Research Note

How long do respondents think online surveys should be? New evidence from two online panels in Germany

Melanie Revilla (iD)
RECSM-Universitat Pompeu Fabra, Spain

Jan Karem Höhne (iD)
University of Mannheim, Germany and RECSM-Universitat Pompeu Fabra, Spain

Abstract

In recent years, the number of surveys, especially online surveys, has increased dramatically. Due to the absence of interviewers in this survey mode (who can motivate the respondents to continue answering), some researchers and practitioners argue that online surveys should not be longer than 20 min. However, so far, there has been little research investigating how long respondents think that online surveys should or could be. In this study, we therefore asked respondents of two online panels in Germany (one probability-based panel and one nonprobability panel) about their opinions on the ideal and maximum lengths of surveys. We also investigated whether socio-demographic, personality-related, and survey-related variables were associated with the ideal and maximum lengths reported by respondents. Finally, we compared the stated and observed survey lengths to evaluate the extent to which respondents are able to accurately estimate survey length. Our results suggest that the ideal length of an online survey is between 10 and 15 min and the maximum length is between 20 and 28 min, depending on the measure of central tendency (mean or median) used and the panel. Moreover, we found significant effects of socio-demographics (gender, age, education, and number of persons in household), of personality traits, and survey-related questions (whether the respondents liked the survey, found it easy, and answered from a PC) on at least one of the dependent variables (ideal or maximum lengths). Finally, we found only small differences (less than two min) between stated and observed lengths.

Corresponding author:
Melanie Revilla, Research and Expertise Centre for Survey Methodology, University Pompeu Fabra, Mercè Rodoreda 24.406, Ramón Trías Fargas, 25-27, 08005 Barcelona, Spain.
Email: melanie.revilla@upf.edu

Introduction

Previous research suggests that comparatively long surveys are associated with several negative effects, especially in the case of online surveys. For instance, Crawford et al. (2001) and Marcus et al. (2007) found that a longer announced survey length reduced response rates. Deutskens et al. (2004) found that longer online surveys lead to higher proportions of "don't know" answers and higher break-off rates. Galesic and Bosnjak (2009) found that questions close to the end had higher item non-response rates, shorter answers to open-ended questions, and less variability in answers to grid questions. Overall, this suggests that shorter online surveys should be preferred but provides no suggestions regarding how long they should or could be.

Revilla and Ochoa (2017) asked respondents of an opt-in online panel in Mexico what they consider to be the ideal and maximum survey length and examined how responses varied with socio-demographic characteristics and survey-related attitudes. They found a median ideal length of 10 min and a median maximum length of 20 min. Moreover, the ideal and maximum lengths were significantly associated with how much respondents liked answering the survey and how much trust they had that their data would be treated confidentially. However, the socio-demographic characteristics did not have a significant impact on either measure, except age in the case of maximum length.

In this research note, we build on the study by Revilla and Ochoa (2017), addressing three aspects that the authors recommended for future research (p. 564):

Using a probability-based online panel. We used data from the probability-based German Internet Panel (GIP) and the nonprobability Respondi panel (Germany).

Considering more countries. While Revilla and Ochoa (2017) conducted their study in Mexico, we conducted our study in Germany.

Using further explanatory variables. We additionally include the "Big Five" personality traits: agreeableness, conscientiousness, extraversion, neuroticism, and openness (measured with the BFI-10 [10 Item Big Five Inventory], as proposed by Rammstedt et al. (2013).

Moreover, even though Revilla and Ochoa (2017) were interested in respondents' opinions on online surveys, the questions they used only mentioned the word "surveys." We therefore tested whether mentioning online surveys instead of surveys affects the outcome, by randomly assigning respondents to a similar question formulation as the one used by Revilla and Ochoa (2017) or a question formulation that explicitly stated "online surveys."

