 16 items.

	Fisher *F*	Degrees of freedom	Significance
Model	5.21	44	0.00
Intercept	452.80	1	0.00
Device	22.19	2	0.00
Format	25.21	4	0.00
Total number of scale points	6.65	2	0.00
Device × formats	2.91	8	0.00
Device × total number of scale points	0.47	4	0.76
Format × total number of scale points	2.71	8	0.01
Device × format × total number of scale points	1.53	16	0.08
Error		4,435	
Total		4,480	
Corrected total		4,479	

interacts significantly with device and with the total number of scale points: especially on smartphones, sliders and visual analogue scales are more likely to lead to nonresponse. We noticed no cross-effects between device and the total number of scale points. The explained variance is low (R^2 = 0.05); the Tukey test indicates that nonresponse for personal computers is significantly lower than for smartphones (mean difference of scores = −0.16, standard error = 0.04) and tablets (mean difference = −0.29, standard error = 0.04). Scores for radio buttons are significantly higher than for sliders (mean difference = 0.45, standard error = 0.06) for the mix of slider and visual analogue scale (mean difference = 0.39, standard error = 0.06) and for visual analogue scales (mean difference = 0.16, standard error = 0.06). Radio and big button formats have the fewest nonresponses. Five-point scales have more nonresponses than 11-point scales (mean difference = 0.16, standard error = 0.04).

Table 5 shows that, for 5-point scales, differences in the total number of nonresponses are smaller on personal computers than on tablets and mobile phones. Sliders and the mix of slider and visual analogue scales have more nonresponses than the other formats. For 7-point scales, Table 5 shows that the tablet still has the highest proportion of nonresponses across all devices, but that the differences with other devices are smaller. For 11-point scales, Table 5 shows that personal computers depend less on the response format. Sliders and visual analogue scales have more nonresponses on mobile phones than on other devices.

Table 5. Mean of the total number of nonresponses across 16 items by device and response format.

	Desktop personal computer	Tablet	Mobile
5-point scale			
Radio buttons	0.06	0.15	0.37
Big buttons	0.36	0.20	0.11
Slider	0.33	1.17	1.02
Visual analogue scale	0.2	0.46	0.31
Mix	0.52	1.01	0.69
7-point scale			
Radio buttons	0.05	0.27	0.15
Big buttons	0.21	0.29	0.12
Slider	0.62	0.71	0.50
Visual analogue scale	0.15	0.53	0.28
Mix	0.28	0.83	0.62
11-point scale			
Radio buttons	0.03	0.37	0.08
Big buttons	0.25	0.33	0.11
Slider	0.13	0.46	0.60
Visual analogue scale	0.13	0.41	0.53
Mix	0.16	0.65	0.31

4.3. *Total number of "not applicable" answers*

Every question has a "not applicable" option presented below the list of options. An analysis of variance shows no significant differences in the total number of respondents selecting the option "not applicable" across devices, total number of scale points, and response formats.

4.4. *Evaluation of respondents*

We compute scores of the respondents based on their opinions with respect to "clarity," "fun," "layout," "usability," "difficulty," and "seriousness" of the questioning. The regression of this score in Table 6 has type of devices, total number of scale points, response formats, socio-demographics, and device experience as explanatory variables. Respondents favored 11-point scales over 5- and 7-point scales. Users of mobile phones rated the questionnaire significantly lower than users of personal computers or tablets.

Crossed effects among devices, total number of scale points, and response formats show that respondents who have chosen their device (by ignoring their assignment) are more critical with the questionnaire, except those who have chosen the mobile. Bar conditions are also favored on mobile phones and tablets, particularly for 5-point scales. Men are more critical than women with the questionnaire, singles more than people in couple, old people more than middle-aged respondents and those with higher education or higher income. Regular users of personal computers and mobiles for surveys express more satisfaction with the questionnaire than others.

5. Conclusion

We examined the influence of response formats and devices on response, nonresponse, and evaluation. Participants owning a personal computer, a tablet, and a mobile phone were assigned to three experiments, combining device (personal computer, tablet, or mobile phone), response format (radio buttons, big buttons or tiles, slider bars, visual analogue scales, or a mix of slider and visual analogue scale) and total number of scale points (5, 7, 11). Many respondents did not use the device they were assigned to but chose one on their own. We find no bias due to this selection effect, however.

Slider bars with drag-and-drop perform worse than radio buttons, big buttons, and visual analogue scales. The handle, presented on the left-hand side of the bar, biases responses to the left-hand side of the scale (Funke et al., 2011). Visual analogue scales perform as well as radio buttons. Respondents favored bars over buttons, but visual analogue scales save space compared to radio buttons, which may be handy for Internet survey designers.

Table 6. Evaluation of the questionnaire.

