

# Innovating the collection of open-ended answers: The linguistic and content characteristics of written and oral answers to political attitude questions

Konstantin Gavras<sup>1</sup>  | Jan Karem Höhne<sup>2,3</sup>  | Annelies G. Blom<sup>1,4</sup>  | Harald Schoen<sup>1</sup>

<sup>1</sup>School of Social Sciences, University of Mannheim, Mannheim, Germany

<sup>2</sup>RECSM-Universitat Pompeu Fabra, Barcelona, Spain

<sup>3</sup>Faculty of Social Sciences, University of Duisburg-Essen, Duisburg, Germany

<sup>4</sup>University of Bergen, Bergen, Norway

## Correspondence

Konstantin Gavras, School of Social Sciences, University of Mannheim, Mannheim, Germany.

Email: [konstantin.gavras@uni-mannheim.de](mailto:konstantin.gavras@uni-mannheim.de)

## Funding information

Deutsche Forschungsgemeinschaft (DFG), Grant/Award Number: 139943784

## Abstract

The rapid increase in smartphone surveys and technological developments open novel opportunities for collecting survey answers. One of these opportunities is the use of open-ended questions with requests for oral instead of written answers, which may facilitate the answer process and result in more in-depth and unfiltered information. Whereas it is now possible to collect oral answers on smartphones, we still lack studies on the impact of this novel answer format on the characteristics of respondents' answers. In this study, we compare the linguistic and content characteristics of written versus oral answers to political attitude questions. For this purpose, we conducted an experiment in a smartphone survey (N = 2402) and randomly assigned respondents to an answer format (written or oral). Oral answers were collected via the open source 'SurveyVoice (SVoice)' tool, whereas written answers were typed in via the smartphone keypad. Applying length analysis, lexical structure analysis, sentiment analysis and structural topic models, our results reveal that written and oral answers differ substantially from each other in terms of lengths, structures, sentiments and topics. We find evidence that

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* published by John Wiley & Sons Ltd on behalf of Royal Statistical Society.

written answers are characterized by an intentional and conscious answering, whereas oral answers are characterized by an intuitive and spontaneous answering.

#### KEYWORDS

open-ended questions, political attitudes, sentiment analysis, smartphone surveys, structural topic modelling, text data, voice data

## 1 | INTRODUCTION

Attitude measurement is an important endeavour in social science research and many adjacent research fields. For instance, political attitudes measured in surveys help to explain and predict individual political behaviour and, in aggregated form, are linked to the stability of political systems (Achen, 1975; Campbell et al., 1960; Claassen, 2020; Torney-Purta & Valsiner, 2005). In the early 20th century, researchers frequently used open-ended questions in interviewer-based surveys to measure political attitudes. The interviewers recorded respondents' answers, although not by transcribing verbatim, but paraphrasing them (Shapiro, 1970; West & Blom, 2017). Over the years, due to cost and time restrictions, however, scholars interested in political attitudes switched to closed-ended questions as default. This technique enables researchers to easily measure political attitudes by obtaining numerical values on a large number of questions. However, closed-ended questions give rise to several types of response bias, such as acquiescence, extreme response styles and response order effects (see DeCastellarnau, 2018), and provide only limited insights into people's attitudes and reasoning.

In an attempt to overcome the shortcomings of closed-ended questions and to gather more in-depth and unfiltered information about respondents' political attitudes, researchers started to reconsider open-ended questions as a proper methodology (Revilla & Ochoa, 2016; Smyth et al., 2009; Zuell et al., 2015). The main reason is that only open-ended questions are capable of collecting in-depth information for fully understanding respondents' attitudes towards the objective under investigation (Eagly et al., 1994; Esses & Maio, 2002; Geer, 1988, 1991). So far, however, researchers have mainly focused on using open-ended questions with a request for written answers. This has some methodological drawbacks, because many respondents find it difficult to express their attitudes in a written format. It might also be burdensome to enter a written answer via a keypad; this particularly applies to open-ended questions that are answered via a smartphone with a virtual on-screen keypad shrinking the viewing space available for substantive content on the screen (Höhne, Cornesse, et al., 2020; Höhne, Schlosser, et al., 2020).

The rapid increase in smartphone use in web surveys (see, for instance, Gummer et al., 2019; Höhne, 2021; Peterson et al., 2017; Revilla & Ochoa, 2016), coupled with technological developments, has introduced new opportunities for collecting respondents' answers to open-ended questions. For instance, built-in microphones enable researchers to collect oral instead of written answers. Especially, smartphones facilitate the recording of oral answers to collect rich information about respondents' political attitudes by triggering an open narration (Gavras & Höhne, 2020; Revilla & Couper, 2019; Revilla et al., 2018). Respondents are able to express their attitudes with almost no further (technical) burden; they only need to press a recording button on the respective survey page and record their answer. Such recording tools resemble the voice messaging option of popular Instant-Messaging Services, such as WhatsApp and WeChat.

There are several reasons to expect that written and oral answers differ with respect to the precision and disclosure of information (see Schober et al., 2015). In addition, requests for written and oral answers may initiate different processes of attitude formation and thus result in different answers (Taber & Young, 2013). However, so far, the pertinent literature on measuring political attitudes is characterized by a substantial gap of knowledge when it comes to requests for written and oral answers to open-ended questions. To the best of our knowledge, there is no research on how answers to political attitude questions are affected by open-ended questions with requests for written and oral answers. We employ different quantitative text analysis methods to test our hypotheses. More specifically, we examine the number of words, assess the lexical structure, estimate sentiment scores and apply structural topic models (STMs). Applying these text analyses allows us to examine the linguistic and content characteristics of written and oral answers to political attitude questions.

