Asking for feedback: Innovating final comment questions in self-administered web surveys

Joshua Claassen
German Centre for Higher Education Research and Science Studies (DZHW)
Leibniz University Hannover

Jan Karem Höhne
German Centre for Higher Education Research and Science Studies (DZHW)
Leibniz University Hannover

Jessica Kuhlmann
University of Siegen

Abstract
Web surveys frequently include so-called "final comment questions" (FCQs) to provide respondents the opportunity to express their experiences with the survey in general and its questions in particular. A comprehensive analysis of FCQs to enhance survey and question design is often impeded by high item nonresponse and low answer quality in the form of short and uninterpretable answers. In this article, we therefore investigate what respondent and FCQ characteristics drive the provision and quality of answers to FCQs. For this purpose, we conducted two web survey studies in two German online panels. The first Study (N = 874) experimentally varied the visual design of the FCQ ("one multi-line answer box" vs. "ten single-line answer boxes"). The second Study (N = 1,001) experimentally varied the answer format of the FCQ ("request for a text answer" vs. "request for a voice answer"). The results reveal that answer provision is mainly driven by respondent characteristics (e.g., age and survey interest), while answer quality is mainly driven by FCQ characteristics (e.g., request for a voice answer). Overall, this article provides researchers with empirically proven FCQ design recommendations to improve the data quality of future web surveys.

Keywords: answer behavior, visual design, built-in microphone, final comment question, open-ended answers, smartphone survey, survey evaluation

Introduction

Self-administered web surveys are a predominant data collection method in social science research. Most frequently, web surveys include closed questions with pre-defined answer options. Open questions requiring narrative answers are less frequently used, even though they come with methodological advantages. For example, they have great potential to gather rich and in-depth information from respondents. Furthermore, they are well suited for explorative research because

This document is a preprint and thus it may differ from the final version.

they do not require extended knowledge about the object under investigation (Braun et al. 2021; Singer and Couper 2017). A special type of open questions are so-called "final comment questions" (FCQs) that are placed at the very end of web surveys. FCQs allow respondents to express themselves freely and beyond closed questions. The following question is an example of an FCQ: *Do you have any comments or suggestions on the survey as a whole or individual questions from it?* Answers to FCQs can be used to investigate personal narratives and subjective experiences regarding the survey topic. For example, Nalavany et al. (2023) conducted a thematic analysis of FCQ answers in a web survey of adults with dyslexia to learn about their psychosocial experiences. FCQs are also crucial in survey-related research to collect information on respondents' evaluations of the web survey including its questions. Importantly, FCQs shed light on various aspects, such as topic coverage, potential critique, methodological problems, and technological issues, that researchers and practitioners are not aware of when fielding a web survey.

Web surveys frequently struggle with depressed response rates (Daikeler et al. 2020) and the collection of high-quality data (Cornesse and Bosnjak 2018). Asking respondents about their survey experience, difficulties, and preferences through FCQs is thus key for designing respondent-centered web surveys (Wilson and Dickinson 2022) that reduce respondent burden and improve data quality (O'Cathain and Thomas 2004; Schonlau 2015). Nonetheless, when considering the existing survey literature, answers to FCQs are rarely analyzed and reported. The reasons for this circumstance are threefold: First, even though text-as-data methods continuously improve (Schonlau and Couper 2016; Schonlau et al. 2021; Gavras et al. 2022), the coding of narrative answers is often done manually, which is time-consuming and costly (Singer and Couper 2017). Second, many respondents do not provide answers to FCQs and thus they frequently struggle with high item nonresponse (Boscher et al. 2022; Decorte et al. 2019; McLauchlan and Schonlau 2016; Schonlau 2015). Third, answers to FCQs are often short, contain general or non-informative content (Höhne and Claassen 2024), and impede proper analysis (LaDonna et al. 2018).

The reasons for high item nonresponse rates of FCQs and their low answer quality remain open, because there is only a small body of research investigating the relationship between answering FCQs and respondent characteristics and survey or question design. Regarding respondent characteristics, Boscher et al. (2022) argue that respondents' cognitive capacities and motivation are important factors driving the provision of answers to FCQs. The authors found supporting evidence for the latter claim on motivation. Relatedly, Decorte et al. (2019) found that respondents' involvement (or interest) in the survey topic is positively associated with answering FCQs. Regarding FCQ characteristics, in contrast, there is – to our best knowledge – no research that experimentally varies the design of FCQs to investigate its relationship with answer provision and quality.

In this article, we address this research gap to provide new evidence on respondent and FCQ characteristics driving the provision and quality of answers to an FCQ. Therefore, we experimentally investigate answer provision and answer quality across respondent characteristics (e.g., education and survey interest) and different FCQ designs (e.g., text or voice answers). For

this purpose, we conducted two web surveys in two German online panels. The first web survey (called Study 1) experimentally varied the answer format of the FCQ between one multi-line answer box (single-box condition) and ten single-line answer boxes (list-style condition). The second web survey (called Study 2) experimentally varied the answer format to the FCQ between a request for a text answer (text condition) and a request for a voice answer (voice condition). While the first web survey was a mixed-device survey, the second one was a smartphone-only survey.

Background and research hypotheses

Research on FCQs is rare and thus, in our upcoming argumentation, we partially build on research on open narrative questions in general. It is frequently argued that open questions (including FCQs) are cognitively more demanding than closed questions. Open questions do not provide respondents with a frame of reference in the form of answer options. In contrast, respondents must rely on themselves and formulate answers in their own words (Zuell et al. 2015). In line with this reasoning and satisficing theory (Krosnick 1991), it is expected that the provision of narrative answers is associated with respondents' educational level (as an indicator of their cognitive capacities). There is empirical evidence for open questions in general that education is positively associated with answer provision (Scholz and Zuell 2012) and answer quality in terms of length and interpretability (Barth and Schmitz 2021; Kunz et al. 2021; Schmidt et al. 2020).

Answering open questions is burdensome, especially on smartphones with virtual on-screen keypads shrinking the viewing space (Revilla and Ochoa 2016). Therefore, open questions require respondents to make greater effort than their closed counterparts. Importantly, respondents' motivation to give (effortful) answers to survey questions is closely related to their interest in the topic of the survey (Anduiza and Galais 2017; Gummer et al. 2021). Thus, survey interest is a key aspect that may help to explain the provision of answers to open questions (Holland and Christian 2009; Zuell and Scholz 2015) as well as their quality in terms of length, substantiveness (or interpretability), and number of topics (Barth and Schmitz 2021; Holland and Christian 2009; Schmidt et al. 2020).

When investigating FCQs, it is important to look at both answer provision and answer quality, because previous research has shown that many respondents do not provide answers to FCQs (McLauchlan and Schonlau 2016; Schonlau 2015). In addition, answers to FCQs often contain general or non-informative content (Höhne and Claassen 2024). The relevance of education and survey interest for answering FCQs is highlighted by the fact that FCQs are placed at the very end of web surveys. Studies suggest that respondent burden increases and answer quality decreases over time (Galesic and Bosnjak 2009; Jeong et al. 2023; Neuert 2024). Thus, respondents' education (co-responsible for their ability to provide narrative answers) and survey interest (co-responsible for their willingness to exert the effort required) are key aspects for explaining the provision and quality of answers to FCQs. We postulate the following two hypotheses that are tested in Studies 1 and 2:

H1: Respondents with a high educational level and high survey interest are more likely to provide an answer to an FCQ.

H2: Respondents with a high educational level and high survey interest are more likely to provide an answer of high quality to an FCQ.

Respondent characteristics, such as education and survey interest, cannot be directly changed by researchers. In contrast, survey and question design features can be varied and thus they provide promising avenues for increasing the provision and quality of answers to FCQs. Specifically, the visual design of answer boxes is auspicious because there is evidence that respondents use their size and appearance as cues to draw conclusions about the answers they are expected to provide. Large answer boxes tend to stimulate narrative answers (Couper et al. 2011, p. 67) and their size is positively associated with the length of answers (Chaudhary and Israel 2016; Christian and Dillman 2004). In contrast to single-box designs, list-style designs in which respondents must enter their answers in multiple single-line answer boxes might be best suited for FCQs because they signal respondents to comprehensively mention all aspects (or topics) that come to mind (Keusch 2014; Meitinger and Kunz 2024; Mohr et al. 2016). However, list-style designs are likely to be associated with a higher burden because respondents are asked to retrieve as much information as possible and to condense this information into a few, concise words per line (Meitinger and Kunz 2024). In correspondence to our reasoning, we expect that the list-style design decreases respondents' willingness to provide an answer to an FCQ but increases the share of substantive answers and the number of topics. We postulate the following three hypotheses that are tested in Study 1:

H3a: Respondents who receive ten single-line answer boxes (list-style condition) are less likely to provide an answer to an FCQ than respondents who receive one multi-line answer box (single-box condition).