Method

Data collection

German Internet Panel (GIP). The GIP is a probability-based online panel, which aims to represent the German population aged 16 to 75. We use data from Wave 42 in July 2019 (Blom et al., 2020). Respondents could answer through the device type of their choice (i.e., PC, smartphone, or tablet). The GIP collects paradata, such as response time. While 4,714 panelists started this wave, 79 broke off before being asked any study-relevant questions, leaving 4,635 panelists for analyses. Supplemental Appendix 1 provides information about the sample composition.

Respondi. Respondi is a nonprobability panel in Germany (www.respondi.com). In this study, the target population was the German population aged 18 to 70. Cross-quotas for age and gender were

Table 1. Formulations of the survey length questions in both experimental groups.

	Group 1 (general)	Group 2 (online)
Stated length	How long do you think that it took you to complete this survey?	How long do you think that it took you to complete this online survey?
Ideal length	What is the ideal length of a survey that you would like to answer?	What is the ideal length of an online survey that you would like to answer?
Maximum length	What is the maximum length of a survey that you would like to answer?	What is the maximum length of an online survey that you would like to answer?

used to represent the German population. Paradata, such as response times were collected using "Embedded Client Side Paradata (ECSP)" (Schlosser & Höhne, 2018). The goal was to obtain 3,000 panelists completing the entire survey: 1,500 through a PC and 1,500 through a smartphone. Panelists were randomly assigned to a device type when invited to the survey. If they entered the survey with a wrong device type (detected through the user-agent-strings), they were blocked and asked to switch to the correct device type. Data were collected in July/August 2019. In total, 3,407 respondents started the survey, but 177 did not answer any of the three main study-relevant questions (see Section 'Surveys and experimental designs'), leaving 3,230 respondents for analyses. Supplemental Appendix 1 provides information about the sample composition.

Surveys and experimental designs

Both online surveys were about 20-min long. They included a variety of political and social topics.

We were mainly interested in three questions placed close to the end of the surveys. For these questions, in each panel, respondents were randomly assigned to one of two experimental groups. Group 1 received a German translation of the question formulations used by Revilla and Ochoa (2017). Group 2 received identical question formulations, with the addition of the word "online." Table 1 provides the formulations of the questions in each group. Respondents were always asked to enter numbers in an open-answer field to express the length in min.

We were also interested in how responses varied across key socio-demographics (gender, age, education, and number of persons in the household) and personality traits (measured with the 10 Item Big Five Inventory by Rammstedt et al., 2013). Furthermore, we used two questions to measure the survey satisfaction: the first asked how much respondents liked answering the survey and the second how easy or difficult respondents found answering the survey. The exact question wording slightly differed across panels (see Supplemental Appendix 2).

Analyses

The analyses were kept as similar as possible to those of Revilla and Ochoa (2017) to facilitate the comparison of the results obtained in both studies.

Ideal and maximum lengths. First, we report the following descriptive statistics for ideal and maximum lengths: minimum, maximum, mean, and median per experimental group and panel. We test for significant differences between experimental groups using *t*-tests for means and K-tests for medians. However, we do not test for significant differences between panels because they differ in key characteristics. Thus, the two panels serve only as case studies, allowing us to investigate whether we would observe similar trends in different contexts.

Table 2. Descriptive statistics for ideal and maximum lengths in minutes.

Length	Group 1 (general)					Group 2 (online)				
	Min	Max	Mean	Median	*n*	Min	Max	Mean	Median	*n*
GIP										
Ideal	1	78	15.3	15.0	2,298	1	180	15.4	15.0	2,291
Maximum	1	180	23.9	20.0	2,299	1	180	23.5	20.0	2,295
Respondi										
Ideal	1	120	13.1	10.0	1,611	1	200	13.0	10.0	1,597
Maximum	2	300	28.7	25.0	1,606	0	720	28.2	20.0*	1,598

GIP: German Internet Panel.
*.01 < *p* < .05

Then, we run for each panel ordinary least squares (OLS) regressions with ideal and maximum lengths as continuous dependent variables. Following Revilla and Ochoa (2017), we include the following independent variables: men (1 = *men*, 0 = *women*), age (14 ascending categories), education level (low and high levels with middle as reference), number of persons in the household (five ascending categories; only available for the GIP), whether the respondent liked answering the survey (five ascending categories for the GIP and seven for Respondi), the easiness of answering the survey (four ascending categories for the GIP and seven for Respondi), and whether the respondent used a PC to answer the survey (1 = *PC*, 0 = *smartphone or tablet* in the GIP and 0 = *smartphone* in the Respondi panel).