## 2 | ON THE PSYCHOLOGY OF WRITTEN AND ORAL ANSWERS

Political attitudes can be defined as an ‘overall categorization of an [political] attitude object along an evaluative dimension’ (Zanna & Rempel, 1988, p. 319). Political psychologists have proposed ideal-typical models of how people form attitudes. According to the model of memory-based processing, respondents form attitudes towards an object on the spot when they are asked to evaluate the object by constructing an evaluation from the information they can retrieve from long-term memory (Zaller & Feldman, 1992). According to the on-line processing model, by contrast, respondents update an on-line tally in the moment when they are exposed to an attitude object. When asked for their attitude, they simply retrieve the on-line tally from long-term memory, without processing any detailed information about the object (Lodge et al., 1989; McGraw et al., 2003). Empirically, the two models appear to be at work with respondents being able to choose which processing model to use in a specific situation (see Kim & Garrett, 2012).

In the context of survey answers to political attitude questions, it also seems reasonable to assume that the answer format affects respondents’ attitude formation. According to Tourangeau et al. (2000), survey answers are based on a four-stage process: First, respondents must comprehend the question meaningfully (comprehension). Second, respondents must recall relevant information from memory (recall). Third, respondents must form a judgment based on the recalled information (judgment). Finally, respondents must communicate their mental judgment (response) in the answer format provided. These four stages occur concurrently, and respondents switch quickly and subconsciously back and forth between earlier and later stages (Tourangeau & Bradburn, 2010, pp. 316–317).

There is one important point that must be taken into consideration when applying this cognitive model to open-ended questions with requests for written and oral answers. In the request for written answers respondents can edit their answers (as a part of the response stage), whereas in the request for oral answers this is (usually) not possible. Even though written answers may allow respondents to recall more information because of their intentional and conscious character, they also allow to consider social desirability aspects, resulting in response bias (Kreuter et al., 2008). Requests for written answers may encourage a memory-based processing for intentionally and consciously forming political attitudes. Due to their intuitive and spontaneous character and their answer delivery process, oral answers cannot be easily edited and may rely on on-line tallies rather than on memory-based processing.

Also, differences between written and oral answers may occur due to the differences in opportunities for respondents to express their attitudes associated with the two answer formats. Open-ended questions with a request for a written answer have been frequently criticized because the attitudes expressed are confounded with respondents' ability to articulate themselves in a written form, particularly in cases of low literacy (Denscombe, 2008; Geer, 1988; Stanga & Sheffield, 1987). Requests for oral answers may, thus, offer an appealing way for researchers and respondents to collect open-ended answers, while minimizing biases due to written articulation and literacy skills.

In summary, requests for written answers differ from requests for oral answers on two main accounts. First, requests for written answers are characterized by activating different attitude formation processes than their oral counterparts. We expect that requests for written answers are more likely to trigger memory-based processing than requests for oral answers. Thus, written answers may be more intentional and conscious, whereas oral answers may be more intuitive and spontaneous. Second, the way of answer delivery (i.e. typing in text via a keypad vs. recording oral input via a built-in microphone) might also lead to differences between the two requests for an answer. Thus, we expect typing in an answer is more burdensome for respondents than simply recording an answer.

### 3 | HYPOTHESES

From the previously outlined theoretical background we derive five hypotheses that describe the potential impact of requests for written and oral answers on respondents' answer behaviour. Regarding our first hypothesis, written answers may trigger a more intentional and conscious attitude formation process. In contrast, oral answers may trigger a more intuitive and spontaneous attitude formation process. In addition, typing in answers on a smartphone keypad is expected to be more burdensome than recording answers. It also allows the editing of the answers during and after the typing. This should lead to shorter answers in terms of the number of words.

**Hypothesis 1** *Written answers result in a lower number of words than oral answers.*

Based on the same theoretical reasoning, we expect that written and oral answers differ with regard to their lexical structure (i.e. the variety of words used and the readability of the answers). Since answers in the request for written answers are more intentional and conscious, we expect that respondents select their words more carefully and forethought than respondents with a request for oral answers, which are more intuitive and spontaneous. Consequently, written answers are expected to be more lexically structured than oral answers; that is, written answers show a higher lexical richness, a higher lexical diversity and a lower level of readability (i.e. answers being more difficult to read) than their oral counterparts.

**Hypothesis 2** *Written answers result in lexically more structured answers than oral answers.*

As explained above, respondents tend to take social norms and values into account when answering survey questions and edit their answer in accordance with perceived social (un)desirability. Both the more intentional and conscious answering as well as the editing possibilities in written answers provide greater scope for social desirability bias than the intuitive and spontaneous answering in oral answers. This may lead to fewer distinctly positive or distinctly negative sentiments in written than in oral answers. Consequently, we expect written answers to be less (positively and negatively) extreme than oral answers.

