

Effects of Embodied Interviewing Agents on Open Narrative Responses

International Journal of
Market Research
2025, Vol. 0(0) 1–11
© The Author(s) 2025



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/14707853251388213

journals.sagepub.com/home/mre



Jan Karem Höhne¹ , Cornelia Neuert²  and
Joshua Claassen¹ 

Abstract

Open narrative questions in web surveys have the great potential to obtain rich and in-depth information from respondents. However, open narrative questions administered through web surveys frequently suffer from short, uninterpretable, or no responses at all. This especially applies to mobile web surveys, such as smartphone surveys, bearing the risk of not obtaining sufficient information to answer the research question(s) under investigation. Advances in Generative Artificial Intelligence (GenAI) make it possible to enhance respondents' web survey experience by resembling in-person interactions in a self-administered setting. Building on these advances, we investigate web surveys in which open narrative questions are asked through embodied interviewing agents, incorporating features of in-person interviews in web surveys. While the presence of an interviewing agent can encourage more considerate and meaningful responses, it can also introduce social desirability. In this study, we therefore address the following research question: *How do embodied interviewing agents affect responses to sensitive open narrative questions?* We conducted a mobile web survey and randomly assigned respondents to interviewing agents varying in gender (male or female) or a text-based web survey interface without an agent. We employed two open narrative questions on sensitive topics: one on women's role in the workplace and one on family relations. The results of the quantitative text analyses indicate that there are no differences with respect to response length. However, open narrative responses to the male interviewing agent include more topics. There are no differences when it comes to sentiments (or extremity of responses), indicating that social desirability plays a minor role.

Keywords

data quality, open narrative questions, quantitative text analytics, response behavior, mobile web survey, virtual interviewers

¹Leibniz University Hannover, German Centre for Higher Education Research and Science Studies (DZHW), Research Infrastructure and Methods Division, Hannover, Germany

²GESIS – Leibniz Institute for the Social Sciences, Survey Design and Methodology Department, Mannheim, Germany

Corresponding Author:

Jan Karem Höhne, Leibniz University Hannover, German Centre for Higher Education Research and Science Studies (DZHW), Research Infrastructure and Methods Division, Lange Laube 12, 30159 Hannover, Germany.

Email: hoehne@dzhw.eu

Introduction and Background

Web surveys are one of the most frequently used methods for collecting information about respondents' attitudes. Typically, they include closed questions with pre-defined categories (e.g., from "agree" to "disagree"). To overcome the methodological shortcomings of closed questions, such as response styles compromising data quality (van Vaerenbergh & Thomas, 2013), and to gather richer information from respondents, many researchers suggest the employment of open narrative questions that require entering responses in their own words (Lenzner et al., 2024; Revilla & Ochoa, 2016; Singer & Couper, 2017; Smyth et al., 2009; Zuell et al., 2015).

Although open narrative questions are a promising way to improve data quality in web surveys and to collect more in-depth information, they come with some methodological challenges. Research has shown that respondents frequently provide short, uninterpretable, or no responses at all (Lenzner et al., 2024; Revilla & Ochoa, 2016; Singer & Couper, 2017; Smyth et al., 2009; Zuell et al., 2015). For example, Lenzner et al. (2024) reported that up to 12% of respondents engage in item-nonresponse and up to 17% of the responses given are uninterpretable. One explanation for this finding is that the absence of interviewers in web surveys impedes the motivation of respondents to provide considerate and meaningful open narrative responses (Lenzner & Neuert, 2017; Meitinger & Behr, 2016). In addition, short, uninterpretable, or no responses are quite common when respondents participate in web surveys through mobile devices, such as smartphones, with virtual on-screen keypads (Höhne et al., 2020; Revilla & Ochoa, 2016). Considering the constantly increasing rate of mobile devices in the form of smartphones in web surveys (Claassen et al., 2025; Gummer et al., 2023) this is a methodological concern for the administration of open narrative questions in such surveys.