In addition, we included the five personality traits of agreeableness, conscientiousness, extraversion, neuroticism, and openness (nine ascending categories each). Each trait was measured with two questions, the responses to which we sum to produce each respondent's final score across each trait. All questions were coded such that a higher value indicates a higher level of the respective personality trait. We also control for the experimental group by including a dummy variable (1 = *online group*, 0 = *general group*).

To deal with outliers in ideal length in the regressions, we excluded respondents who reported an ideal length longer than 60 min, resulting in two exclusions in the GIP and four in the Respondi panel. For the regressions of maximum length, we excluded respondents who reported a maximum length longer than 180 min, resulting in 11 exclusions in the Respondi panel (see Supplemental Appendix 3 for robustness checks).

Stated and observed lengths. Following Revilla and Ochoa (2017), we report the minimum, maximum, mean, and median for stated (reported by respondents) and observed (determined using paradata) lengths as well as the (absolute) difference between the two measures. We test for significant differences between experimental groups using *t*-tests for means and K-tests for medians.

Due to multitasking behavior and/or interruptions during the survey, the observed length is very long for some respondents. We used the highest value of stated length as the threshold to define outliers in observed length and excluded these outliers from the analyses.

Main results

Ideal and maximum lengths: descriptive statistics

Table 2 presents the descriptive statistics for the ideal and maximum lengths for each experimental group and each panel.

Table 3. Regression analyses of ideal and maximum lengths.

Independent variables	GIP		Respondi	
	Ideal	Maximum	Ideal	Maximum
Men	−.09	.43	.56*	2.16**
Age	.13**	.07**	.06**	−.02
Low Education	.70**	.30	1.60**	.03
High Education	−.70**	−.53	−.75**	−.80
No. persons household	−.36**	−.49**	NA	NA
Liked answering the survey	1.30**	2.10**	.83**	2.14**
Easiness answering the survey	−.53**	−.66**	−.31*	.25
PC	−.05	.17	.45*	1.35*
Agreeableness	.12	.02	.18*	.04
Conscientiousness	.27**	−.01	.37**	.40
Extraversion	−.17**	−.34**	.15*	−.21
Neuroticism	−.04	.04	.01	−.02
Openness	−.01	.25**	.12*	.67**
Online group	.04	−.48	−.34	−1.07
Constant	4.93**	15.85**	1.18	7.16
Adjusted R^2	.1590	.0381	.1070	.0316
n	4,514	4,524	3,112	3,102

GIP: German Internet Panel.
*$p < .05$, **$p < .01$.

First, the mean ideal length is between 13.0 and 15.4 min and the median ideal length is between 10 and 15 min (depending on the group and panel). In contrast, the mean maximum length is between 23.5 and 28.7 min and the median maximum length is between 20.0 and 25.0 min. These results are quite similar to those of Revilla and Ochoa (2017, p. 560, Table 1) with slightly longer ideal and maximum lengths in this study. For ideal length, they reported a mean of 12.6 min and a median of 10.0 min. For maximum length, they reported a mean of 22.7 min and a median of 20.0 min.

Second, there are no significant differences between experimental groups, except for maximum length (median) in the Respondi panel. Thus, adding the word "online" in the question, in general, does not affect the reported ideal and maximum lengths.

However, there are some variations across the panels. In terms of ideal length, the GIP respondents have higher means (about 2 min longer) and medians (about 5 min longer) than the Respondi respondents in both experimental groups. Regarding maximum length, in contrast, the GIP respondents have lower means (about 5 min shorter, in both experimental groups) and lower medians (about 5 min shorter, in the first experimental group).