**Hypothesis 3** *Written answers result in less (positively and negatively) extreme answers than oral answers.*

Due to the intentional and conscious memory-based processing, we expect written answers to consist of a larger number of topics than oral answers that are based on an intuitive and spontaneous on-line processing. However, written answers are also more burdensome, potentially preventing respondents from writing down all relevant aspects that they have in mind. Even though it remains unclear whether and to what extent the increased burden of entering written answers counterbalances the effect of intentionality and conscientiousness, we expect no differences in the number of topics mentioned by respondents between written and oral answers.

**Hypothesis 4** *Written answers do not result in a different number of topics than oral answers.*

As stated in Hypothesis 4, we assume the response burden associated with written answers to counterbalance their intentional and conscious character, which, in turn, results in a similar number of topics mentioned in written and oral answers. However, this does not necessarily apply to the range of topics mentioned by respondents. Given the memory-based processing and the information retrieval from long-term memory associated with written answers it is likely that under this condition respondents draw on a broader range of topics than when providing oral answers to open-ended questions. In addition, we do not expect that the burden of entering written answers affect the range of topics and, thus, we assume that requests for written answers produce a broader range of topics than requests for oral answers.

**Hypothesis 5** *Written answers result in a broader range of topics than oral answers.*

## 4 | METHOD

### 4.1 | Experimental design

We conducted an experiment in a smartphone survey and randomly assigned respondents to one of two request conditions. The first experimental group ( $n = 1694$ ) received six open-ended questions with a request for written answers (written condition). The second experimental group ( $n = 1679$ ) received the same six open-ended questions but with a request for oral answers (oral condition). We additionally randomized respondents to different conditions (e.g. posing motivational messages) within the request conditions (i.e. written or oral). However, these sub-conditions are not subject of this article.

### 4.2 | Study procedure

Data collection was conducted by the survey company Forsa (<http://www.forsa.de>) and took place in Germany in December 2019 and January 2020. Forsa drew a quota sample from their nonprobability access panel based on age, gender, education and region (East and West Germany). The quota plan was designed to match the German population on these demographic characteristics. The quotas were calculated based on the German Mikrozensus, Germany's official statistics, as population benchmark (FDZ, 2020).

The email invitation to the web survey included information on the estimated duration of the survey (about 15 min), the respective device type (i.e. smartphone) to use for survey completion,

and a link to the survey. The first survey page outlined the general topic and procedure of the survey, and included a statement stating that the study adheres to the EU, national and federal data protection laws and regulations. Furthermore, respondents were explicitly asked whether they are willing to provide written and oral answers. Only respondents agreeing to this screening question were forwarded to our survey.

After the screening process at the beginning of the survey, respondents were randomly assigned to either request condition. Respondents received modest financial compensation from Forsa, which was proportional to the length of the survey and credited to their study account after finishing the entire survey. We also collected User-Agent-Strings using the open-source tool ‘Embedded Client Side Paradata (ECSP)’ (Schlosser & Höhne, 2018, 2020) to identify device properties, such as device type and operating system, throughout the survey. This was done to ensure smartphone participation.

### 4.3 | Sample

Forsa sent out 19,754 survey invitations to opt-in panelists to participate in the smartphone survey, of which 2538 panelists were screened out because the quotas were already achieved, they declined to consent to give text and voice answers, or because they tried to access the survey with another device than a smartphone. A total of 3373 panelists started the survey, but 971 of them dropped out before finishing the entire survey. This leaves us with 2402 respondents available for statistical analysis.

Of these 2402 respondents, 1477 participated in the written condition and 925 in the oral condition. This corresponds to a participation rate of about 12% among all invitees. The different sample sizes across the two conditions stem from differential dropout rates during the survey with about 13% dropping out in the written condition and about 45% dropping out in the oral condition. In addition, there is between 25% and 28% item non-response in the oral condition. The results of logistic regressions on item non-response indicate no significant differences, except for voice message usage (see Appendix A in the Supplementary Materials). We did not estimate logistic regressions on item non-response in the written condition, because of low variance in the dependent variable, as item non-response is less than 5%. To make sure that the dropouts did not bias the experimental assignment, we compared the sample compositions of the written and oral conditions. Table 1 reports the sample composition.

In addition, we estimated the effect of differential dropouts across the written and oral conditions using logistic regressions. The results in Figure 1 indicate no significant differences, except for Greens and AfD voters between both conditions.