H3b: Respondents who receive ten single-line answer boxes (list-style condition) are more likely to provide a substantive answer to an FCQ than respondents who receive one multi-line answer box (single-box condition).

H3c: Respondents who receive ten single-line answer boxes (list-style condition) are more likely to provide a higher number of topics to an FCQ than respondents who receive one multi-line answer box (single-box condition).

The trend towards smartphone usage in web surveys has been increasing in recent years (Gummer et al. 2019, 2023; Peterson et al. 2017; Revilla et al. 2016). For example, in the first regular wave of the probability-based German Internet Panel (September 2012), only 4% of respondents participated with a smartphone. In the March 2024 wave of the German Internet Panel, already 45% of respondents participated with a smartphone. This smartphone increase introduces novel measurement opportunities that potentially reduce respondent burden and improve data quality (Revilla 2022). Specifically, it enables researchers to ask open questions with requests for voice instead of text answers releasing respondents from the burden of typing in their answers via a virtual on-screen keypad (Ruan et al. 2017). Thus, FCQs with a request for voice answers are a promising way to improve respondent experience and data quality (Gavras et al. 2022; Höhne et

al. 2024; Revilla and Couper 2021; Revilla et al. 2020; Schober et al. 2015). By partially simulating everyday conversations, requests for voice answers have the potential to encourage respondents to engage in open narrations, reducing the respondent burden associated with answers to FCQs (Gavras and Höhne 2022; Gavras et al. 2022; Höhne et al. 2024). This potentially helps to increase the share of substantive answers to FCQs, since respondents are more likely to elaborate on their arguments.

Despite the growing use of voice inputs in everyday life (e.g., voice messages sent via messenger apps), respondents' willingness to provide voice answers in web surveys designed for smartphones is still rather low (Höhne 2023; Lenzner and Höhne 2022; Revilla et al. 2018). For example, in a smartphone survey conducted by Revilla et al. (2020), between 5% and 60% of respondents receiving a request for open voice answers did not provide answers. Requests for text answers resulted in only 2% of respondents not providing an answer. However, Gavras et al. (2022) and Höhne et al. (2024) show that voice answers tend to be of higher quality than their text counterparts. For example, voice answers result in more topics, indicating that voice answers are richer and more in-depth than text answers. In line with this reasoning, we postulate the following three hypotheses that are tested in Study *2*:

H4a: Respondents who receive a request for a voice answer are less likely to provide an answer to an FCQ than respondents who receive a request for a text answer.

H4b: Respondents who receive a request for a voice answer are more likely to provide a substantive answer to an FCQ than respondents who receive a request for a text answer.

H4c: Respondents who receive a request for a voice answer are more likely to provide a higher number of topics to an FCQ than respondents who receive a request for a text answer.

Method

Study 1

Data collection

Data for Study *1* were collected in the non-probability SoSci Panel (www.scoscipanel.de), which is a project of the Institute for Communication Science and Media Research at the Ludwig-Maximilian-University Munich and the German Society for Journalism and Communication Science (DGPuK). It does not pursue any commercial goals. Researchers are eligible to submit study proposals that undergo a review process evaluating the methodological soundness of the studies. By acceptance of the proposals, respondents of the panel (recruited via an opt-in subscription process) are invited to take part in the web surveys. The invitation process via email is administered by researchers of the DGPuK. Web survey data collection is free of charge.

The web survey ran from 16th May 2022 to 5th June 2022 (with a reminder sent on 25th May 2022). Email invitations included information on the topic (i.e., new communication forms in web surveys), the estimated duration of the web survey (approx. 20 min), and a link to the web survey. Respondents could use the device of their choice for web survey completion. The first page of the web survey provided additional details on the web survey and its structure. We also included

a statement of confidentiality, expounding that the study adheres to EU and national data protection laws and regulations. Respondents took part voluntarily without the provision of incentives.

Sample

An invitation email was sent to 5,676 respondents (out of these emails, 68 could not be successfully delivered). A total of 1,146 respondents started the web survey, out of which 874 respondents finished it. The AAPOR Response Rate 1 was about 20% (AAPOR 2023). On average, these respondents were 49 years old and 65% of them were female. In terms of education, 20% had completed intermediate secondary school or less (low/medium education level), while 79% had completed college preparatory secondary school or university-level education (high education level). In total, 63% of respondents participated with a computer, 3% with a tablet, and 35% with a smartphone.

Experimental design

Respondents were randomly assigned to one out of two experimental groups. The first group (n = 436) received an FCQ with one multi-line answer box (single-box condition). The second group (n = 438) received an FCQ with ten single-line answer boxes (list-style condition). To evaluate the effectiveness of random assignment we compared the two experimental groups with respect to age [$t(866) = -0.05$, $p = 0.959$], gender [$\chi^2(1) = 0.43$, $p = 0.511$], education [$\chi^2(1) = 0.00$, $p = 0.99$], and participation device [$\chi^2(2) = 0.46$, $p = 0.79$]. The two groups did not differ statistically with respect to these variables.

Final comment question (FCQ)

At the end of the web survey, we asked respondents the following FCQ:

Finally, we would like to give you the opportunity to say something about our survey. Do you have any comments or suggestions on the survey in general or individual questions in particular?

Respondents received the FCQ with one multi-line answer box (single-box condition) or ten single-line answer boxes (list-style condition). In the latter condition, we used placeholders (Aspect 1, Aspect 2, etc.) to indicate that respondents are supposed to enter one aspect per line. Appendix A includes screenshots of the FCQ.

Analytical strategy

In a first step, we report statistics for answer provision, substantive answers, and topic number across the single-box condition and the list-style condition. To investigate answer provision across respondent characteristics and FCQ designs, we then run logistic regressions with answer provision as dichotomous dependent variable (1 = Yes). We estimate two sequential models. In line with hypothesis 1 and hypothesis 3a, we use high education (1 = Yes; low/medium as reference), survey interest (1 “Not at all interesting” to 7 “Very interesting”), and the experimental condition (1 = List-style condition; single-box condition as reference) as main independent variables in the first model. Following previous research on respondent behavior (Barth and

Schmitz 2021; Lenzner and Höhne 2022; Zuell and Scholz 2015), we add the following variables as control variables in the second model: female (1 = Yes), age (in years), participation device (1 = smartphone, computer/tablet as reference), survey evaluation in terms of difficulty (1 = "Very easy" to 7 = "Very difficult") and topic sensitivity (1 = "Not at all intimate" to 7 "Very intimate"). Appendix C contains English translations of the original question wordings and response categories.

To examine hypothesis 2 as well as hypotheses 3b and 3c, we investigate the provision of substantive answers and the number of topics. Following Revilla and Ochoa (2016), we define non-substantive answers as answers that do not require respondents to think, such as when saying "no comment", "no", or nonsense in the form of "hahaha." Based on these considerations, the third author coded a dichotomous variable: substantive answers (1 = Yes). To estimate interrater reliability, the first author independently coded a random subset of 30% (n = 70) of the answers. We then computed unweighted Cohen's Kappa, resulting in a value of 0.92 (agreement rate = 97%), which indicates an almost perfect agreement (Landis & Koch 1977). Following the coding of substantive answers, the third author manually went through all substantive answers and coded the number of mentioned topics. To estimate interrater reliability, the first author independently coded another random subset of 30% (n = 61) of the substantive answers. We then computed weighted Cohen's Cappa, resulting in a value of 0.73 (agreement rate = 85%). This indicates a substantial agreement (Landis and Koch 1977).

We only include respondents who provided answers to the FCQ (n = 232) in the analysis of substantive answers and only respondents who provided substantive answers to the FCQ (n = 202) in the analysis of topic number (Table D1 in Appendix D reports descriptive statistics). For substantive answers, as dichotomous dependent variable (1 = "yes"), we run logistic regressions. Furthermore, we run zero-truncated poisson regressions for the number of topics as dependent variable, because topic number is a count variable without the occurrence of the value 0. Again, we estimate two sequential models for each of the two dependent variables, using the same model specifications as in the previous analysis on hypotheses 1 and 3a.