Advances in Generative Artificial Intelligence (GenAI), such as image and text-to-speech generation, may help to improve open narrative response provision in web surveys in general and mobile web surveys in particular by facilitating the creation of interviewing agents. Such agents may help overcome the lack of human interviewers. More specifically, agents can administer open narrative questions through videos implemented in the web survey and vary regarding human (e.g., gender and age) and speech characteristics (e.g., timbre and speech rate). The incorporation of interviewing agents – with life-like appearance and speech – has the great potential to incorporate the quality-improving aspects associated with human interviewers. However, at the same time, they may risk social desirability bias, which is rather low in self-administered web surveys (Kreuter et al., 2008).

There are only very few studies that have employed interviewing agents and even fewer that specifically focus on social desirability bias (Conrad et al., 2015, 2020; Lind et al., 2013). One intriguing example is the study by Conrad et al. (2020) that reported a kind of "race-of-agent" effect. When respondents suggested that they were opposed to giving African Americans preferences in hiring and promotion, they were asked a follow-up question: "Do you oppose preferences in hiring and promotion for African Americans strongly or not strongly?" When the agent that asked the question was White, significantly more respondents responded strongly opposing preferences (about 65%) than when the agent was African American (about 59%). In addition, Lind et al. (2013) compared disclosure of sensitive information between in-person interviews, Audio-Computer-Assisted-Self-Interviews (ACASI), and interviewing agents. The authors reported disclosure differences for various questions. For example, for self-reported reading of newspapers, they found higher self-report rates for in-person interviews (about 49%) than for interviewing agents (about 37%) and ACASI (about 18%).

Considering the literature, there is a lack of research investigating interviewing agents that administer open narrative questions. Existing studies mostly looked at closed and open numeric questions (i.e., requiring respondents to provide digits). When responding to open narrative questions respondents may consider social norms during and after initial response entering, resulting in socially desirable behavior (Gavras et al., 2022; Höhne et al., 2024). Potentially, this is reinforced when interviewing agents (who look and speak like a human) administer the questions because they can introduce a kind of social presence (Conrad et al., 2015; Kreuter et al., 2008). At the same time, interviewing agents may help to enhance the provision of open narrative responses as they potentially give web surveys a human touch, thereby restoring the quality-improving aspects associated with human interviewers. This especially applies to mobile web surveys that are frequently associated with short, uninterpretable, or no responses at all. In this study, we therefore investigate the association between interviewing agents and response behavior using two sensitive open narrative questions (one on women's role in the workplace and one on family relations) implemented in a mobile web survey. Asking respondents such questions can be perceived as sensitive because women's role allocation and division of duties are subject to an ongoing societal discussion and change (Flood et al., 2021; Fraser, 2007). We address the following research question: *How do embodied interviewing agents affect responses to sensitive open narrative questions?*

Method

Data Collection

Data was collected in the nonprobability Respondi/Bilendi online panel in Germany in November and December 2023. The survey language was German. The online panel drew a three-by-two cross-quota sample based on age (young, middle, and old) and gender (male and female). In addition, they drew three quotas on school education (low, medium, and high). The quotas were calculated using the German Microcensus.

Respondents were invited by email, which included information on the device to be used for completion (smartphone or tablet) and a link that redirected respondents to the web survey. To restrict completion to mobile devices, the online panel detected respondents' device type. Respondents who attempted to access the web survey using a non-mobile device were prevented from proceeding and were asked to switch to a mobile device. On the first web survey page, respondents were informed about the survey topic and procedure, the duration (between 5 and 10 min), and that the study adheres to existing data protection laws and regulations.

We created videos of interviewing agents reading questions to be presented during web survey completion, utilizing the synthetic media software HeyGen (<https://www.heygen.com/>). This software generates AI-based agent videos through image diffusion, text-to-speech generation, and speech animation, allowing us to vary agent characteristics, such as gender, appearance, and speech. The videos were implemented in the web survey as required by the experimental design. For web survey administration, we used the Unipark software (<https://www.unipark.com/>).