Furthermore, there is a lot of variations within experimental groups and panels. For instance, in Group 2 in the Respondi panel, the ideal length varies from 1 to 200 min and the maximum length varies from 0 to 720 min.

Ideal and maximum lengths: regressions analyses

Next, we investigate the variables that affect ideal and maximum lengths using regression analyses. Table 3 presents the results.

In the ideal length models, the effects of the variables men (Respondi), age (GIP and Respondi), education (GIP and Respondi), and number of persons in the household (GIP) are statistically

Table 4. Descriptive statistics for stated length, observed length, and their (absolute) difference in minutes.

Length	Group 1 (general)					Group 2 (online)				
	Min	Max	Mean	Median	*n*	Min	Max	Mean	Median	*n*
GIP										
Stated	1.0	90	13.2	12.0	2,099	1.0	60	13.4	12.0	2,091
Observed	2.2	78.1	13.3	11.9		3.2	67.7	13.5	11.9	
Difference	−52.8	68.1	0.1	−0.3		−25.8	51.9	0.1	−0.2	
Absolute Difference	0.0	68.1	4.1	2.9		0.0	51.9	4.1	2.9	
Respondi										
Stated	1.0	185	12.4	10.0	1,621	1.0	360	12.1	10.0	1,601
Observed	1.5	297.9	11.3	8.7		1.8	211.3	11.4	8.7	
Difference	−176.8	237.9	−1.2	−2.2		−306.2	196.3	−0.7	−1.9	
Absolute Difference	0.0	237.9	5.5	3.3		0.0	306.2	5.6	3.3	

Results are reported for those who answered the stated length question and for those paradata were available to determine observed length. "Difference" is computed for each respondent as observed minus stated lengths. "Absolute Difference" is computed for each respondent by taking the absolute value of the difference observed minus stated lengths. GIP: German Internet Panel.

significant. In the maximum length models, the effects of the variables men (Respondi), age (GIP), and number of persons in the household (GIP) are statistically significant. This suggests that the ideal and maximum lengths are more strongly related to respondents' main socio-demographic characteristics than was found by Revilla and Ochoa (2017), where the only significant effect was that of age in the maximum length model.

Regarding the survey-related variables, similar to the findings in Revilla and Ochoa (2017), the more respondents liked answering the survey, the longer the ideal and maximum lengths were. Moreover, the easier respondents found the survey, the shorter the reported ideal (GIP and Respondi) and maximum (GIP) lengths were.

While we did not find a significant effect of PC use in the GIP,[1] we found a significant positive effect of PC use on both ideal and maximum lengths in the Respondi panel. To put it differently, respondents answering with a PC tended to state longer ideal and maximum lengths, whereas respondents answering with a smartphone tended to state shorter ideal and maximum lengths. The difference between both panels might be due to the device randomization in the Respondi panel.

Furthermore, we found some significant effects of the personality traits on ideal and maximum lengths, except for neuroticism. Finally, we did not find significant effects for the experimental group on the ideal or maximum lengths.

Stated length versus observed length

Following Revilla and Ochoa (2017), we also compare the stated and observed survey lengths. Table 4 presents the results.

First, we did not expect differences between the experimental groups since respondents were asked the same question on estimating the length of the current survey. This is supported by the analyses.

Second, there are differences between panels and between respondents within each panel. In the GIP, the means and medians of the stated and observed lengths are similar. In the Respondi panel, the observed length is slightly shorter than the stated length. However, the mean and median differences are quite small (up to 1.2 and 2.2 min, respectively). In general, these findings suggest that respondents' length estimations are quite close to what is observed using paradata.

Discussion

This study suggests that the ideal length of an online survey is between 10 and 15 min and the maximum length between 20 and 28 min, depending on the measure of central tendency used and the panel examined. These findings are in line with those reported by Revilla and Ochoa (2017) for a nonprobability panel in Mexico, even though if we obtained somewhat longer ideal and maximum lengths.