Study 2

Data collection

Data for Study 2 were collected in the Forsa Omninet Panel (omninet.forsa.de) in Germany in November 2021. Forsa drew a cross-quota sample from their online panel based on age (young, middle, and old) and gender (female and male). In addition, they drew quotas on education (low, middle, and high). The quotas were calculated based on the German Microcensus as a population benchmark.

The email invitation included information on the device to be used for survey participation (smartphone) and a link that re-directed respondents to the web survey. The first web survey page introduced the topic and outlined the overall procedure. In addition, it included a statement of confidentiality assuring that the study adheres to EU and national data protection laws and regulations. Respondents also received financial compensation for their participation from Forsa.

Sample

Forsa invited 6,745 respondents to take part in the web survey. No respondents were screened out because of full quotas or because they tried to access the web survey with a device other than a smartphone. A total of 1,681 respondents started the web survey, but 680 of them broke off before they were asked any study-relevant questions. In the text condition 159 (about 24%) respondents broke off, whereas in the voice condition 521 (about 51%) respondents broke-off. This leaves us with 1,001 respondents for statistical analyses. The AAPOR Response Rate 1 was about 15% (AAPOR 2023). On average, these respondents were 48 years old, and 49% of them were female. In terms of education, 72% had completed intermediate secondary school or less (low/medium education level), while 28% had completed college preparatory secondary school or university-level education (high education level).

Experimental design

Respondents were randomly assigned to one out of two experimental groups. The first group (n = 500) received an FCQ with a request for a text answer (text condition). The second group (n = 501) received an FCQ with a request for a voice answer (voice condition). To guarantee that differential break-off did not affect the effectiveness of random assignment, we compared the two experimental groups with respect to age [$t(999) = -0.54$, $p = 0.588$], gender [$\chi^2(1) = 0.22$, $p = 0.636$], and education [$\chi^2(1) = 0.88$, $p = 0.349$]. The two groups did not differ statistically with respect to these variables.

Final comment question (FCQ)

At the end of the web survey, we asked respondents the following FCQ:

Finally, we would like to give you the opportunity to say something about our survey. Do you have any comments or suggestions on the survey in general or individual questions in particular?

Respondents received the FCQ with a request for a text answer (text condition) or voice answer (voice condition). Respondents also received instructions on how to provide text and voice answers at the very beginning of the survey. Voice answers were collected using the open source "SurveyVoice (SVoice)" tool by Höhne et al. (2021). Appendix B includes screenshots of the FCQ and English translations of the answer instructions.

Analytical strategy

Before data analysis, the recordings of respondents' voice answers were automatically transcribed by OpenAI's automatic speech recognition system Whisper (Radford et al. 2023). As a quality assurance measure, a student assistant listened to 20% of the recordings (n = 48) and systematically notated any differences between the recordings and the transcripts. The differences were assessed by the first author, revealing only minor discrepancies and an overall high transcription quality.

In a first step, we report statistics for answer provision, substantive answers, and topic number across the text and voice condition. To investigate answer provision across respondent characteristics and FCQ designs, we then run logistic regressions with answer provision as the

dichotomous dependent variable (1 = Yes). We estimate two sequential models. Similar to our analysis in Study 1 and in line with hypothesis 1 and hypothesis 4a, we use high education (1 = Yes, Low/medium as reference), survey interest (1 "Not at all interesting" to 7 "Very interesting"), and the experimental condition (1 = Voice condition; text condition as reference) as main independent variables in the first model. In the second model, we add the same control variables as in Study 1. Appendix C contains English translations of the original question wordings and response categories.

We investigate the provision of substantive answers and the number of topics[1] to examine hypothesis 2 as well as hypotheses 4b and 4c. Using the same coding scheme as in Study 1, the third author coded a dichotomous variable: substantive answers (1 = Yes). To estimate interrater reliability, the first author independently coded a random subset of 30% (n = 138) of the answers. We then computed unweighted Cohen's Kappa, resulting in a value of 1.0 (agreement rate = 100%), which indicates a perfect agreement (Landis and Koch 1977). To determine the topic number, the third author then manually went through all substantive answers and coded the number of topics mentioned. To estimate interrater reliability, the first author independently coded another random subset of 30% (n = 106) of the substantive answers. We subsequently computed weighted Cohen's Cappa, resulting in a value of 0.73 (agreement rate = 79%). This indicates a substantial agreement (Landis and Koch 1977).

As in Study 1, we only include respondents who provided answers to the FCQ (n = 460) in the analysis of substantive answers and only respondents who provided substantive answers to the FCQ (n = 353) in the analysis of topic number (Table D1 in Appendix D reports descriptive statistics). We run logistic regressions with the provision of substantive answers as the dichotomous dependent variable (1 = Yes). Furthermore, we run zero-truncated poisson regressions with the number of topics as the dependent variable. As before, we estimate two sequential models for each of the two dependent variables with the same model specifications as in the previous analysis in Study 1.

Results

Study 1

Descriptive statistics

In a first step, we report the share of provided answers, substantive answers, and the average number of topics across experimental conditions and in total. Table 1 presents the results. The results show that only about 27% of respondents provided an answer to the FCQ. Answer provision does not differ significantly between the single-box condition and the list-style condition. Of the respondents providing an answer, 87% provided a substantive answer. The proportion of substantive answers is significantly higher in the list-style condition (93%) than in the single-box condition (82%). On average, respondents mentioned 1.4 topics per answer. There are no significant differences between the experimental conditions.

[1] Using the same data, Höhne and Claassen (2024) already analyzed the answer length of text and voice answers to the FCQ. The authors report that voice answers are more than three times longer than text answers.

These results provide preliminary evidence that the list-style condition is positively associated with substantive answers, as proposed by hypothesis 3b. However, there is no preliminary evidence for hypotheses 3a, which proposed a negative association between the list-style condition and answer provision, and 3c, which proposed a positive association between the list-style condition and topic number.

Table 1. Statistics across the single-box and list-style conditions and in total (Study 1)

	Answer provision	Substantive answers	Number of topics
Total	26.5 %	87.1 %	1.4
Single-box condition	28.7 %	82.4 %	1.4
List-style condition	24.4 %	92.5 %	1.5
Test statistics	Z(1) = 1.42, p = 0.078	Z(1) = 2.29, p = 0.011	t(200) = -1.32, p = 0.095

Note. We report proportions for answer provision and substantive answers and means for topic number. We computed directed Z-tests for answer provision ($p_{\text{single-box condition}} > p_{\text{list-style condition}}$) and substantive answers ($p_{\text{list-style condition}} > p_{\text{single-box condition}}$) and a directed Student's t-test with equal variances for topic number ($\mu_{\text{list-style condition}} > \mu_{\text{single-box condition}}$).

Regression analyses

In a next step, we investigate variables associated with answer provision, substantive answers, and topic number. To do so, we run separate logistic regressions with answer provision (1 = Yes) and substantive answers (1 = Yes) as dependent variables, respectively. In addition, we run zero-truncated poisson regressions with topic number as the dependent variable (ranging from 1 to 4 topics). For each dependent variable, we estimate two sequential models. In the first model, we include the list-style condition (single-box condition as reference), survey interest, and high education (low/medium education as reference) as independent variables. In the second model, we add female (male as reference), age, survey difficulty, topic sensitivity, and smartphone participation as control variables. Table 2 presents the results.

With respect to answer provision (hypotheses 1 and 3a), only the second model is statistically significant [M1: LR $\chi^2(3) = 2.55$, $p = 0.467$, Pseudo $R^2 = 0.00$; M2: LR $\chi^2(8) = 35.55$, $p < 0.001$, Pseudo $R^2 = 0.04$] indicating that the list-style condition, survey interest, and high education cannot explain answer provision. In the second model, in contrast to our hypotheses, the list-style condition, survey interest, and high education are again not associated with answer provision. To put it differently, the likelihood of providing an answer to the FCQ does not differ between respondents with high and low/medium education, between respondents with high and low survey interest, and between respondents receiving ten single-line answer boxes (list-style condition) and one multi-line answer box (single-box condition), respectively. However, we now find that age is positively associated with answer provision implying that older respondents may be more eager to comply with the web survey instructions.