Sample Characteristics

In total, Respondi/Bilendi invited 35,969 respondents to take part in the web survey. Out of these, 3,340 respondents started the web survey resulting in a response rate of about 9%. However, 1,324 respondents were screened out because of full quotas or not using a smartphone or tablet for web survey participation, and 145 respondents dropped out. Another 718 respondents participated in

a study that is not part of this article. This leaves us with 1,153 respondents for statistical analyses. These respondents were between 18 and 86 years old with a mean age of 49 years, and 51% of them were female. In terms of school education, 43% had completed lower secondary school (low education), 23% intermediate secondary school (medium education), and 33% college preparatory secondary school or more (high education).

Experimental Design

We randomly assigned respondents to one of three experimental groups. The first group ($n = 376$) was asked two open narrative questions by a male agent (*male interviewer condition*). The second group ($n = 395$) was asked the same two questions by a female agent (*female interviewer condition*). The third group ($n = 382$) was asked the same two questions using a text-based web survey interface without an agent (*text control condition*). Figure 1 shows screenshots of the three conditions.

Table 1 reports the sample composition across the three conditions. As shown, the three conditions did not differ regarding age, gender, and education.

Open Narrative Questions and Interviewing Agents

We asked two sensitive open narrative questions dealing with women's role in the workplace and family relations. Each question was accompanied by a text field in which respondents could enter a response. Respondents had to click play to start the video (the agent remained static until the respondent clicked play). Before exposing respondents to the agents (*first and second conditions*), they received a short introduction explaining them that an agent will ask the questions (the

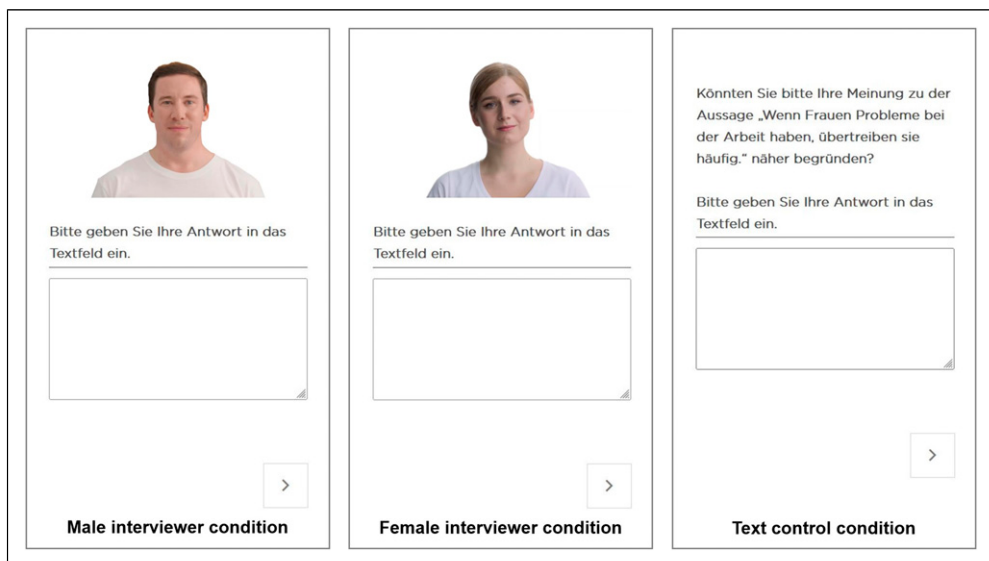


Figure 1. Screenshots of the Two Interviewing Agents and the Text-Based Web Survey Interface. Note. First open narrative question on women's role in the workplace. Male agent on the left (*first condition*), Female agent in the middle (*second condition*), and text-based web survey interface on the right (*third condition*). The agents were created using the synthetic media software HeyGen (<https://www.heygen.com/>)

Table 1. Sample Composition Across the Three Conditions

Variables	Male interviewer condition	Female interviewer condition	Text control condition	Test statistics
Age	49	49	50	$F(2, 1144) = 0.63, p = .54$
Female	50	53	51	$\chi^2(2) = 0.62, p = .73$
Medium education	22	26	22	$\chi^2(4) = 3.87, p = .42$
High education	36	31	34	

Note. We report means for age. For the remaining variables, we report percentages.

