Examining the Effects of Embodied Interviewing Agents on Open Narrative Responses

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

Cornelia Neuert
GESIS – Leibniz Institute for the Social Sciences

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

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 or no responses at all. This bears 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?* For this purpose, we conducted a 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: 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 interviewing agents 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 surveys; Virtual interviewers

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 methodological shortcomings of closed questions, such as response styles compromising data quality (van Vaerenbergh & Thomas,

---

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

2013), and to gather richer information from respondents, many researchers suggest the employment of open narrative questions (i.e., requiring respondents to enter responses in their own words) (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, they come with some methodological challenges. Research has shown that respondents frequently provide short or no responses at all (Revilla & Ochoa, 2016; Singer & Couper, 2017; Smyth et al., 2009; Zuell et al., 2015). This especially applies when they participate through smartphones with virtual on-screen keypads (Höhne et al., 2020; Revilla & Ochoa, 2016). In addition, the number of not interpretable responses (i.e., vague responses lacking the details for meaningful analyses) is often high. Lenzner et al. (2024), for example, 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).

Advances in Generative Artificial Intelligence (GenAI) may help to overcome the lack of human interviewers in web surveys by installing embodied interviewing agents. Such agents can administer questions to give web surveys a human touch. They can vary regarding human (e.g., gender and age) and speech characteristics (e.g., acoustic color and speech rate). These characteristics cannot be easily changed for human interviewers. It is even possible that respondents select an agent of their choice (Conrad et al., 2020). The incorporation of visually realistic agents has the great potential to bring back 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 a few studies that have employed interviewing agents. For example, Conrad et al. (2015) discovered that agents with advanced dialog capabilities, as opposed to those with limited dialog capabilities, elicited more precise responses. Respondents interacted in a more social way, maintained more eye contact, asked more frequently for clarifications, and rated the agents as more personal and less distant when they included advanced dialog capabilities. For agents with greater facial animation, respondents showed a somewhat higher engagement level (Conrad et al., 2015; Lind et al., 2013). Conrad et al. (2020) reported a decrease in socially desirable responding when the characteristics of the respondent and agent matched. When respondents shared the same ethnicity as the agent, they were more inclined to report being slightly overweight, which is considered a socially undesirable response, compared to when respondents did not share the same ethnicity.

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 (that 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). In this study, we investigate the association between interviewing agents and response behavior using two open narrative questions dealing with women's role in the workplace and family relations. Asking respondents such questions can be seen as sensitive since women's role allocation and division of duties are subject to an ongoing societal discussion and change (Fraser, 2007; Flood et al., 2021). We address the following research question: *How do embodied interviewing agents affect responses to sensitive open narrative questions?*

We conducted a web survey and randomly assigned respondents to 1) a male agent, 2) a female agent, or 3) a text-based web survey interface without an agent. Importantly, the agents appeared visually realistic and were equipped with a natural interplay regarding body language, facial expression, and speech. However, the agents had no dialog capabilities, but simply read questions aloud.

Method

Data collection

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

Respondents were invited by email, which included information on the device (smartphone or tablet) to be used for completion and a link that redirected respondents to the web survey. To restrict completion to mobile devices, the online panel detected respondents' devices. 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 minutes), and that the study adheres to existing data protection laws.

We created videos of interviewing agents reading questions to be presented during web survey completion, utilizing the synthetic media software HeyGen (www.heygen.com). This allowed us to vary agent characteristics, such as gender and appearance. 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, 3,340 respondents started the web survey. However, 1,324 respondents were screened out because of full quotas or not fulfilling the participation requirements, 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 19 and 87 years old with a mean age of 50 years, and 51% of them were female. In terms of school education, 42% had completed lower secondary school (low education), 23% intermediate secondary school (middle education), and 35% college preparatory secondary school or more (high education).

Experimental design

We randomly assigned respondents to one out 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.