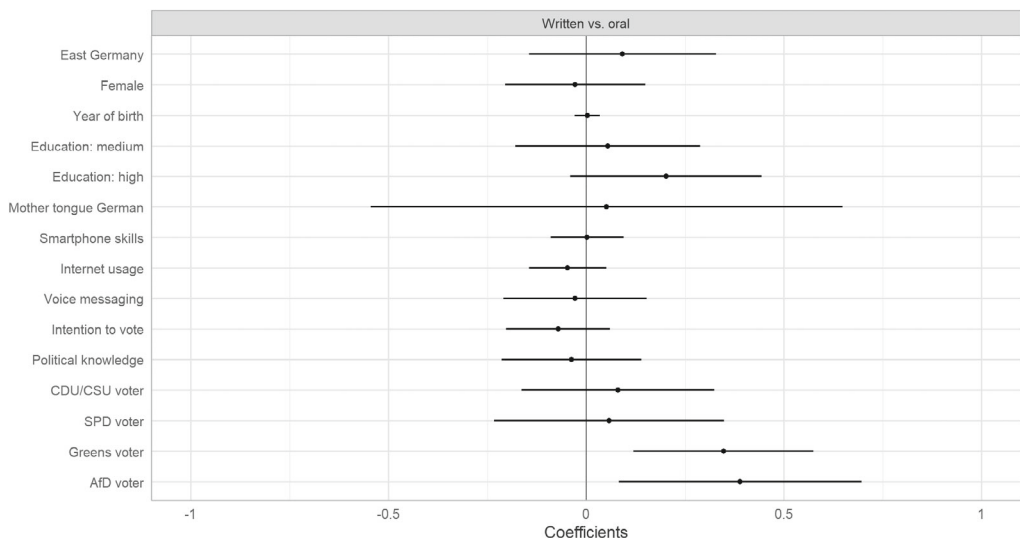
### 4.4 | Political attitude questions

We asked six open-ended questions: the most important political problem in Germany, attitudes towards the German chancellor, and attitudes towards four German political parties (i.e. CDU/CSU [Christian Democratic Union/Christian Social Union], SPD [Social Democratic Party], Greens [Alliance 90/The Greens], and AfD [Alternative for Germany]). The question wording was adopted from established social surveys in Germany. All questions were in German, which was the mother tongue of about 98% of the respondents (see Appendix B in the Supplementary Materials for English translations of all open-ended questions). To improve comparability between

**TABLE 1** Sample composition by request conditions (written and oral)

Socio-demographic characteristics	Written condition	Oral condition
Year of birth	1975–1979	1975–1979
Female	0.49	0.49
Low education level	0.22	0.24
Medium education level	0.33	0.33
High education level	0.45	0.42
Region (West Germany)	0.84	0.85
German mother tongue	0.98	0.98
Smartphone skills	4.67	4.67
Internet usage	2.39	2.42
Voice message usage	0.63	0.64
Intention to vote	3.77	3.78
Political knowledge	0.49	0.49
CDUCSU voter	0.20	0.21
SPD voter	0.11	0.13
Greens voter	0.32	0.26
AfD voter	0.12	0.09
N	1,477	907

*Note:* Except for year of birth, smartphone skills, intention to vote and internet usage, the table reports proportions of the respective characteristics. Year of birth has been surveyed in 5-year intervals (we compared medians). Smartphone skills has been surveyed on a decrementally aligned and end-labelled scale with seven points (we compared means). Internet usage and intention to vote have been surveyed on a decrementally aligned and fully labelled scale with five points (we compared means).

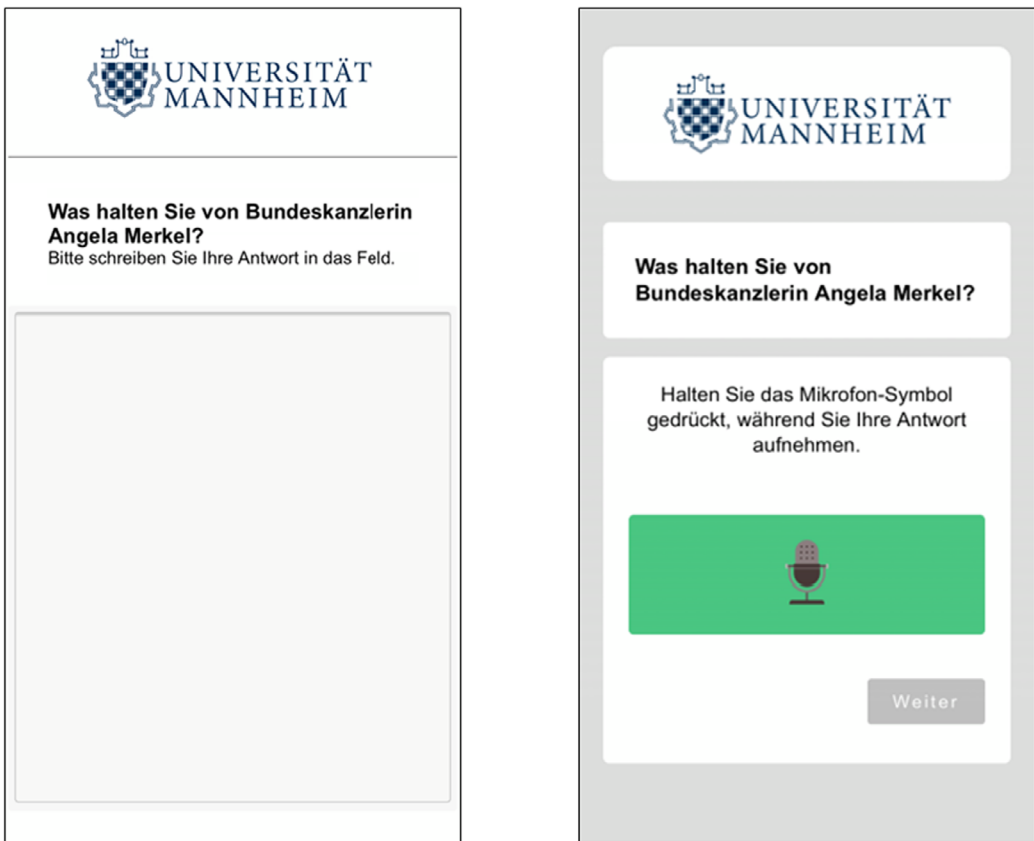
**FIGURE 1** Logistic regressions for comparing the written (written = 1) and oral conditions in terms of differential dropout

*Note:* 95% confidence intervals. Sample size is N = 2354 respondents.

the written and oral requests for an answer, we employed an optimized survey layout, preventing horizontal scrolling. We presented only one question per page (single question presentation). Respondents were informed that they could skip questions but were not provided with an explicit non-substantive answer option, such as “don’t know” or “no opinion”.