[Supplemental Online Material \(SOM\) 1](#) provides English translations of the introduction and all questions used in this study). The agents introduced themselves by saying their name (Alex in both cases) and thanking respondents for their participation. In the third condition, respondents simply received the open narrative questions in text form. These two questions were formulated as follows:

- 1) Could you please explain your response to the statement “Women exaggerate problems they have at work” in more detail? *Please enter your response in the text field.*
- 2) In your opinion, what is the ideal division of labor between men and women in terms of work and family? *Please enter your response in the text field.*

The first open narrative question was preceded by three closed questions on women’s role in the workplace, whereas the second open narrative question was preceded by three additional closed questions on family relations. The six closed questions were also administered by the interviewing agents (*first and second conditions*), but the data from these questions was not analyzed in the present study.

Results

To investigate our research question, we analyze respondents’ open narrative responses using quantitative text analytics. Importantly, we only consider given responses, not considering item-nonresponse. Similar to [Lenzner et al. \(2024\)](#), we distinguish between complete (i.e., leaving the text field blank) and partial item-nonresponse (i.e., instances in which respondents implicitly refused to respond). Item-nonresponse was slightly higher for the first question (about 18%) than for the second question (about 16%).¹ It did not vary between the three conditions: women’s role in the workplace [$\chi^2(2) = 0.42, p = .81$] and family relations [$\chi^2(2) = 0.01, p = .99$].

In a first step, we look at the number of characters included in the open narrative responses using base R. The number of characters informs about the consideration and effort respondents put into their responses ([Gavras et al., 2022](#); [Höhne et al., 2024](#)). Relatedly, we then conduct Structural Topic Models (STMs, [Roberts et al., 2014](#)) employing the *stm* package in R. The *stm* package infers the number of topics mentioned. Importantly, we only consider words mentioned in more than ten responses. We drop stop words and count the number of topics for all responses to which (at least) 10% of the individual responses are attributed. In doing so, we follow research on modeling topics in open narrative responses ([Gavras et al., 2022](#); [Höhne et al., 2024](#)). We employ the following diagnostic criteria ([Roberts et al., 2019](#); [Wallach et al., 2009](#); [Weston et al., 2023](#)): held-out likelihood,

residuals, semantic coherence, and level of lower bound. For the two questions, 20 topics (first question) and 15 topics (second question) can be considered appropriate. The [Supplemental Online Material \(SOM\) 2](#) includes the diagnostic plots. Finally, we investigate the sentiments of the open narrative responses (Pang & Lee, 2008). Sentiments inform about the extremity of responses and thus the scope of socially desirable responses (Gavras et al., 2022; Höhne et al., 2024). Specifically, we use the German sentiment vocabulary SentiWS v2.0 developed by Remus et al. (2010), containing about 3,500 basic word forms and about 30,500 inflections. In SentiWS, words (including inflections) are assigned scores – varying between -1 (negative) and $+1$ (positive) – that suggest the strength of the sentiment-afflicted words.

To investigate whether the interviewing agents affect open narrative responses, we conduct linear hierarchical regressions using number of characters, number of topics, and sentiment ratio as dependent variables (questions nested within respondents). In correspondence with our research question, we include the following independent variables in Model 1, respectively: male interviewer condition (1 = yes) and female interviewer condition (1 = yes) with text control condition as reference. In Model 2, we then control for question content in the form of the family relations question (1 = yes) and demographic characteristics in the form of age (in years), female respondent (1 = yes), and medium education (1 = yes) and high education (1 = yes) with low education as reference. Following previous research (Lenzner & Höhne, 2022; Zuell & Scholz, 2015), we additionally include respondents' survey evaluations in terms of interest (1 “not at all interesting” to 7 “very interesting”), difficulty (1 “very easy” to 7 “very difficult”), personal feeling (1 “not at all personal” to 7 “very personal”), and satisfaction (1 “not at all” to 7 “very much”). This results in two models for each dependent variable.