Table 2. Regression analyses Study 1

	Answer provision (Logistic)		Substantive answers (Logistic)		Number of topics (Poisson)	
	M1 Coefficient (SE)	M2 Coefficient (SE)	M1 Coefficient (SE)	M2 Coefficient (SE)	M1 Coefficient (SE)	M2 Coefficient (SE)
Intercept	-0.75*	-2.88***	2.09*	2.53	-0.71	-2.12
	(0.35)	(0.59)	(0.90)	(1.56)	(0.48)	(0.80)
List-style condition (reference: single-box condition)	-0.22	-0.22	1.16*	1.21*	0.25	0.23
	(0.16)	(0.16)	(0.46)	(0.47)	(0.21)	(0.21)
Survey interest	-0.01	-0.01	-0.17	-0.19	0.08	0.08
	(0.05)	(0.06)	(0.14)	(0.14)	(0.07)	(0.07)
High education(reference: low/medium education)	-0.13	0.14	0.37	0.24	-0.15	-0.13
	(0.19)	(0.20)	(0.46)	(0.49)	(0.25)	(0.25)
Female		0.17		0.24		0.40
		(0.17)		(0.44)		(0.25)
Age		0.03***		-0.01		0.01
		(0.01)		(0.02)		(0.01)
Survey difficulty		0.08		0.17		0.13
		(0.07)		(0.21)		(0.09)
Topic sensitivity		-0.00		0.08		0.10
		(0.05)		(0.13)		(0.07)
Smartphone participation (reference: computer/tablet)		0.16		-0.97*		-0.01
		(0.18)		(0.45)		(0.24)
N	846	846	223	223	194	194
McFadden's Pseudo R^2	0.00	0.04	0.06	0.09	0.01	0.04

Note. *p < .05, **p < .01, ***p < .001. M1 = model 1, M2 = model 2. SE = standard error. Listwise deletion of missing values.

Turning to the provision of substantive answers (hypotheses 2 and 3b), both models are statistically significant [M1: LR $\chi^2(3) = 9.48$, $p = 0.024$, Pseudo $R^2 = 0.06$; M2: LR $\chi^2(8) = 15.67$, $p = 0.047$, Pseudo $R^2 = 0.09$] but the second model has greater explanatory power indicated by the higher Pseudo R^2. In both models and in line with our descriptive results, the list-style condition is positively associated with substantive answers supporting hypothesis 3b. In other words, respondents receiving ten single-line answer boxes (list-style condition) are more likely to provide substantive answers than respondents receiving one multi-line answer box (single-box condition). In addition, smartphone participation is negatively associated with substantive answers indicating that answering open-ended questions on smartphones with virtual on-screen keypads is associated with higher respondent burden impeding the provision of high-quality answers (Revilla & Ochoa 2016). In contrast, survey interest and high education are not associated with substantive answers. Thus, we do not find evidence supporting hypothesis 2.

Finally, with respect to topic number (hypotheses 2 and 3c), both models have a low Pseudo R^2 and fail to reach statistical significance [M1: LR $\chi^2(3) = 3.85$, $p = 0.279$, Pseudo $R^2 = 0.01$; M2: LR $\chi^2(8) = 12.63$, $p = 0.125$, Pseudo $R^2 = 0.04$] implying that the independent variables we included in the model are not associated with topic number. Importantly, in contrast to our hypotheses, there is no association between survey interest, high education, the list-style condition, and topic number. To put it differently, topic number does not differ between respondents with high and low/medium education, between respondents with high and low survey interest, and between respondents receiving ten single-line answer boxes (list-style condition) and one multi-line answer box (single-box condition), respectively.

Study 2

Descriptive statistics

As for Study 1, we first report the percentages of provided answers, substantive answers, and the average number of topics across experimental conditions and in total. Table 3 presents the results. The results show that about 46% of respondents provided an answer to the FCQ. Answer provision does not differ significantly between the text and voice conditions. Of the respondents providing an answer, 77% provided a substantive answer. The proportion of substantive answers is significantly higher in the voice condition (83%) than in the text condition (70%). Respondents mentioned on average 1.4 topics per answer. Voice answers contained about 30% more topics than their text counterparts.

These results provide preliminary evidence that the voice condition is positively associated with substantive answers, as proposed by hypothesis 3b, and topic number, as proposed by hypothesis 3c. However, there is no preliminary evidence for hypothesis 3a, which proposed a negative association between the voice condition and answer provision.

Table 3: Statistics across text and voice conditions and in total (Study 2)

	Answer provision	Substantive answers	Number of topics
Total	46.0 %	76.7 %	1.4
Text condition	44.2 %	69.7 %	1.2
Voice condition	47.7 %	83.3 %	1.6
Test statistics	$Z(1) = 1.11$, $p = 0.867$	$Z(1) = 3.44$, $p < 0.001$	$t(335.08) = -5.97$, $p < 0.001$

Note. We report proportions for answer provision and substantive answers and means for topic number. We computed directed Z-tests for answer provision ($p_{\text{text condition}} > p_{\text{voice condition}}$) and substantive answers ($p_{\text{voice condition}} > p_{\text{text condition}}$) and a directed Student's t-test with unequal variances for topic number ($\mu_{\text{voice condition}} > \mu_{\text{text condition}}$)

Regression analyses

In a next step, we investigate variables associated with answer provision, substantive answers, and topic number. To do so, we run separate logistic regressions with answer provision (1 = yes) and substantive answers (1 = yes) as the dependent variables, respectively, and zero-truncated poisson regressions with topic number as the dependent variable (ranging from 1 to 5 topics). For each dependent variable, we estimate two sequential models. In the first model, we include the voice condition (text condition as reference), survey interest, and high education (low/medium education as reference) as independent variables. In the second model, we add female (male as reference), age, survey difficulty, and topic sensitivity as control variables. Table 4 presents the results.

Looking at answer provision (hypotheses 1 and 4a), both models are statistically significant [M1: LR $\chi^2(3) = 46.94$, $p < 0.001$, Pseudo $R^2 = 0.03$; M2: LR $\chi^2(7) = 78.68$, $p < 0.001$, Pseudo $R^2 = 0.06$] but the explanatory power of the second model is higher indicated by the increase of the Pseudo R^2. In both models, survey interest is positively associated with answer provision while high education is not associated with answer provision. These results partly support hypothesis 1, which proposed a positive association between survey interest, high education, and answer provision. In addition, in contrast to hypothesis 4a, there is no association between the voice condition and answer provision. However, as in Study 1, we find that age is positively associated with answer provision implying that older respondents may be more eager to comply with the web survey instructions.

With respect to substantive answers (hypotheses 2 and 4b), both models are again statistically significant [M1: LR $\chi^2(3) = 15.44$, $p = 0.001$, Pseudo $R^2 = 0.03$; M2: LR $\chi^2(7) = 27.80$, $p < 0.001$, Pseudo $R^2 = 0.06$]. However, of all independent variables in our two models, only the voice condition is associated (positively) with substantive answers. To put it differently, in line with our descriptive results and hypothesis 4b, respondents receiving an FCQ with a voice answer request are more likely to provide substantive answers than respondents receiving an FCQ with a text answer request. In contrast to hypothesis 2, the likelihood of providing substantive answers does not differ between respondents with high and low/medium education and between respondents with high and low survey interest, respectively.

Table 4. Regression analyses Study 2

	Answer provision (Logistic)		Substantive answers (Logistic)		Number of topics (Poisson)	
	M1 Coefficient (SE)	M2 Coefficient (SE)	M1 Coefficient (SE)	M2 Coefficient (SE)	M1 Coefficient (SE)	M2 Coefficient (SE)
Intercept	-1.93***	-3.13***	0.32	0.12	-1.58***	-1.33*
	(0.28)	(0.45)	(0.48)	(0.79)	(0.41)	(0.57)
Voice condition (reference: text condition)	0.16	0.16	0.76***	0.78***	0.90***	0.92***
	(0.13)	(0.13)	(0.23)	(0.23)	(0.19)	(0.19)
Survey interest	0.31***	0.29***	0.07	0.13	0.11	0.11
	(0.05)	(0.05)	(0.08)	(0.09)	(0.06)	(0.07)
High education (reference: low/medium education)	0.08	0.26	0.42	0.24	0.20	0.14
	(0.15)	(0.15)	(0.26)	(0.27)	(0.16)	(0.16)
Female		0.21		-0.22		-0.26
		(0.13)		(0.23)		(0.15)
Age		0.02***		-0.02		-0.01
		(0.00)		(0.01)		(0.01)
Survey difficulty		0.02		0.12		0.01
		(0.05)		(0.09)		(0.05)
Topic sensitivity		-0.02		0.14		0.04
		(0.04)		(0.06)		(0.05)
N	1001	1001	460	460	353	353
McFadden's Pseudo R^2	0.03	0.06	0.03	0.06	0.05	0.06

Note. *p < .05, **p < .01, ***p < .001. M1 = model 1, M2 = model 2. SE = standard error. Listwise deletion of missing values.