For recording respondents’ oral answers, we implemented a voice recording tool in the browser-based smartphone survey. For this purpose, we used the open source ‘SurveyVoice (SVoice)’ tool developed by Höhne et al. (2021) that records respondents’ oral answers via the built-in microphone of smartphones, irrespective of the operating system (e.g. Android and iOS). Figure 2 shows the design of the open-ended questions with requests for written and oral answers respectively.

The open-ended questions with requests for written and oral answers were kept exactly the same across the experimental conditions. The questions were preceded by short answer instructions that were tailored to the respective request for an answer (see Appendix B in the Supplementary Materials for English translations of the instructions).



**FIGURE 2** Example of the open-ended question on the German chancellor (Angela Merkel) with requests for written (on the left side) and oral answers (on the right side)

*Note:* The ‘Next’ (‘Weiter’) button of the open-ended question with a request for a written answer is not displayed because of space limitations. We did not limit the number of characters in the open answer box or the recording time in the ‘SurveyVoice (SVoice)’ tool. [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



## 4.5 | Analytical strategy

Respondents' oral answers had to be transcribed (into written answers) before statistical analyses. For this purpose, we used Google's Transcribe API 'Speech-to-Text' that automatically transcribes audio files into text (Google, 2020). As shown by Proksch et al. (2019), the performance of the API does not substantially differ from human transcription. They found an average cosine similarity of  $r > 0.9$  between Google-transcribed and human-transcribed political speeches in German.

In the following, we outline the analytical strategy associated with our research hypotheses. In our analyses, we exclusively focus on given answers by respondents—not considering dropouts and item non-responses—and compare them with respect to length, lexical structure, sentiment and topic.

*Hypothesis 1:* In order to test our first hypothesis, we initially count the number of words between the two request conditions. The reason for using words instead of characters is that (strong) accents and dialects can affect the number of characters (e.g. omitting the final letters of a word) when automatically transcribing oral answers. This would decrease the accuracy of the answer length. We use the `quanteda` package in R and count the number of 'tokens' (or words) for each answer of respondents. We then calculate the mean number of tokens and finally conduct Welch two-sample  $t$  tests with unequal variances to test for significant differences in the number of words between written and oral answers.

*Hypothesis 2:* Following Benjamin (2012), we investigate the lexical richness (defined as the number of different words when taking the total number of words into account), lexical diversity (defined as the ratio of unique words divided by the number of total words) and readability (defined as the complexity of word and sentence structure) to infer the level of lexical structure in written and oral answers. In order to ease the interpretation and to prevent confounding results, we analyse the three aspects of lexical structure separately (John & Paul, 2002). We compare the following three measures and conduct Welch two-sample  $t$  tests with unequal variances:

1. Yule's K: This indicator measures the lexical richness of respondents' answers. K ranges from 0 to  $\infty$  with higher values indicating a lower lexical richness (Yule, 1944).
2. Type-token ratio (TTR): This indicator measures lexical diversity of respondents' answers. TTR ranges from 0 to 1 with higher values indicating a lower level of lexical diversity (Templin, 1957).
3. Flesch reading ease (FRE): This indicator measures the readability of respondents' answers. FRE ranges from  $-\infty$  to 121.22 with higher values indicating an easier readability (Flesch, 1948).

*Hypothesis 3:* We conduct sentiment analyses to investigate the level of extremity of respondents' answers (Pang & Lee, 2008). For this purpose, we use the German sentiment vocabulary SentiWS (Remus et al., 2010) in which words are assigned scores ranging from -1 (very negative) to 1 (very positive). The scores indicate the strength of the sentiment-afflicted words. We estimate the extremity of answers using the following formula (Lowe et al., 2011):

$$S = \log \frac{\text{pos} + 0.001}{|\text{neg}| + 0.001}$$

where `pos` denotes the weighted sum of positive sentiment words and `|neg|` denotes the absolute weighted sum of negative sentiment words. We add a small penalty (0.001) to prevent calculation

problems when dividing by zero. Finally, we compare the logged mean scores between the written and oral conditions and conduct Student two-sample  $t$  tests with equal variances, except for the question on the Greens where we apply a Welch two-sample  $t$  test with unequal variances.

*Hypothesis 4:* We employ structural topic models (STMs, Roberts et al., 2014) using the `stm` package in R to determine the number of topics mentioned by respondents. STMs facilitate the examination of the impact of experimental treatments on the content of, for instance, written and oral answers and, thus, they are suited to analyse differences in topics (Roberts et al., 2014, p. 3). Pietsch and Lessmann (2018) as well as Nelson et al. (2021) were able to show that automated coding approaches are important complements to traditional hand-coding approaches. When aiming to uncover overall topic distributions and exploring topic occurrence, STMs are considered a viable alternative to hand-coding of answers to open-ended questions. We only include words that were mentioned in at least 10 answers and remove all stop words. As there is no established threshold for removing seldom occurring words, we conducted robustness checks with words appearing in at least 5 and 20 answers. The main conclusions did not change. For each answer, we count the number of topics to which at least 10% of the individual answers are attributed. Again, there is no established threshold to identify topic assignment. Therefore, we conducted robustness checks with 5% and 20% as lower shares for topic assignment. The main conclusions did not change. We use the following diagnostic criteria to determine the appropriate number of topics for each question (Roberts et al., 2019; Wallach et al., 2009): high held-out likelihood, low residuals, medium semantic coherence and low level of lower bound (see Appendix C in the Supplementary Materials for diagnostic plots). Following these criteria, we calculate the number of topics for both request conditions and conduct Student two-sample  $t$  tests with equal variances. Finally, we descriptively compare the topics mentioned between the two request conditions and report the overlap between written and oral answers.