The [Supplemental Online Material \(SOM\) 3](#) reports survey evaluations across the three conditions. Analyses were conducted in R Studio (version 2024.04.1). For replication purposes, we released a dataset and analysis code via Harvard Dataverse (see <https://doi.org/10.7910/DVN/FGPNC4>).

As shown in [Table 2](#), both interviewer conditions (male and female) are not associated with the length of respondents' open narrative responses (see Model 1). Neither of the two coefficients reaches statistical significance ($p > .05$). This implies that respondents assigned to an interviewer condition provide responses of similar length as respondents assigned to the text control condition. Model 2 reveals that both open narrative questions result in a similar number of characters. This is indicated by the coefficient on the family relations question. Younger, female, and higher educated respondents give longer open narrative responses. Respondents evaluating the survey as being more interesting and less personal also provide longer responses.

The results on topic number draw a somewhat different picture. We now find that the male interviewer condition is associated with the number of topics (see Model 1). Specifically, the positive coefficient indicates that the male interviewer condition results in a higher number of topics than the text control condition. In Model 2, the family relations question is positively associated with topic number, which implies that it includes more topics than the question on women's role in the workplace. Female respondent is positively associated with topic number, whereas difficulty is negatively associated.

Finally, the two interviewer conditions are not associated with the sentiments of respondents' open narrative responses (see Model 1). Both coefficients do not reach statistical significance. This indicates that, compared to the text control condition, the agents do not obtain (more or less) extreme responses. As shown in Model 2, the sentiments of the responses to the family relations question are more positive than those to the question on women's role in the workplace. Female respondents

Table 2. Hierarchical Regressions With Questions Nested in Respondents for the Three Text Analytic Measures

	Number of characters		Number of topics		Sentiment ratio	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
	Coefficients	Coefficients	Coefficients	Coefficients	Coefficients	Coefficients
Male interviewer condition	−4.44 (3.85)	−3.53 (3.84)	0.11* (0.05)	0.13* (0.05)	0.27 (0.33)	0.15 (0.27)
Female interviewer condition	−2.31 (3.81)	−0.96 (3.80)	0.06 (0.05)	0.06 (0.05)	0.46 (0.33)	0.45 (0.27)
Family relation question		0.45 (3.07)		0.64*** (0.04)		5.16*** (0.22)
Age		−0.61*** (0.10)		−0.00 (0.00)		0.00 (0.01)
Female respondent		12.94*** (3.11)		0.14*** (0.04)		−0.47* (0.22)
Medium education		4.28 (4.03)		−0.02 (0.05)		−0.04 (0.29)
High education		21.97*** (3.57)		−0.01 (0.05)		−0.15 (0.25)
Interest		7.12*** (1.71)		−0.03 (0.02)		−0.02 (0.13)
Difficulty		−2.53 (1.95)		−0.06* (0.03)		0.21 (0.15)
Personal feeling		−1.95* (0.91)		0.00 (0.01)		0.01 (0.06)
Satisfaction		0.85 (1.93)		0.03 (0.03)		0.05 (0.14)
Intercept	71.49*** (2.72)	52.06*** (11.90)	2.10*** (0.04)	1.74*** (0.16)	0.04 (0.24)	−2.51** (0.84)
Observations	1,911	1,868	1,704	1,662	1,032	1,005
Adjusted R ²	0.00	0.06	0.00	0.13	0.00	0.35

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. Standard errors in parentheses. Dependent variables: Number of characters, number of topics, and sentiment ratio.

provide less positive responses, which is indicated by the negative coefficient. In contrast to all previous models, the adjusted R²-value is relatively high (>0.30).

Discussion and Conclusion

The goal of this study was to investigate whether embodied interviewing agents affect responses to sensitive open narrative questions. We conducted a mobile web survey in which we randomly assigned respondents to a male agent, female agent, or a text-based web survey interface. Respondents were then asked two sensitive open narrative questions dealing with women's role in the workplace and family relations. The overall results indicate that agents are robust against social

desirability bias and can obtain more information from respondents. However, the amount of information seems to depend on agent characteristics.