Finally, with respect to topic number (hypothesis 2 and 4c), both models are statistically significant [M1: LR $\chi^2(3) = 33.58$, $p < 0.001$, Pseudo $R^2 = 0.05$; M2: LR $\chi^2(7) = 39.03$, $p < 0.001$, Pseudo $R^2 = 0.06$] and do not differ substantially with respect to their explanatory power as indicated by the marginal increase of the Pseudo R^2. In line with our descriptive results and hypothesis 4c, respondents receiving an FCQ with a voice answer request mention on average more topics than respondents receiving an FCQ with a text answer request. However, in contrast to hypothesis 2, there is no association between survey interest, high education, and topic number.

Discussion and conclusion

The aim of this article was to investigate respondent and FCQ characteristics driving the provision and quality of answers to an FCQ. Therefore, we conducted two experimental studies varying the visual design (Study 1) and answer format of the FCQ (Study 2). Similar to previous research (Boscher et al. 2022; Decorte et al. 2019; Höhne and Claassen 2024; McLauchlan and Schonlau 2016; Schonlau 2015), our results show that many respondents provide no answers or only answers of low quality to an FCQ (Studies 1 and 2). Interestingly, answer provision is mainly driven by respondents' age (Studies 1 and 2) and survey interest (Study 2). In contrast, answer quality in terms of substantive answers (Studies 1 and 2) and topic number (Study 2) is mainly driven by FCQ characteristics. In the following, we discuss our empirical findings in light of our hypotheses.

First looking at respondent characteristics, we found no empirical evidence for an association between high education, survey interest, and answer provision in Study 1, in contrast to our first hypothesis. When it comes to Study 2 the results are somewhat mixed: High education is not associated with answer provision, while survey interest is. These findings suggest that respondents' motivation (measured through survey interest) is more important for explaining answer provision than respondents' cognitive capacities (measured through their educational level). Interestingly, we additionally found that older respondents are more likely to provide an answer to an FCQ than younger respondents (Studies 1 and 2). Previous research indicates that age is positively associated with agreeableness and conscientiousness (Allemand et al. 2008; Soto and John 2012). Thus, older respondents may be more eager to comply with the web survey instructions. However, this presumption lacks empirical evidence and needs further, more refined investigation. For instance, it would we worthwhile to measure the Big Five personality traits in future studies investigating answer provision to FCQs.

We found no empirical evidence in Study 1 and Study 2 for an association between high education, survey interest, and answer quality, in contrast to our second hypothesis. We also found no other respondent characteristics to be associated with answer quality, except smartphone participation in Study 1. Thus, future studies should include additional covariates that could possibly explain individual variance in answer quality, such as the Big Five personality traits (Sturgis and Smith 2023) and survey experience (Zhang et al. 2020). Furthermore, it would be important to consider additional indicators of data quality beyond substantive answers and topic number. For instance, answers to FCQs could be coded in terms of whether respondents elaborate on mentioned topics (Smyth et al. 2009).

Now turning to the characteristics of FCQs, we initially look at Study 1 varying the visual design of the FCQ ("one multi-line answer box" vs. "ten single-line answer boxes"). In contrast

to hypothesis 3a, we found no differences with respect to answer provision between the list-style and single-box condition. These results are also supported by the regression analyses controlling for gender, age, survey difficulty, topic sensitivity, and smartphone participation. However, in line with hypothesis 3b, the share of substantive answers was about 10% higher in the list-style condition than in the single-box condition. Furthermore, the list-style condition was positively associated with substantive answers in our regression analyses indicating the robustness of the finding. Our findings indicate that providing respondents with multiple single-line answer boxes does not increase respondent burden but helps respondents to formulate substantive answers that can be interpreted meaningfully. These findings are partly in contrast to previous research indicating that providing respondents with multiple single-line answer boxes increases respondent burden (Keusch 2014). However, it is important to note that our conclusions regarding respondent burden lack proper empirical evidence and thus we encourage future research to examine respondent burden associated with different FCQ characteristics. For instance, investigating response times may shed light on respondent burden. In addition, while previous studies have found respondents to mention more topics when receiving multiple single-line answer boxes (Keusch 2014; Meitinger and Kunz 2024), we found no difference between the list-style and single-box condition (contrary to hypothesis 3c). This finding was supported by our regression analyses.

Turning now to Study 2 varying the answer format of the FCQ ("request for a text answer" vs. "request for a voice answer"), it is important to note that previous studies have consistently found lower answer provision rates for open questions with requests for voice compared to text answers (Gavras et al. 2022; Gavras and Höhne 2022; Revilla and Couper 2021; Revilla et al. 2020). In contrast to hypothesis 4a, we found no differences in terms of answer provision between the voice and text conditions. This was supported by our regression analyses controlling for gender, age, survey difficulty, and topic sensitivity. However, in line with a recent study comparing the quality of text and voice answers to open probing questions (Lenzner et al. under review), the share of substantive answers in the voice condition was about 15% higher than in the text condition, providing initial support for hypothesis 4b. This finding also holds when controlling for gender, age, survey difficulty, and topic sensitivity in the regression analyses, providing new evidence on the data quality benefits of open questions with requests for voice answers (see also Gavras and Höhne 2022). Similar to previous studies on open questions with voice and text answer requests (Gavras et al. 2022; Höhne et al. 2024), we found that voice answers contained about 30% more topics than their text counterparts. Again, this also holds when controlling for gender, age, survey difficulty, and topic sensitivity. This finding can be explained by the different answer processes of text and voice answers: Voice answers are less burdensome than text answers (i.e., respondents only have to press a recording button and record their answer) and trigger open narrations resulting in longer answers containing more topics than text answers (Gavras et al. 2022; Höhne et al. 2024).

This article has some methodological limitations providing avenues for future research. First, both samples were drawn from nonprobability online panels, and thus, we cannot infer to the general population. We encourage future research to go a step further using data collected from a probability-based panel to check the robustness and generalizability of our results. Second, related to the previous point, the distribution of respondents' educational level in Study 1 is skewed. Almost 80% of respondents were highly educated impeding proper testing of our

hypotheses 1 and 2. It would be worthwhile to conduct more refined subgroup analyses with respect to education in future studies. Therefore, it would be necessary to collect a more heterogenous sample when it comes to education. Third, following previous research (see, for example, the literature review by Roberts et al. 2019), we used respondents' educational level as an indicator of their cognitive abilities. Future research should consider more refined indicators of respondents' cognitive abilities, such as vocabulary tests measuring respondents' verbal intelligence (Lenzner 2012). Finally, in line with previous research we found low answer provision rates (about 25% in Study 1 and 45% in Study 2). In addition, answer provision was not associated with FCQ characteristics. Future studies may investigate how to increase answer provision rates with respect to FCQs by, for instance, examining whether and to what extent motivational prompts and additional incentives (e.g., per provided answer) increase answer provision.

This article contributes to the state of research and provides new evidence for the ongoing methodological discussion about FCQs. Most importantly, it shows that answer provision is mainly driven by respondent characteristics, while answer quality is mainly driven by FCQ characteristics. Thus, our findings provide crucial information on how to best design FCQs across respondent groups. In addition, it is key to tailor the design of FCQs to the planned analyses after data collection. For instance, if researchers aim to conduct content analysis of respondents' narratives and survey experiences, answers to the FCQ should be detailed and elaborated. Thus, it may be worthwhile to employ an FCQ with a request for a voice answer facilitating open narrations. In contrast, if researchers aim to obtain a rough overview of whether respondents encountered technical issues while answering the web survey, answers to the FCQ should be rather short and concise. In this case, it may be worthwhile to employ a list-style design facilitating short and substantive answers.

References

Allemand, M., Zimprich, D., and Hendriks, A. A. J. (2008), "Age Differences in Five Personality Domains Across the Life Span," *Developmental Psychology* , 44, 758-770. https://doi.org/10.1037/0012-1649.44.3.758

Anduiza, E., and Galais, C. (2017), "Answering Without Reading: IMCs and Strong Satisficing in Online Surveys," *International Journal of Public Opinion Research* , 29, 497-519. https://doi.org/10.1093/ijpor/edw007

American Association for Public Opinion Research (2023), "Standard Definitions: Final Disposition of Case Codes and Outcome Rates for Surveys," 10th edition. AAPOR.