Table 1. Sample composition across the three conditions

Variables	Male interviewer condition	Female interviewer condition	Text control condition	Test statistics
Age	50	50	51	F(2,1144) = 0.63, p = 0.54
Female	50	53	51	$\chi^2(2) = 0.62$, p = 0.73
Middle education	22	26	22	$\chi^2(4) = 4.08$, p = 0.40
High education	38	32	34	

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

Open narrative questions and interviewing agents

We asked two 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 (Appendix A 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 data from the six closed 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. This includes complete (i.e., leaving the text field blank) and partial item-nonresponse (i.e., instances in which respondents implicitly refused to respond; e.g., "yes" or "nonsense"). Item-nonresponse was slightly higher for the first question (about 18%) than for

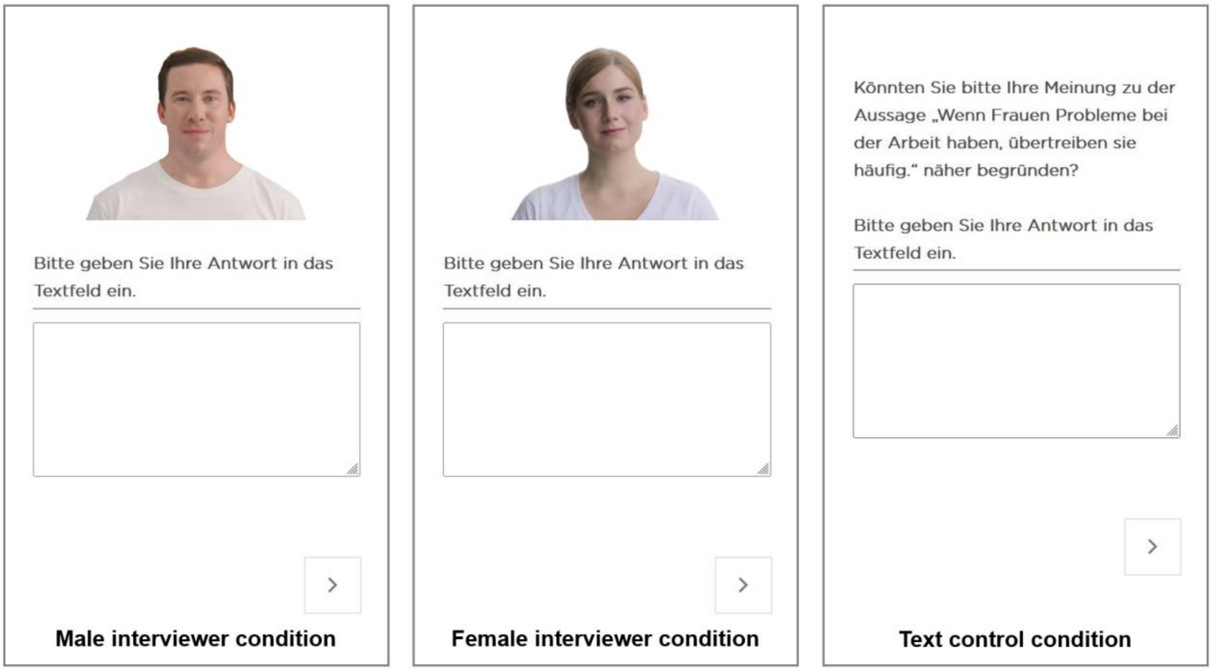

Figure 1. Screenshots of the two interviewing agents and the text-based web survey interface
Note. First open narrative question on women's role in the workplace. Male agent on the left (*first condition*), female agent in the middle (*second condition*), and text-based web survey interface on the right (*third condition*).

the second question (about 16%).[1] It did not vary between the three conditions: women's role in the workplace [$\chi^2(2) = 0.42$, $p = 0.81$] and family relations [$\chi^2(2) = 0.01$, $p = 0.99$].

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 both questions, 20 topics can be considered appropriate. Appendix B 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

[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 14% for the question on family relations.

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 relation question (1 = yes) and demographic characteristics in the form of age (in years), female respondent (1 = yes), and middle 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.

Appendix C reports survey evaluations across the three conditions. Analyses were conducted in R Studio (version 2024.04.1).

As shown in Table 1, both interviewer conditions (male and female) are not associated with the length of respondents' open narrative responses (see Model 1). None of the two coefficients reaches statistical significance ($p > 0.05$). 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 family relation 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 different picture. We now find that the male and female interviewer conditions are associated with the number of topics (see Model 1). Specifically, the positive coefficients indicate that both interviewer conditions result in a higher number of topics than the text control condition. In Model 2, the family relation question is negatively associated with topic number, which implies that it includes less topics than the question on women's role in the workplace. Female respondent is positively associated with topic number, whereas age and middle education are both negatively associated. The survey evaluations show no associations at all.