*Hypothesis 5:* We calculate the effective number of topics (ENTs; see Laakso & Taagepera, 1979) determining the range of topics in both request conditions (written and oral):

$$\text{ENT} = \frac{1}{\sum_{i=1}^n t_i^2},$$

where  $n$  denotes the number of topics determined by the diagnostics of the STM and  $t_i^2$  denotes the squared proportion of each topic within the total number of topics. The ENT score combines the mean and the variance of the topic distribution within one single measure so that it is difficult to test for significant differences between both request conditions.

As robustness checks, we additionally conducted multi-level regressions (with questions nested in respondents) for the analyses on Hypotheses 1–3. The main conclusions did not change.

## 5 | RESULTS

### **Hypothesis 1** *Length of the answers.*

With respect to our first hypothesis, we now compare respondents' average answer length in terms of number of words between requests for written and oral answers. Table 2 reports the statistical results. Concerning the length of the answers, we find that oral answers are significantly longer than their written counterparts. This applies to all political attitude questions, except for the one on the most important political problem. In some cases, the oral answers are up to 40%

TABLE 2 Average answer length of written and oral answers

Questions	Written answers	Oral answers	Difference
Problem	27.3	31.2	-3.9
Chancellor	27.9	36.0	-8.1***
CDU/CSU	22.8	31.8	-9.0***
SPD	22.3	33.0	-10.7***
Greens	23.4	39.1	-15.7***
AfD	21.9	35.4	-13.5***

Note: \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . Written condition:  $n = 1414$  to  $1453$ . Oral condition:  $n = 667$  to  $695$ . We used the number of words for determining answer length. Difference: written answers minus oral answers.

TABLE 3 Lexical structure (lexical richness, lexical diversity and readability) of written and oral answers

Questions	Yule's K			Type-token ratio (TTR)			Flesch reading ease (FRE)		
	Written answers	Oral answers	Difference	Written answers	Oral answers	Difference	Written answers	Oral answers	Difference
Problem	1413.4	640.7	772.7***	0.93	0.90	0.03***	-9.4	-6.3	-3.1
Chancellor	793.5	327.5	466.0***	0.94	0.90	0.04***	49.7	30.2	19.5***
CDU/CSU	1018.9	395.2	623.7***	0.94	0.90	0.04***	49.0	39.2	9.8***
SPD	1024.7	337.0	687.7***	0.94	0.90	0.04***	50.4	36.2	14.2***
Greens	906.1	356.3	549.8***	0.94	0.89	0.05***	43.3	24.2	19.1***
AfD	1353.7	408.9	944.8***	0.95	0.90	0.05***	47.5	32.7	14.8***

Note: \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . Written condition:  $n = 1414$  to  $1453$ . Oral condition:  $n = 667$  to  $695$ . Yule's K measures lexical richness, Type-Token Ratio (TTR) measures lexical diversity, and Flesch Reading Ease (FRE) measures readability. Difference: written answers minus oral answers.

longer than the written ones. Thus, these results provide strong supporting evidence for our first hypothesis.

### Hypothesis 2 Lexical structure of the answers.

Lexical structure is a multi-dimensional concept, including lexical richness (measured by Yule's K), lexical diversity (measured by TTR) and readability (measured by FRE). For this reason, we evaluate the three measures separately. Remember that for Yule's K and TTR higher values indicate lower lexical richness and lower lexical diversity, respectively, and for FRE higher values indicate easier readability.

As Table 3 shows, written and oral answers are characterized by different lexical structures. This systematically applies to all six political attitude questions. Considering the measures of lexical richness (Yule's K) we find that oral answers, compared to written answers, are characterized by a significantly larger variety of words, indicating a richer set of vocabulary. In addition, the lexical diversity measure TTR shows a more diverse set of vocabulary for oral answers. Thus, the results indicate that lexical richness and diversity are higher for oral than written answers, which contradicts our hypothesis. The results on readability draw a similar picture. The FRE is significantly lower for oral than written answers. This applies to five of six questions, indicating that

TABLE 4 Sentiment ratio of written and oral answers

Questions	Written answers	Oral answers	Difference
Problem	-1.23	-1.78	0.55***
Chancellor	1.69	2.11	-0.42
CDU/CSU	0.02	1.01	-0.99***
SPD	-0.80	-0.32	-0.48
Greens	0.66	1.47	-0.81***
AfD	-2.54	-1.09	-1.45***

Note: \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . Written condition:  $n = 1414$  to  $1453$ . Oral condition:  $n = 667$  to  $695$ . Difference: written answers minus oral answers.

oral answers, compared to written answers, are also more difficult to read. Overall, the results provide no evidence for our second hypothesis. We discuss this point further in the discussion and conclusion section.