Barth, A., and Schmitz, A. (2021), "Interviewers' and Respondents' Joint Production of Response Quality in Open-Ended Questions. A Multilevel Negative-Binomial Regression Approach," *Methods, Data, Analysis* , 15, 43-76. https://doi.org/10.12758/mda.2020.08

Boscher, C., Steinle, J., Raiber, L., Fischer, F., and Winter, M. H. J. (2022), "Möchten Sie uns abschließend noch etwas mitteilen? Auswertung der offenen Abschlussfrage in einem sozialwissenschaftlichen Survey," *HeilberufeScience* , 13, 171-178. https://doi.org/10.1007/s16024-022-00376-0

Braun, V., Clarke, V., Boulton, E., Davey, L., and McEvoy, C. (2021), "The Online Survey as a Qualitative Research Tool," *International Journal of Social Research Methodology* , 24, 641-654. https://doi.org/10.1080/13645579.2020.1805550

Chaudhary, A. K.. and Israel. G. D. (2016), "Influence of Importance Statements and Box Size on Response Rate and Response Quality of Open-Ended Questions in Web/Mail Mixed-Mode Surveys," *Journal of Rural Social Sciences* , 31, 140-159.

Cornesse, C., and Bosnjak, M. (2018), "Is There an Association Between Survey Characteristics and Representativeness? A Meta-Analysis," *Survey Research Methods* , 12, 1-13. https://doi.org/10.18148/srm/2018.v12i1.7205

Couper, M. P., Kennedy, C., Conrad, F. G., and Tourangeau, R. (2011), "Designing Input Fields for Non-narrative Open-Ended Responses in Web Surveys," *Journal of Official Statistics* , 27, 65-85.

Christian, L. M., and Dillman, D. A. (2004), "The Influence of Graphical and Symbolic Language Manipulations on Responses to Self-Administered Questions," *Public Opinion Quarterly* , 68, 57-80. https://doi.org/10.1093/poq/nfh004

Daikeler, J., Bosnjak, M., and Manfreda, K. L. (2020), "Web Versus Other Survey Modes: An Updated and Extended Meta-Analysis Comparing Response Rates," *Journal of Survey Statistics and Methodology* , 8, 513-539. https://doi.org/10.1093/jssam/smz008

Decorte, T., Malm, A., Sznitman, S. R., Hakkarainen, P., Barratt, M. J., Potter, G. R., Werse, B.,Kamphausen, G., Lenton, S., and Frank, V. A. (2019), "The Challenges and Benefits of Analyzing Feedback Comments in Surveys: Lessons from a Cross-National Online Survey of Small-Scale Cannabis Growers," *Methodological Innovations* , 12. https://doi.org/10.1177/2059799119825606

Galesic, M., and Bosnjak, M. (2009). "Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey," *Public Opinion Quarterly* , 73, 349-360. https://doi.org/10.1093/poq/nfp031

Gavras, K., and Höhne, J. K. (2022), "Evaluating Political Parties: Criterion Validity of Open Questions with Requests for Text and Voice Answers," *International Journal of Social Research Methodology* , 25, 135-141. https://doi.org/10.1080/13645579.2020.1860279

Gavras, K., Höhne, J. K., Blom, A. G., and Schoen, H. (2022), "Innovating the Collection of Open-Ended Answers: The Linguistic and Content Characteristics of Written and Oral Answers to Political Attitude Questions," *Journal of the Royal Statistical Society (Series A)* , 185, 872-890. https://doi.org/10.1111/rssa.12807

Gummer, T., Höhne, J. K., Rettig, T., Roßmann, J., and Kummerow, M. (2023), "Is There a Growing Use of Mobile Devices in Web Surveys? Evidence from 128 Web Surveys in Germany," *Quality & Quantity* , 57, 5333-5353. https://doi.org/10.1007/s11135-022-01601-8

Gummer, T., Quoß, F., and Roßmann, J. (2019), "Does Increasing Mobile Device Coverage Reduce Heterogeneity in Completing Web Surveys on Smartphones?," *Social Science Computer Review* , 37, 371-384. https://doi.org/10.1177/0894439318766836

Gummer, T., Roßmann, J., and Silber, H. (2021), "Using Instructed Response Items as Attention Checks in Web Surveys: Properties and Implementation," *Sociological Methods & Research* , 50, 238-264. https://doi.org/10.1177/0049124118769083

Höhne, J. K. (2023), "Are Respondents Ready for Audio and Voice Communication Channels in Online Surveys?," *International Journal of Social Research Methodology* , 26, 335-342. https://doi.org/10.1080/13645579.2021.1987121

Höhne, J. K., and Claassen, J. (2024), "Examining final comment questions with requests for written and oral answers," *International Journal of Market Research*. https://doi.org/10.1177/14707853241229329

Höhne, J. K., Gavras, K., and Claassen, J. (2024), "Typing or Speaking? Comparing Text and Voice Answers to Open Questions on Sensitive Topics in Smartphone Surveys," *Social Science Computer Review.* https://doi.org/10.1177/08944393231160961

Höhne, J. K., Gavras, K., and Qureshi, D.D. (2021), "SurveyVoice (SVoice): A Comprehensive Guide for Collecting Voice Answers in Surveys," *Zenodo*. https://doi.org/10.5281/zenodo.4644590

Holland, J. L., and Christian, L. M. (2009), "The Influence of Topic Interest and Interactive Probing on Responses to Open-Ended Questions in Web Surveys," *Social Science Computer Review* , 27, 196-212. https://doi.org/10.1177/0894439308327481

Jeong, D., Aggarwal, S., Robinson, J., Kumar, N., Spearot, A., and Park, D. S., (2023), "Exhaustive or Exhausting? Evidence on Respondent Fatigue in Long Surveys," *Journal of Development Economics* , 161 https://doi.org/10.1016/j.jdeveco.2022.102992

Keusch, F. (2014), "The Influence of Answer Box Format on Response Behavior on List-style Open-Ended Questions," *Journal of Survey Statistics and Methodology* , 2, 305-322. https://doi.org/10.1093/jssam/smu007

Kunz, T., Quoß, F., and Gummer, T. (2021), "Using Placeholder Text in Narrative Open-ended Questions in Web Surveys," *Journal of Survey Statistics and Methodology* , 9, 992-1012. https://doi.org/10.1093/jssam/smaa039

Krosnick, J. A. (1991), "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys," *Applied Cognitive Psychology* , 5, 213-236. https://doi.org/10.1002/acp.2350050305

LaDonna, K. A., Taylor, T., and Lingard, L. (2018), "Why Open-Ended Survey Questions are Unlikely to Support Rigorous Qualitative Insights," *Academic Medicine* , 93, 347-349. https://doi.org/10.1097/acm.0000000000002088

Landis, J. R., and Koch, G. G. (1977), "The Measurement of Observer Agreement for Categorical Data," *Biometrics* , 33, 159-174. https://doi.org/10.2307/2529310

Lenzner, T. (2012), "Effects of Survey Question Comprehensibility on Response Quality," *Field Methods* , 24, 409-428. https://doi.org/10.1177/1525822X12448166

Lenzner, T., and Höhne, J. K. (2022), "Who is Willing to Use Audio and Voice Inputs in Smartphone Surveys, and Why?," *International Journal of Market Research* , 64, 594-610. https://doi.org/10.1177/14707853221084213

Lenzner, T., Höhne, J. K., and Gavras, K. (Under review), "Innovating Web Probing: Comparing Written and Oral Answers to Open-ended Probing Questions in a Smartphone Survey," *Journal of Survey Statistics and Methodology*.