The response length analyses did not reveal any differences between the three conditions, whereas the topic model analyses did. Although the presence of interviewing agents does not encourage respondents to provide prolonged responses, the male agent somehow leveraged more information or topics from respondents. Following conversational principles, such as the “maxim of quantity” (Sudman et al., 1996), respondents may keep their responses short making conversational contributions as informative as possible but not more informative than necessary. The degree to which respondents follow conversational principles may depend on agent characteristics, including gender, speech, and animation, shaping the agent’s realism and life-likeness. This is only an attempted explanation that needs further investigation. One way to shed light on this matter is to, for example, utilize (web) probing techniques to gain more insights into the response process of respondents (Lenzner et al., 2024; Lenzner & Neuert, 2017; Meitinger & Behr, 2016).

Having interviewing agents that are visually realistic harbors the danger of resembling social presence (Conrad et al., 2015; Kreuter et al., 2008). Open narrative responses allow respondents to consider social norms and edit the responses accordingly, which opens the floor to social desirability bias (Gavrus et al., 2022; Höhne et al., 2024). Since text-based web surveys commonly suffer less from socially desirable response behavior (they lack cues of social presence; Kreuter et al., 2008), it is assumable that they obtain more extreme responses than web surveys including interviewing agents. This especially applies when asking questions dealing with sensitive topics. Thus, our findings on sentiments indicate that interviewing agents may not necessarily pose a threat to data quality due to social desirability bias.

The advent of Large Language Models (LLMs) allows respondents to easily circumvent the effort associated with responding to open narrative questions as they can prompt LLMs to respond to such questions and simply copy the LLM-based response in the text field. We therefore recommend that survey researchers and practitioners consider this threat to data quality and integrity when administering open narrative questions in web surveys. One simple way is to prohibit copying and pasting in web surveys providing protection to some extent. Another way is to detect LLM-based text in open narrative questions. For example, Claassen et al. (2025) fine-tuned various prediction models that relied on the transformer model BERT. The authors achieved impressive prediction performances. However, prediction performance decreased if the models made predictions about questions on which they were not fine-tuned (cross-corpus). A token analysis revealed that LLM-based text contains specific terms and wordings that distinguish it from the text of human respondents.

This study has some limitations that provide avenues for future research. Similar to previous studies on interviewing agents (see, for example, Conrad et al., 2015; Conrad et al., 2020; Lind et al., 2013), data is based on a nonprobability sample. Although we used quotas on age, gender, and education for building a sample that matches the population on specific benchmarks, this may reduce the generalizability of our results. For example, the general acceptance of interviewing agents in web surveys could be higher among panelists from nonprobability online panels, as they are more frequently confronted with new forms of questions and methods. We therefore recommend investigating interviewing agents in probability-based panels. Another point is that we mainly considered text analytic measures, such as sentiment analysis and topic models, but we did not look at data quality. In our opinion, it is key to investigate data quality more closely. For example, it might be worthwhile to use the sentiment scores to evaluate the correlation between these scores and appropriate criterion variables. This analysis was beyond the scope of this study but would allow researchers to draw conclusions about data quality in terms of criterion validity. Finally, we only

looked at two open narrative questions dealing with sensitive topics. Future research could investigate a more diverse set of questions that varies in terms of sensitivity and topics. In addition, it might be worthwhile to investigate further agent characteristics (e.g., appearance and clothing), agent assignment strategies (e.g., letting respondents decide between agents; see Conrad et al., 2020), and questions on respondents' perceived question sensitivity.

This study provides novel insights into interviewing agents administering open narrative questions. Our study shows that agents, compared to text-based web surveys, perform well without increasing item-nonresponse or provoking socially desirability bias. In contrast, web surveys including agents can be more informative, which is indicated by the number of topics obtained by the male agent. Nonetheless, research on agents is still in its infancy and thus we advocate for further studies putting them to the test. This especially applies to studies that attempt to equip agents with conversational skills by, for example, leveraging Large Language Models (LLMs).