	Standardized coefficient	Significance
Radio buttons	Ref.	
Big buttons	−0.00	0.94
Sliders	−0.01	0.81
Visual analogue scales	−0.04	0.19
Mix	−0.05	0.16
5-point	−0.00	0.99
7-point	Ref.	
11-point	0.08	0.04
Desktops	Ref.	
Tablets	−0.06	0.10
Mobiles	−0.20	0.00
Self-selection of device	−0.05	0.05
Big buttons × 5-point	−0.01	0.69
Big buttons × 11-point	−0.03	0.26
Big buttons × tablet	0.00	0.89
Big buttons × mobile	0.03	0.24
Slider × 5-point	0.02	0.55
Slider × 11-point	−0.01	0.63
Slider × tablet	0.02	0.54
Slider × mobile	0.04	0.15
Visual analogue scale × 5-point	0.00	0.99
Visual analogue scale × 11-point	−0.00	0.96
Visual analogue scale × tablet	0.05	0.04
Visual analogue scale × mobile	0.04	0.08
Mix × 5-point	0.05	0.04
Mix × 11-point	0.01	0.73
Mix × tablet	0.03	0.29
Mix × mobile	0.05	0.04
5-point × tablet	−0.01	0.85
5-point × mobile	−0.02	0.34
11-point × tablet	−0.03	0.25
11-point × mobile	−0.03	0.29
Self-selection of device × tablet	0.03	0.25
Self-selection of device × mobile	0.06	0.01
Sex (1 = man)	−0.09	0.00
Age	0.15	0.00
Age >65 (yes = 1)	−0.06	0.00
High education	−0.15	0.00
High income	−0.03	0.03
Single	−0.05	0.00
Interview duration	−0.02	0.15
Experience with desktops	0.02	0.34
Experience with tablets	−0.01	0.70
Experience with mobiles	−0.00	0.95
Experience with surveys on desktops	0.07	0.00
Experience with surveys on tablets	0.02	0.34
Experience with surveys on mobiles	0.06	0.00

Analysis of variance Fischer test $F(4722, 44) = 9.71$, $p < 0.01$. $R^2 = 0.08$. We found no significant crossed effects with socio-demographics and experimental factors, except that men prefer big buttons and higher educated people dislike 11-point scales.

Respondents find 11-point scales not more difficult than 5- and 7-point scales; they even favor them. However, 11-point scales take too much space on small screens such as those of mobile phones. In our experiment, we used the same length for 5-, 7-, and 11-point scales; but, usually longer

response scales take more space to the detriment of visibility. We showed that respondents, except regular users of mobile surveys, prefer personal computers over mobile devices to fill out the questionnaires. Then, habit can compensate for less comfort of use. We have then clarified some ergonomics involved in surveys. We have shown that response is sensitive to device, formats, and scales. This gives guidelines for survey organizations.

Acknowledgments

The authors thank two anonymous reviewers and guest editor Natalja Menold for helpful comments, and Randy Streng and Frans Louwen from Growth from Knowledge for their help with the data.

References

Amin, A., Lugtig, P., and Toepoel, V. (2016). The effects of adding a mobile-compatible design to the American life panel. *Survey Practice*, *4*(3): 1–8.

Beuckelaer (de), A., Toonen, S., and Davidov, E. (2013). On the optimal number of scale points in graded pair comparisons. *Quality and Quantity*, *47*(5): 2869–2882.

Bruijne (de), M. and 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.

Bruijne (de), M. and Wijnant, A. (2014). Improving response rates and questionnaire design for mobile web surveys. *Public Opinion Quarterly*, *78*(4): 951–962.

Couper, M. P., Tourangeau, F., Conrad, F., et al. (2004). What they see is what we get. Response options for web surveys. *Social Science Computer Review*, *22*(1): 111–127.

Funke, F. (2015). A web experiment showing negative effects of slider scales compared to visual analogue scales and radio buttons. *Social Science Computer Review*, *34*(2): 244–254.

Funke, F., Reips, U.-D., and Thomas, R. K. (2011). Sliders for the smart: type of rating scale on the web interacts with education level. *Social Science Computer Review*, *29*(2): 221–231.

Jenkins, C. R. and Dillman, D. A. (1995). Towards a theory of self-administered questionnaire design. In L. Lyberg, P. Biemer, M. Collins, et al. eds., *Survey Measurement and Process Quality*. New York: Wiley, 165–196.

Krosnick, J. A. and Fabrigar, L. R. (1997). Designing rating scales for effective measurement in surveys. In L. Lyberg, P. Biemer, M. Collins, et al. eds., *Survey Measurement and Process Quality*. New York: Wiley, 141–164.

Maitland, A. (2009). How many scale points should i include for attitudinal questions? *Survey Practice*, *2*(5): 1–2.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *101*(2): 343–352.

Poggio, T., Bosnjak, M., and Weyandt, K. (2015). Survey participation via mobile devices in a probability-based online-panel: prevalence, determinants, and implications for nonresponse. *Survey Practice*, *8*(2): 1–7.

Preston, C. C. and Colman, A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, *104*(1): 1–15.

Schouten, B., Van Den Brakel, J., Buelens, B., et al. (2013). Disentangling mode-specific selection and measurement bias in social surveys. *Social Science Research*, *42*(6): 1555–1570.

Sikkel, D., Steenbergen, R., and Gras, S. (2014). Clicking vs . dragging. different uses of the mouse and their implications for online surveys. *Public Opinion Quarterly*, *78*(1): 177–190.

Smith, T. (1995). Little things matter: a sampler of how differences in questionnaire format can affect survey responses. In American Statistical Association ed., *Proceedings of the American Statistical Association, Survey Research Methods Section*. New York: Springer, 187–201.

Toepoel, V., Das, J. W. M., and Van Soest, A. H. O. (2009). Design of web questionnaires: the effect of layout in rating scales. *Journal of Official Statistics*, *25*(4): 509–528.

Toepoel, V. and Dillman, D. A. (2011). Words, numbers, and visual heuristics in web surveys. *Social Science Computer Review*, *29*(2): 193–207.

Toepoel, V. and 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.

Tourangeau, R., Couper, M., and Conrad, F. (2013). Up means good: the effect of screen position on evaluative ratings in web surveys. *Public Opinion Quarterly*, *77*(1): 69–88.

Weathers, D., Sharma, S., and Niedrich, R. W. (2005). The impact of the number of scale points, dispositional factors, and the status qua decision heuristic on scale reliability and response accuracy. *Journal of Business Research*, *58*(11): 1516–1524.