### Hypothesis 3 *Sentiments of the answers.*

Our results on the sentiments and the (positive and negative) extremity of written and oral answers are mixed. Table 4 reports the statistical results. Only for three of six questions (i.e. problem, CDU/CSU and Greens) we find significantly more extreme answers for oral than written answers. Oral answers to the question on the most important problem are more negative than written answers. However, oral answers to the questions on the CDU/CSU and Greens are more positive than their written counterparts. In contrast to our expectation, for the question on the AfD, we observe a reversed relation with written answers being significantly more negative than oral answers. In the case of the question on the German chancellor, we find that oral answers are more positive and, in the case of the question on the SPD, we find that written answers are more negative. However, the latter two questions do not differ significantly across the request conditions. Thus, we only find partial evidence for our third hypothesis.

Gavras and Höhne (2020) found that sentiment scores of written and oral answers to open-ended questions on attitudes towards German political parties can be used to predict voting behaviour. In order to provide further descriptive evidence, we estimated the correlation matrix of the sentiment scores (see Appendix D in the Supplementary Materials). The results indicate that the sentiment scores are moderately correlated. This applies to both the written and oral conditions and the overall correlation structure does not vary substantially between conditions.

### Hypothesis 4 *Topics of the Answers.*

With respect to our fourth hypothesis, we now investigate the average number of topics. Following the diagnostic criteria, about 26 topics per question can be considered appropriate; min = 20 (SPD) and max = 30 (chancellor and Greens). Table 5 reports the statistical results. In contrast to our expectation, written and oral answers result in a different number of topics. For five of six questions (i.e. problem, CDU/CSU, SPD, Greens and AfD) we find that respondents mentioned significantly more topics in the oral than in the written condition. Only for the question on the chancellor, we find the same number of topics in both request conditions. Overall, these findings do not provide supporting evidence for our fourth hypothesis.

TABLE 5 Average number of topics in written and oral answers

Questions	Written answers	Oral answers	Difference
Problem	1.70	2.07	-0.37***
Chancellor	1.11	1.11	0.00
CDU/CSU	1.06	1.61	-0.55***
SPD	1.64	1.81	-0.17***
Greens	0.80	1.27	-0.47***
AfD	1.43	1.61	-0.18***

Note: \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . Written condition:  $n = 1279$  to  $1371$ . Oral condition:  $n = 605$  to  $664$ . We only considered topics that encompass at least 10% of the individual answers when calculating the average number of topics. Difference: written answers minus oral answers.

For a better understanding of these results, we now investigate the content of the topics mentioned by respondents. For analytical purposes, we restrict this analysis to the 10 most frequently mentioned topics in written and oral answers respectively. Table 6 shows that depending on the request for an answer respondents mention different topics. On average, the overlap between the two request conditions is lower than 50% and runs from 30% (AfD) to 70% (chancellor). See Appendix E in the Supplementary Materials for the prevalence of the topics.

Taking a closer look at the topics, in the written answers topics refer to rather historic events and specific policies. For instance, Hartz IV (SPD) and refugees (chancellor). In the oral answers, in contrast, the topics refer to rather recent events and broader aspects, such as self-absorption (SPD) and general rejection in terms of being terrible and horrible (AfD). Interestingly, in the oral answers, respondents articulate their question answering process, expressing how they came up with their answers, indicating a kind of think-aloud (chancellor, CDU/CSU, SPD, Greens and AfD). As robustness check, we calculated the average number of topics and the effective number of topics when excluding the 'question answering process' topic. Overall, the main conclusion did not change. The topics mentioned most frequently in both request conditions refer to rather general aspects, such as inequality (problem) and long government (chancellor). Overall, these results provide empirical evidence that respondents refer to different topics depending on the respective request for an answer.

#### **Hypothesis 5** *Range of Topics in the Answers.*

After considering the number of topics, we now turn to the range of topics between written and oral answers. As shown in Table 7 and in line with our fifth hypothesis, the ENTs is consistently higher for written than for oral answers. The only exception is the question on the chancellor with slightly higher ENTs for oral answers. Thus, the results provide supporting evidence for our fifth hypothesis.

## **6 | DISCUSSION AND CONCLUSION**

The aim of this study was to investigate the linguistic and content characteristics of written and oral answers to open-ended questions on political attitudes. We investigated answers to six political attitude questions: The most important problem in Germany, attitudes towards the German chancellor, and attitudes towards four German political parties. For this purpose, we conducted