ORCID iDs

Jan Karem Höhne  <https://orcid.org/0000-0003-1467-1975>

Cornelia Neuert  <https://orcid.org/0000-0001-9855-5618>

Joshua Claassen  <https://orcid.org/0009-0002-5492-4439>

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability Statement

The data including analysis code is available for replication purposes via Harvard Dataverse (see <https://doi.org/10.7910/DVN/FGPNC4>).

Ethical Approval and Informed Consent Statements

The study presented in the manuscript was conducted in accordance with established ethical standards. Participants were recruited via an online access panel provider. Membership in the online panel and participation in the surveys is voluntary. The panel members are invited by email to participate in the respective survey. In this invitation, they are informed about the survey topic(s), the length of the survey, and the incentive they receive as a reward for participation. On the welcome page of the web survey, all participants were explicitly informed that they may withdraw from the study at any point without providing any justification. All participants gave their informed consent prior to their participation in the survey.

Supplemental Material

Supplemental material for this article is available online.

Note

1. Complete item-nonresponse was about 3% for the question on women's role in the workplace and about 2% for the question on family relations. In contrast, partial item-nonresponse was about 16% for the question on women's role in the workplace and about 13% for the question on family relations.

References

- Claassen, J., Gummer, T., Rettig, T., & Höhne, J. K. (2025). *How does smartphone participation in web surveys differ across Europe? Evidence from 12 European countries*. ResearchGate. <https://doi.org/10.13140/RG.2.2.23342.42565>
- Claassen, J., Höhne, J. K., Bach, R., & Haensch, A. C. (2025). *Identifying bots through LLM-Generated text in open narrative responses: A proof-of-concept study*. ResearchGate. <https://doi.org/10.13140/RG.2.2.29164.68488>
- Conrad, F. G., Schober, M. F., Jans, M., Orlowski, R. A., Nielsen, D., & Levenstein, R. (2015). Comprehension and engagement in survey interviews with virtual agents. *Frontiers in Psychology*, 6, 1578. <https://doi.org/10.3389/fpsyg.2015.01578>
- Conrad, F. G., Schober, M. F., Nielsen, D., & Reichert, H. (2020). Virtual interviewers, social identities, and survey measurement error. In K. Olson, J. D. Smyth, J. Dykema, A. L. Holbrook, F. Kreuter, & B. T. West (Eds.), *Interviewer effects from a total survey error perspective* (pp. 149–164): CFC Press. <https://doi.org/10.1201/9781003020219>
- Flood, M., Dragiewicz, M., & Pease, B. (2021). Resistance and backlash to gender equality. *Australian Journal of Social Issues*, 56(3), 393–408. <https://doi.org/10.1002/ajs4.137>
- Fraser, N. (2007). Feminist politics in the age of recognition: A two-dimensional approach to gender justice. *Studies in Social Justice*, 1(1), 23–35. <https://doi.org/10.26522/ssj.v1i1.979>
- Gavras, K., Höhne, J. K., Blom, A., & 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*, 185(3), 872–890. <https://doi.org/10.1111/rssa.12807>
- Gummer, T., Höhne, J. K., Rettig, T., Roßmann, J., & Kummerow, M. (2023). Is there a growing use of mobile devices in web surveys? Evidence from 128 web surveys in Germany. *Quality and Quantity*, 57(4), 5333–5353. <https://doi.org/10.1007/s11135-022-01601-8>
- Höhne, J. K., Gavras, K., & Claassen, J. (2024). Typing or speaking? Comparing text and voice answers to open questions on sensitive topics in smartphone surveys. *Social Science Computer Review*, 42(4), 1066–1085. <https://doi.org/10.1177/08944393231160961>
- Höhne, J. K., Schlosser, S., Couper, M. P., & Blom, A. (2020). Switching away: Exploring on-device media multitasking in web surveys. *Computers in Human Behavior*, 111, 106417. <https://doi.org/10.1016/j.chb.2020.106417>
- Kreuter, F., Presser, S., & Tourangeau, R. (2008). Social desirability bias in CATI, IVR, and web surveys: The effects of mode and question sensitivity. *Public Opinion Quarterly*, 72(5), 847–865. <https://doi.org/10.1093/poq/nfn063>
- Lenzner, T., & Höhne, J. K. (2022). Who is willing to use audio and voice inputs in smartphone surveys, and why? *International Journal of Market Research*, 64(5), 594–610. <https://doi.org/10.1177/14707853221084213>
- Lenzner, T., Höhne, J. K., & Gavras, K. (2024). Innovating web probing: Comparing written and oral answers to open-ended probing questions in a smartphone survey. *Journal of Survey Statistics and Methodology*, 12(5), 1295–1317. <https://doi.org/10.1093/jssam/smae031>
- Lenzner, T., & Neuert, C. (2017). Pretesting survey questions via web probing – Does it produce similar results to face-to-face cognitive interviewing? *Survey Practice*, 10(4), 1–13. <https://doi.org/10.29115/SP-2017-0020>
- Lind, L. H., Schober, M. F., Conrad, F. G., & Reichert, H. (2013). Why do survey respondents disclose more when computers ask the questions? *Public Opinion Quarterly*, 77(4), 888–935. <https://doi.org/10.1093/poq/nft038>