McLauchlan, C., and Schonlau, M. (2016), "Are Final Comments in Web Survey Panels Associated with Next-Wave Attrition?," *Survey Research Methods* , 10, 211-224. https://doi.org/10.18148/srm/2016.v10i3.6217

Meitinger, K., and Kunz, T. (2024), "Visual Design and Cognition in List-Style Open-Ended Questions in Web Probing," *Sociological Methods & Research* , 53, 940-967. https://doi.org/10.1177/00491241221077241

Mohr, A. H., Sell, A., and Lindsay, L. (2016), "Thinking Inside the Box: Visual Design of the Response Box Affects Creative Divergent Thinking in an Online Survey," *Social Science Computer Review* , 34, 347-359. https://doi.org/10.1177/0894439315588736

Nalavany, B. A., Kennedy, R., Lee, M. H., Carawan, L. W., and Knight, S. M. (2023), "Insights from a Web-Based Survey into the Psychosocial Experiences of Adults with Dyslexia: Findings from a Final Comment Question," *Dyslexia* , 29, 441-458. https://doi.org/10.1002/dys.1756

Neuert, C. E. (2024), "The Effect of Question Positioning on Data Quality in Web Surveys," *Sociological Methods & Research* , 53, 279-295. https://doi.org/10.1177/0049124120986207

O'Cathain, A., and Thomas, K. J. (2004), "‚Any Other Comments?'. Open Questions on Questionnaires – a Bane or Bonus to Research?," *BMC Medical Research Methodology* , 4. https://doi.org/10.1186/1471-2288-4-25

Peterson, G., Griffin, J., LaFrance, J., and Li, J. (2017), "Smartphone Participation in Web Surveys," in *Total Survey Error in Practice*, eds. Biemer, P. P., de Leeuw, E., Eckman, S., Edwards, B., Kreuter, F., Lyberg, L. E., Tucker, N. C., and West, B. T., pp. 203-233, Hoboken: John Wiley & Sons.

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., and Sutskever, I. (2023) "Robust Speech Recognition via Large-Scale Weak Supervision," in *Proceedings of the 40th International Conference on Machine Learning*, Honolulu Hawaii, USA, 23–29 July 2023, pp.28492–28518. https://dl.acm.org/doi/10.5555/3618408.3619590

Revilla, M. (2022), "How to Enhance Web Survey Data Using Metered, Geolocation, Visual, and Voice Data?," *Survey Research Methods* , 16, 1-12. https://doi.org/10.18148/srm/2022.v16i1.8013

Revilla, M., and Couper, M. P. (2021), "Improving the Use of Voice Recording in a Smartphone Survey," *Social Science Computer Review* , 39, 1159-1178. https://doi.org/10.1177/0894439319888708

Revilla, M., Couper, M. P., Bosch, O. J., and Asensio, M. (2020), "Testing the Use of Voice Input in a Smartphone Web Survey," *Social Science Computer Review* , 38, 207-224. https://doi.org/10.1177/0894439318810715

Revilla, M., Couper, M. P., and Ochoa, C. (2018), "Giving Respondents Voice? The Feasibility of Voice Input for Mobile Web Surveys," *Survey Practice* , 11, . https://doi.org/10.29115/SP-2018-0007

Revilla, M., and Ochoa, C. (2016), "Open Narrative Questions in PC and Smartphones: Is the Device Playing a Role?," *Quality & Quantity* , 50, 2495-2513. https://doi.org/10.1007/s11135-015-0273-2

Revilla, M., Toninelli, D., Ochoa, C., and Loewe, G. (2016), "Do Online Access Panels Need to Adapt Surveys for Mobile Devices?," *Internet Research* , 26, 1209-1227. https://doi.org/10.1108/IntR-02-2015-0032

Roberts, C., Gilbert, E., Allum, N., and Eisner, L. (2019), "Satisficing in Surveys: A Systematic Review of the Literature," *Public Opinion Quarterly* , 83, 598-626. https://doi.org/10.1093/poq/nfz035

Ruan, S., Wobbrock, J. O., Liou, K., Ng, A., and Landay, J. A. (2017), "Comparing Speech and Keyboard Text Entry for Short Messages in Two Languages on Touchscreen Phones," in

Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies , 1, pp. 1-23. https://doi.org/10.1145/3161187

Scholz, E. and Zuell, C. (2012), "Item Non-Response in Open-Ended Questions: Who Does Not Answer on the Meaning of Left and Right?," *Social Science Research* , 41, 1415-1428. http://dx.doi.org/10.1016/j.ssresearch.2012.07.006

Schmidt, K. , Gummer, T., and Roßmann, J. (2020), "Effects of Respondent and Survey Characteristics on the Response Quality of an Open-Ended Attitude Question in Web Surveys," *Methods, Data, Analyses* , 14, 3-34. https://doi.org/10.12758/mda.2019.05

Schober, M. F., Conrad, F. G., Antoun, C., Ehlen, P., Fail, S., Hupp, A. L., Johnston, M., Vickers, L., Yan, H. Y., and Zhang, C. (2015), "Precision and Disclosure in Text and Voice Interviews on Smartphones," *PLoS One* , 10, 1-20. https://doi.org/10.1371/journal.pone.0128337

Schonlau, M. (2015), "What Do Web Survey Panel Respondents Answer When Asked 'Do You Have Any Other Comment?,'" *Survey Methods: Insights from the Field*. https://doi.org/10.13094/SMIF-2015-00013

Schonlau, M., Gweon, H., and Wenemark, M. (2021), "Automatic Classification of Open-Ended Questions: Check-All-That-Apply Questions," *Social Science Computer Review* , 39, 562-572. https://doi.org/10.1177/0894439319869210

Schonlau, M. and Couper, M. P. (2016), "Semi-Automated Categorization of Open-Ended Questions," *Survey Research Methods* , 10, 143-152. https://doi.org/10.18148/srm/2016.v10i2.6213

Singer, E., and Couper, M. P. (2017), "Some Methodological Uses of Responses to Open Questions and Other Verbatim Comments in Quantitative Surveys," *Methods, Data, Analyses* , 11, 115-134. https://doi.org/10.12758/mda.2017.01

Smyth, J. D., Dillman, D. A., Christian, L. M., and McBride, M. (2009) "Open-Ended Questions in Web Surveys: Can Increasing the Size of Answer Boxes and Providing Extra Verbal Instructions Improve Answer Quality?," *Public Opinion Quarterly* , 73, 325-337. https://doi.org/10.1093/poq/nfp029

Soto, C. J., and John, O. P. (2012) "Development of Big Five Domains and Facets in Adulthood: Mean-Level Age Trends and Broadly Versus Narrowly Acting Mechanisms," *Journal of Personality* , 80, 881-914. https://doi.org/10.1111/j.1467-6494.2011.00752.x

Sturgis, P., and Brunton-Smith, I. (2023), "Personality and Survey Satisficing," *Public Opinion Quarterly* , 87, 689-718. https://doi.org/10.1093/poq/nfad036

Wilson, L. and Dickinson, E. (2022), *Respondent centered surveys. Stop, listen and then design*, London: Sage.

Zhang, C., Antoun, C., Yan, H. Y., and Conrad, F. G. (2020), "Professional Respondents in Opt-in Online Panels: What Do We Really Know?," *Social Science Computer Review* , 38, 703-719. https://doi.org/10.1177/0894439319845102

Zuell, C., Menold, N., and Körber, S. (2015), "The Influence of the Answer Box Size on Item Nonresponse to Open-Ended Questions in a Web Survey," *Social Science Computer Review* , 33, 115-122. https://doi.org/10.1177/0894439314528091

Zuell, C., and Scholz, E. (2015), "Who is Willing to Answer Open-Ended Questions on the Meaning of Left and Right?," *Bulletin de Méthodologie Sociologique* , 127, 26-42. https://doi.org/10.1177/0759106315582199

Appendix A

Screenshots of the FCQ (Study 1)