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, neither agents in the form of male nor female interviewers obtain (more or less) extreme responses. As shown in Model 2, the sentiments of the responses to the family relation question are more positive than those to the question on women's role in the workplace. Female respondents provide less positive responses, which is indicated by the negative coefficient. In contrast to all previous models, the R^2-value is relatively high (> 0.30).

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	-1.66	0.21	0.19***	0.21***	0.31	0.22
	(3.11)	(3.06)	(0.06)	(0.06)	(0.33)	(0.27)
Female interviewer condition	-1.49	-0.03	0.19***	0.20***	0.52	0.48
	(3.08)	(3.03)	(0.05)	(0.06)	(0.33)	(0.27)
Family relation question		-0.10		-0.17***		5.19***
		(2.45)		(0.04)		(0.22)
Age		-0.57***		-0.00*		0.00
		(0.08)		(0.00)		(0.00)
Female respondent		10.63***		0.19***		-0.46*
		(2.48)		(0.05)		(0.22)
Middle education		2.56		-0.12*		-0.07
		(3.24)		(0.06)		(0.29)
High education		18.91***		-0.10		-0.11
		(2.84)		(0.05)		(0.25)
Interest		6.90***		0.01		-0.02
		(1.36)		(0.03)		(0.12)
Difficulty		-2.18		-0.05		0.17
		(1.56)		(0.03)		(0.15)
Personal feeling		-2.34**		-0.00		0.01
		(0.72)		(0.01)		(0.06)
Satisfaction		0.68		0.03		0.06
		(1.54)		(0.03)		(0.14)
Intercept	67.66***	53.25***		1.99***		-2.56**
	(2.20)	(9.52)		(0.18)		(0.84)
Observations	1,911	1,880	1,704	1,674	1,031	1,013
Adjusted R^2	0.00	0.08	0.01	0.04	0.00	0.35

Note. *p < 0.05, **p < 0.01, ***p < 0.001. Standard errors in parentheses. Dependent variables: Number of characters, number of topics, and sentiment ratio.

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 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 obtain more information from respondents, while being robust against social desirability.

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, they somehow leverage more topics or information. Open narrative questions asked by interviewing agents may induce respondents to keep their responses short and more on point. In doing so, respondents may follow conversational principles, such as the maxim of quantity, which requires to make conversational contributions as informative as possible but not more informative than necessary (Sudman et al., 1996, pp. 62–63). Thus, response length does not necessarily align with topic number. 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 (Gavras 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.

This study has some limitations that provide avenues for future research. Similar to previous studies on interviewing agents (see, for example, Conrad et al., 2020; Conrad et al., 2015; 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, 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, 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 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.

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 social desirability. In contrast, web surveys including agents are more informative. 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.

References

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). CRC Press. https://doi.org/10.1201/9781003020219

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

Flood, M., Dragiewicz, M., & Pease, B. (2021). Resistance and backlash to gender equality. Australian Journal of Social Issues, 56, 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, 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 (Series A), 185, 872-890. https://doi.org/10.1111/rssa.12807

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, 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. 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, 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, 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. 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, 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, 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, 363–80. 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. Odijk, S. Piperidis, M. Rosner, & D. Tapias (Eds.), Proceedings of the 7th international conference on language resources and evaluation (pp. 1168–1171). European Language Resources Association. http://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 & 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, 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, 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, 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. San Francisco: 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). 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, 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 de Méthodologie Sociologique, 127, 26-42. https://doi.org/10.1177/0759106315582199

Appendix A

English translation of the interviewing agent introduction (first and second conditions)

In the following, an interviewing agent based on artificial intelligence (AI) will ask you questions via video. You can play the videos as often as you like. To play the videos, simply click on the play button on the video.

You can give your responses either by selecting a response category or by entering them in your own words in a text field below the video.

If you do not have the opportunity to watch the videos at the moment, you can, of course, complete the web survey later.

English translation of the questions used in this study

Gender

You are ...