- Meitinger, K., & Behr, D. (2016). Comparing cognitive interviewing and online probing: Do they find similar results? *Field Methods*, 28(4), 363–380. <https://doi.org/10.1177/1525822X15625866>
- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2(1–2), 1–135. <https://doi.org/10.1561/1500000011>
- Remus, R., Quasthoff, U., & Heyer, G. (2010). SentiWS - A publicly available German-language resource for sentiment analysis. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odiijk, S. Piperidis, M. Rosner, & D. Tapias (Eds.), *Proceedings of the 7th international conference on language resources and evaluation* (pp. 1168–1171): Valetta. European Language Resources Association. https://www.lrec-conf.org/proceedings/lrec2010/pdf/490_Paper.pdf
- Revilla, M., & Ochoa, C. (2016). Open narrative questions in PC and smartphones: Is the device playing a role? *Quality and Quantity*, 50(6), 2495–2513. <https://doi.org/10.1007/s11135-015-0273-2>
- Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). Stm: An R package for structural topic models. *Journal of Statistical Software*, 91(2), 1–40. <https://doi.org/10.18637/jss.v091.i02>
- Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science*, 58(4), 1064–1082. <https://doi.org/10.1111/ajps.12103>
- Singer, E., & Couper, M. P. (2017). Some methodological uses of responses to open questions and other verbatim comments in quantitative surveys. *Methods, Data, Analyses*, 11(2), 115–134. <https://doi.org/10.12758/mda.2017.01>
- Smyth, J. D., Dillman, D. A., Christian, L. M., & McBride, M. (2009). Open-ended questions in web surveys. *Public Opinion Quarterly*, 73(2), 325–337. <https://doi.org/10.1093/poq/nfp029>
- Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology*: Jossey-Bass Publishers.
- van Vaerenbergh, Y., & Thomas, D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research*, 25(2), 195–217. <https://doi.org/10.1093/ijpor/eds021>
- Wallach, H. M., Murray, I., Salakhutdinov, R., & Mimno, D. (2009). Evaluation methods for topic models. In L. Bottou & M. Littman (Eds.), *Proceedings of the 26th international conference on machine learning* (pp. 1105–1112). Association for Computing Machinery. <https://doi.org/10.1145/1553374.1553515>
- Weston, S. J., Shryock, I., Light, R., & Fisher, P. A. (2023). Selecting the number and labels of topics in topic modeling: A tutorial. *Advances in Methods and Practices in Psychological Science*, 6(2), Article 251524592311601. <https://doi.org/10.1177/25152459231160105>
- Zuell, C., Menold, N., & 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(1), 115–122. <https://doi.org/10.1177/0894439314528091>
- Zuell, C., & Scholz, E. (2015). Who is willing to answer open-ended questions on the meaning of left and right? *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 127(1), 26–42. <https://doi.org/10.1177/0759106315582199>