Adding the word "online" did not affect the overall results. This could be because respondents have similar opinions about survey length, irrespective of the data collection mode, or because they expected that the questions refer to online surveys, even though this was not stated explicitly. In our opinion, the second explanation is more plausible, because respondents were members of online panels. Furthermore, the first question on stated length specifically asked about the current survey, which, in turn, may have shifted respondents' interpretive framework of the next questions to a framework centered on online surveys.

Moreover, we found some large differences in what respondents consider to be the ideal and maximum survey lengths, suggesting that some respondents can be more receptive to receiving longer online surveys. In the regression analyses, we found significant effects of the main socio-demographics and survey-related questions available in our study (except neuroticism and the experimental group) on at least one of the four dependent variables.

Furthermore, there are only minor discrepancies between the stated and observed lengths.

This study has some limitations that represent avenues for future research. First, several of the variables were not measured in the same way across panels. Second, we only used data from one country (Germany). Thus, no cross-national conclusions can be drawn. Third, even though we controlled for the device type (i.e., PC or mobile device) in the regressions, since mobile device participation is of high interest for many researchers, it would be interesting to focus more specifically on survey length for mobile respondents in future studies (e.g., looking at descriptive analyses depending on the device type). Finally, it would be useful to use a more comprehensive set of paradata, such as browser tab and window switching, to control for multitasking behavior and/or interruptions during the survey that artificially inflate the observed length. Otherwise, differences between stated and observed lengths could be due to imprecise time measures and not due to respondents being unable to properly estimate survey length.

Overall, this study provides important information helping researchers and practitioners to decide about the length of online surveys. In general, 10–15 min seem to be the ideal length, whereas more than 30 min seems too long. Nevertheless, there are large variations between respondents. Depending on the population of interest, fielding surveys of different lengths might be appropriate.

ORCID iDs

Melanie Revilla https://orcid.org/0000-0002-8026-8319
Jan Karem Höhne https://orcid.org/0000-0003-1467-1975

Supplemental material

Supplemental material for this article is available online.

Note

1. We ran the same analyses using a smartphone dummy (1 = *smartphone*; 0 = *PC or tablet*) instead of a PC dummy. In both panels, the effect of the smartphone dummy was not significant.

References

Blom, A. G., Fikel, M., Friedel, S., Höhne, J. K., Krieger, U., & Rettig, T., & SFB 884 "Political Economy of Reforms." (2020). *German Internet Panel, wave 42 (July 2019)*. doi: 10.4232/1.13465

Crawford, S. D., Couper, M. P., & Lamias, M. J. (2001). Web surveys: Perceptions of burden. *Social Science Computer Review*, *19*, 146–162. http://doi.org/10.1177/089443930101900202

Deutskens, E., de Ruyter, K., Wetzels, M., & Oosterveld, P. (2004). Response rate and response quality of internet-based surveys: An experimental study. *Marketing Letters*, *15*, 21–36. http://doi.org/10.1023/B:MARK.0000021968.86465.00

Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly*, *73*, 349–360. http://doi.org/10.1093/poq/nfp031

Marcus, B., Bosnjak, M., Lindner, S., Pilischenko, S., & Schuetz, A. (2007). Compensating for low topic interest and long surveys: A field experiment on nonresponse in web surveys. *Social Science Computer Review*, *25*, 372–383. http://doi.org/10.1177/0894439307297606

Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C., & Kovaleva, A. (2013). A short scale for assessing the Big Five dimensions of personality. *Methods, Data, Analyses*, *7*, 233–249. http://doi.org/10.12758/mda.2013.013

Revilla, M., & Ochoa, C. (2017). Ideal and maximum length for a web survey. *International Journal of Market Research*, *59*, 557–565. http://doi.org/10.2501/IJMR-2017-039

Schlosser, S., & Höhne, J. K. (2018). *Embedded Client Side Paradata (ECSP)*. Zenodo. http://doi.org/10.5281/zenodo.1218941