Response categories: 1 "male," 2 "female," and 3 "diverse"

Birthyear (transformed to age in years)

In which year are you born?

Please enter your year of birth: [open numeric field]

School education (original German response categories)

What is your highest general school-leaving qualification?

Response categories: 1 "(Noch) kein Schulabschluss," 2 "Abschluss einer Förderschule (Sonderschule, Hilfsschule)," 3 "Volks- oder Hauptschulabschluss bzw. Polytechnische, Oberschule der ehem. DDR mit Abschluss der 8. oder 9. Klasse," 4 "Mittlere Reife, Realschulabschluss, Fachoberschulreife oder Mittlerer Schulabschluss bzw. Polytechnische Oberschule der ehem. DDR mit Abschluss der 10. Klasse," 5 "Fachhochschulreife, 6 "Abitur, allgemeine oder fachgebundene Hochschulreife bzw. Erweiterte Oberschule der ehem. DDR mit Abschluss 12. Klasse," 7 "Anderer Schulabschluss, und zwar: [offenes Antwortfeld]"

Survey evaluation: interest

How interesting did you find it to answer the questions asked?

Response categories: 1 "very interesting" to 7 "not at all interesting"

Survey evaluation: difficulty

How easy or difficult did you find it to take part in the survey?

Response categories: 1 "very easy" to 7 "very difficult"

Survey evaluation: personal feeling

How personal did you find it to answer the questions asked?

Response categories: 1 "very personal" to 7 "not at all personal"

Survey evaluation: satisfaction
How much did you like the survey overall?
Response categories: “very much” to 7 “not at all”

Appendix B

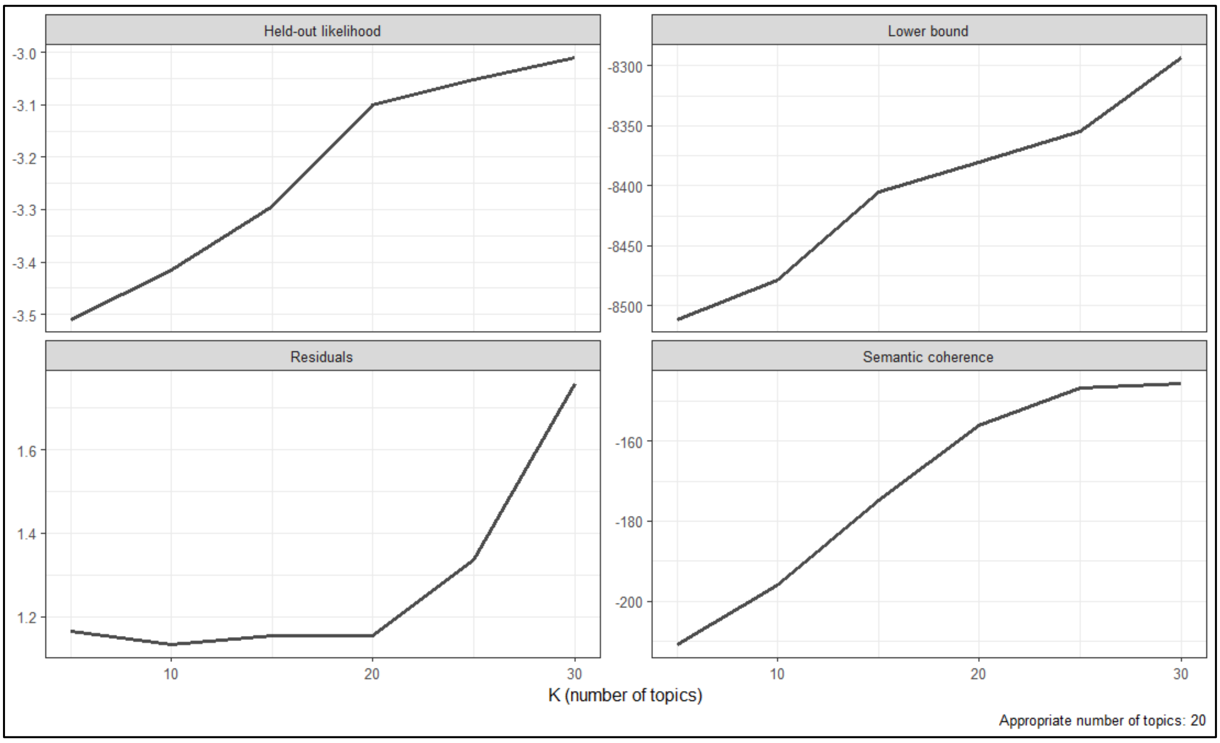

Figure B1. Diagnostic plots for the open narrative question on women’s role in the workplace