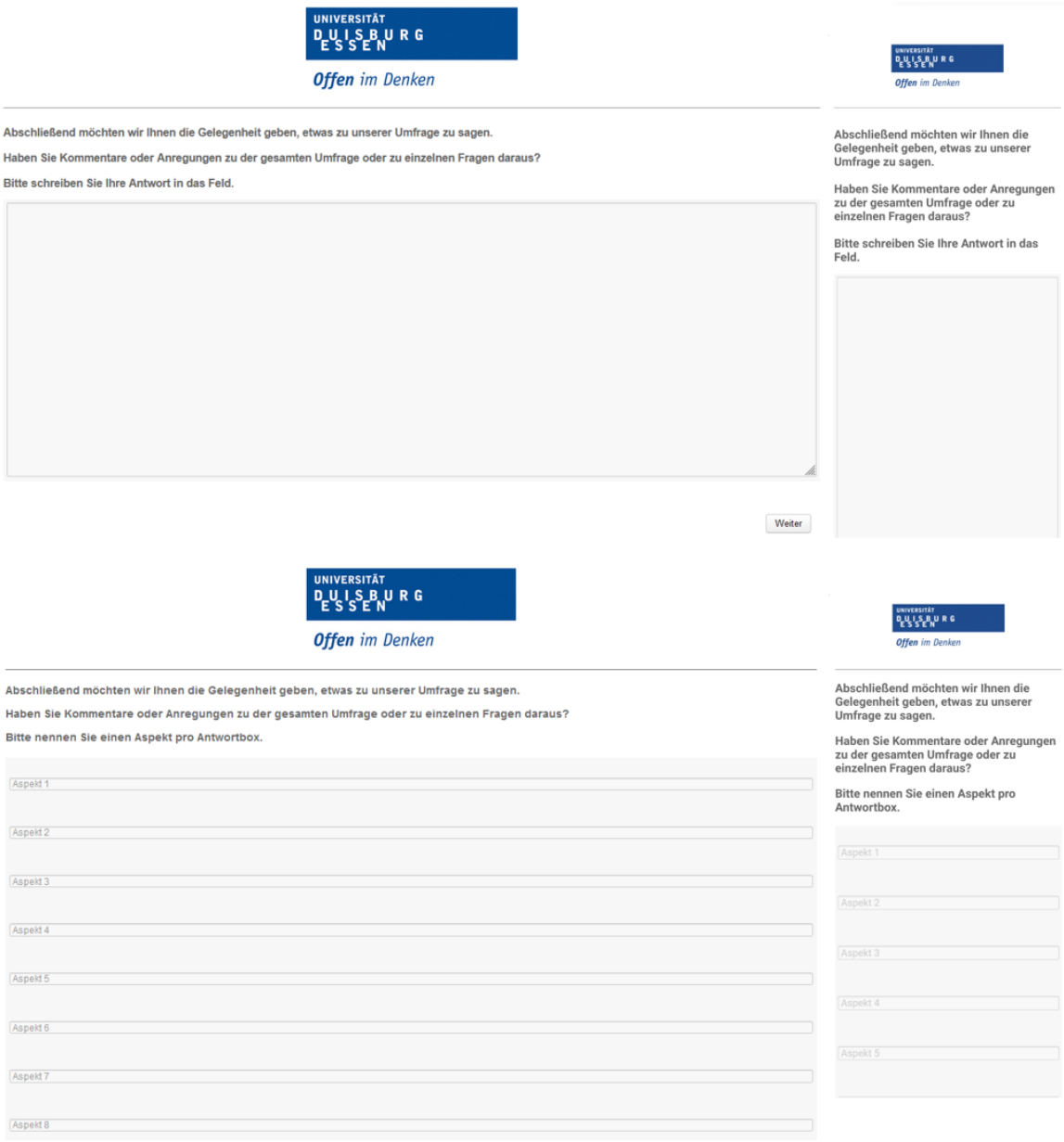

Figure A1. Exemplary screenshots of the FCQ

Note. Multi-line answer box (single-box condition) at the top and ten single-line answer boxes (list-style condition) at the bottom with desktop presentation on the left and smartphone presentation on the right.

Appendix B

Screenshots of the FCQ and English translations of the instructions on how to answer the open questions with requests for text and voice answers (Study 2).

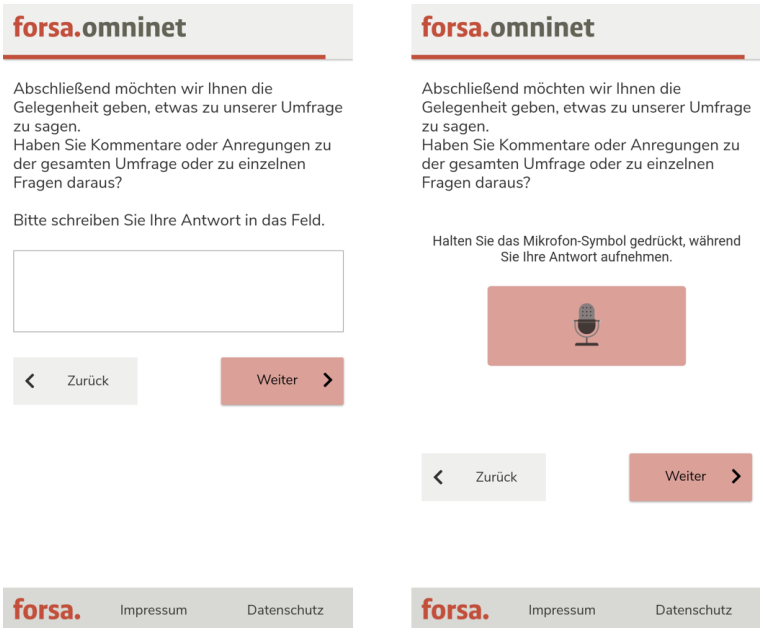

Figure B1. Exemplary screenshots of the FCQ

Note. Request for a text answer (text condition) on the left and request for a voice answer (voice condition) on the right.

Instruction for the text condition

Today we would like to ask you some questions about various social and political issues. You will be asked several times to provide the answers in your own words.
You can enter your answers in the text field via the keyboard of your smartphone.
After successful entry, click on "Next" to continue with the survey as usual.
Of course, your answers will be treated completely confidentially.

Instruction for the voice condition

Today we would like to ask you some questions about various social and political issues. You will be asked several times to give your answers verbally in your own words. You can record your answers via the microphone of your smartphone (similar to WhatsApp or other messaging apps).
Press and hold the microphone icon while recording your answer.
Once you have recorded your answer, you can stop pressing the microphone icon. A tick will indicate that you have successfully recorded your answer. If you want to re-record your answer (e.g., due to recording problems), click on "Delete recording" and simply record your answer again.
After successful recording, click on "Next" to continue with the survey as usual.
Of course, your answers will be kept completely confidential.

Note. These instructions were placed at the beginning of the web survey. The original German wordings of the instructions are available from the first author on request.

Appendix C

English translations of the survey interest and education question and control variables used in the regression analyses.

Survey interest and education question

Survey interest (Studie 1 and 2): How interesting did you find the survey overall? Response categories: 1 'Very interesting' to 7 'Not at all interesting' (recoded into 1 'Not at all interesting' to 7 'Very interesting').

Education (Study 1): What is your highest general school-leaving qualification? Response categories: 1 'Still a student', 2 'Without elementary/main school leaving certificate', 3 'Elementary/main school leaving certificate, 8th or 9th grade, Polytechnische Oberschule (POS) leaving after 8th grade', 4 'Realschulabschluss, 10th grade graduation, graduation from the Polytechnische Oberschule (POS) of the GDR', 5 'Advanced technical college entrance qualification (12th grade)', 6 'General or subject-specific higher education entrance qualification (12th or 13th grade, extended secondary school (EOS), also EOS with apprenticeship)' (categories 5 and 6 recoded into 1 'High education' and categories 1 to 4 into 0 'Low/medium education').

Education (Study 2): Received from survey company.

Control variables used in the regression analyses

Female (Study 1): You are… Response categories: 1 'Male', 2 'Female', 3 'Divers'

Female (Study 2): Received from survey company.

Age (Study 1): In which year were you born (e.g. 1987)? Open text field (recalculated into age).

Age (Study 2): Received from survey company.

Survey difficulty (Studies 1 and 2): How easy or difficult did you find it to answer the questions asked? Response categories: 1 'Very easy' to 7 'Very difficult'.

Topic sensitivity (Studies 1 and 2): How personal did you find answering the questions asked? Response categories: 1 'Very personal' to 7 'Not at all personal' (recoded into 1 'Not at all personal' to 7 'Very personal').

Smartphone participation (Study 1): Extracted from paradata.

Appendix D

Frequency of provided and substantive answers to the FCQ by Study and experimental condition.

Table D1. Frequency and percentages of item nonresponse, provided answers, and (non-) substantive answers across studies

Type of answer	Study 1			Study 2		
	Single-box	List-style	Total	Text	Voice	Total
Item nonresponse	311 (71%)	331 (76%)	642 (73%)	279 (56%)	262 (52%)	541 (54%)
Provided answers	125 (29%)	107 (24%)	232 (27%)	221 (44%)	239 (48%)	460 (46%)
Substantive answers	103 (82%)	99 (93%)	202 (87%)	154 (70%)	199 (83%)	353 (77%)
Non-substantive answers	22 (18%)	8 (7%)	30 (13%)	67 (30%)	40 (17%)	107 (23%)
N (n)	436	438	874	500	501	1001

Note. N (n) = item nonresponse + provided answers. Provided answers = substantive answers + non-substantive answers.