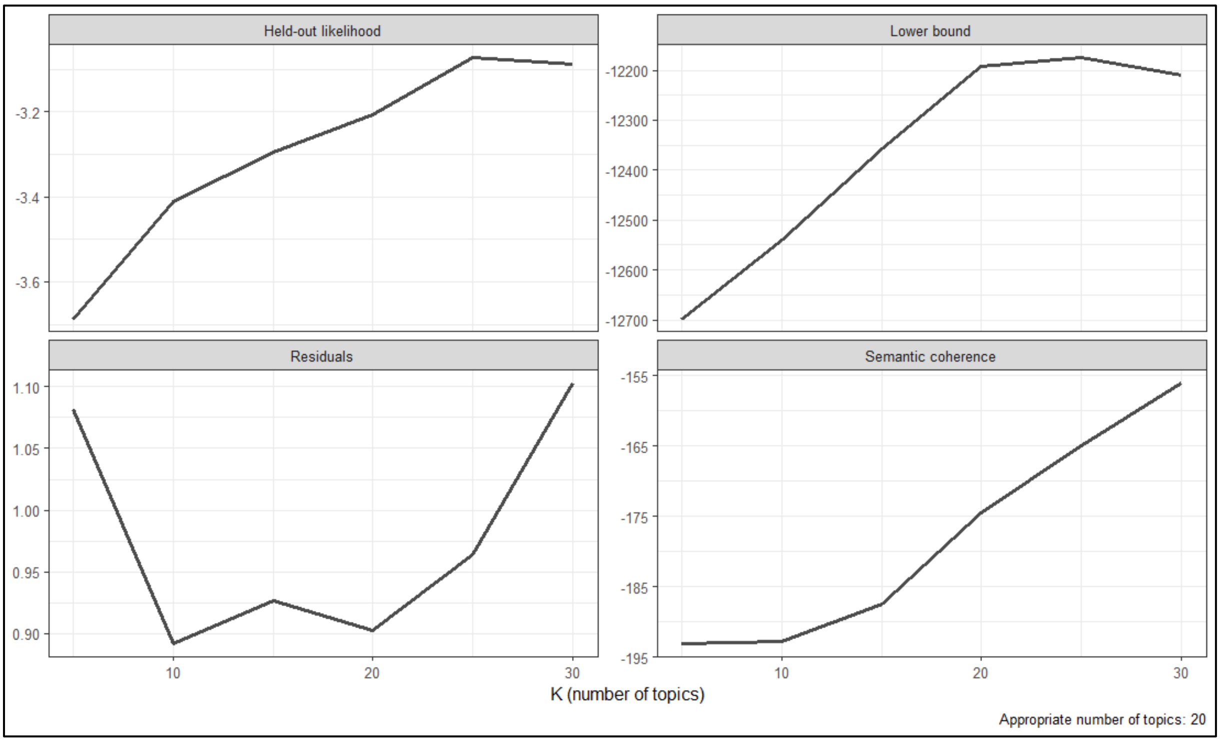

Figure B2. Diagnostic plots for the open narrative question on family relations

Appendix C

Table C1. Respondents' survey evaluations across the three conditions

Variables	Male interviewer condition	Female interviewer condition	Text control condition	Test statistics
Interest	5.5	5.5	5.7	$F(2,1150) = 2.13$, $p = 0.12$
Difficulty	1.7	1.6	1.6	$F(2,1148) = 1.65$, $p = 0.19$
Personal feeling	4.5	4.2	3.8	$F(2,1147) = 12.88$, $p < 0.01$
Satisfaction	5.7	5.7	5.9	$F(2,1146) = 2.55$, $p = 0.08$

Note. We report means for 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 ("not at all" to 7 "very much").