TABLE 6 Topics of written and oral answers

Questions	Written answers	Oral answers	Overlap (%)
Problem	Poverty, social change, immigration, <b>political parties, inequality, question repetition</b> , dissatisfaction, living conditions, energy transition, <b>political extremism</b>	<b>Question repetition</b> , taxes, <b>political extremism</b> , state capacity to act, <b>political parties</b> , egoism, political decisions, poverty, <b>inequality</b> , credibility	40
Chancellor	<b>Good chancellor, long government, achievements</b> , refugees, <b>need for change, good job</b> , successor, <b>competency, taking decisions</b> , political issues	<b>Competency</b> , question answering process, <b>need for change, achievements, long government</b> , voting, clear line, <b>good chancellor, good job, taking decisions</b>	70
CDU/CSU	<b>Good ideas, party leader, outdated</b> , conservative, <b>sister parties</b> , policies, interest representation, people's party, big money, <b>differences to other parties</b> ,	<b>Sister parties</b> , question answering process, <b>good ideas</b> , cheap talk, <b>outdated</b> , Bavarian government, <b>party leader</b> , eligible, centrist, <b>differences to other parties</b>	50
SPD	<b>Always voted, ineligible, good ideas, lost credibility</b> , good concepts, Hartz IV, political profile, reflect own concepts, lack of direction, party leadership	<b>Good ideas</b> , betrayal of simple man, <b>always voted</b> , difficult times, <b>ineligible</b> , differences to other parties, question answering process, self-absorbed, <b>lost credibility</b> , youth organization leader	40
Greens	<b>Credible party, good ideas, important topics, only alternative</b> , government, party leader, climate change, well-developed, dangerous, high polling numbers	Likeable, cars, question answering process, climate politics, <b>important topics, credible party, good ideas</b> , Grand Coalition, unrealistic demands, <b>only alternative</b>	40
AfD	<b>Ineligible</b> , right-wing extremism, <b>party ban</b> , fascist, take demands seriously, good ideas, <b>Grand Coalition</b> , ignoring history, appealing to unsatisfied voters, populist thinking	Question answering process, out of question, <b>ineligible</b> , terrible, necessary, <b>party ban</b> , horrible, Germany, <b>Grand Coalition</b> , free speech	30

Note: Written condition: n = 1279 to 1371. Oral condition: n = 605 to 664. We labelled the 10 most frequently occurring topics in both request conditions. The labels are ordered by the frequency of each topic. The topics in bold indicate topics being mentioned in both request conditions. Overlap: proportion of topics mentioned in both request conditions.

an experiment in a smartphone survey and randomly assigned respondents to a request condition (i.e. written and oral). The results reveal (substantial) differences between written and oral answers with respect to length, lexical structure, sentiments and topics. Overall, the results indicate that requests for written and oral answers may initiate different attitude formation processes. In addition, they seem to differ in terms of response burden.

Our results show that the request for written answers initiates a more intentional and conscious memory-based processing, which manifests itself in shorter answers. In addition,

TABLE 7 Effective number of topics (ENTs) in written and oral answers

Questions	Written answers	Oral answers	Difference
Problem	23.6	19.5	4.1
Chancellor	26.7	27.0	-0.3
CDU/CSU	22.2	20.4	1.8
SPD	18.5	17.8	0.7
Greens	28.1	25.0	3.1
AfD	23.7	21.3	2.4

Note: Written condition:  $n = 1279$  to  $1371$ . Oral condition:  $n = 605$  to  $664$ . The ENT is calculated from the estimated shares of the topics in the STM. Due to the nature of the analytical method, it is not possible to test for statistically significant differences. Difference: written answers minus oral answers.

respondents mention a broader range of topics referring to rather historic events and specific policies in written answers. This is also in line with our previously stated expectation that respondents recall (specific) information from long-term memory. In the request for oral answers, in contrast, respondents rather appear to form their answers more intuitively and spontaneously, relying on on-line tallies. Our results show that the answers provided by the respondents are longer in terms of words. Furthermore, respondents refer to rather broad aspects and recent events.

The length differences between written and oral answers do not only point to different forms of processing, but also to differences in response burden. Overall, written answers appear to be more burdensome than oral answers. Accordingly, differences in the answer delivery process may contribute to the length differences. More specifically, entering written answers via a virtual on-screen keypad of smartphones is more intricate than recording a voice answer by simply pressing a recording button. However, at this point, we cannot determine the impact of this feature due to a lack of suitable data. We therefore propose to employ evaluative questions in future smartphone surveys with requests for written and oral answers to shed light on the overall burden of both types.

The results on lexical structure contradict our hypothesis. One reason might be that the measures used are sensitive to the length of texts (Koizumi, 2012). Written answers are significantly shorter than oral answers so that the results on lexical richness and diversity might be due to methodological artefacts. In addition, the readability scores (FRE) are actually developed for texts with correct punctuation, which is often not provided by respondents negatively affecting the results (Flesch, 1948). Therefore, we recommend employing (or developing) additional methods to investigate differences in the lexical structure of written and oral answers to open-ended questions in future studies.

There is also some evidence that written and oral answers differ with respect to their proneness to social desirability bias. Written answers comprise somewhat fewer (positive and negative) extreme sentiments than their oral counterparts. One possible explanation is that requests for written answers facilitate respondents to consider social norms and values. Before the final answer submission (by clicking the 'Next' button), they can edit written parts that seem to be overly positive or negative so that they comply with socially acceptable points of views. Oral answers, in contrast, do not facilitate respondents to revise specific parts to account for social desirability. For instance, the 'SurveyVoice (SVoice)' tool (Höhne et al., 2021) allows respondents to delete their entire oral answer, but it does not facilitate re-recording individual parts. In this study, however, we did not ask overly sensitive questions, such as questions on sexual practices,

extremist attitudes or diseases. In our opinion, it would be worthwhile to compare the sentiments of written and oral answers to open-ended questions on sensitive topics. If our findings can be replicated, open-ended questions with requests for oral answers might be a promising way to decrease social desirability bias.

In contrast to our expectation, we found significant differences in the number of topics between written and oral answers. For five of six questions, oral answers resulted in a larger number of topics than written answers. This finding suggests that the response burden of entering written answers does not counterbalance the intentional and conscious memory-based processing associated with written answers but exceeds it. More specifically, open-ended questions with requests for written answers seem to prevent respondents from mentioning all relevant aspects that they may have in mind. In our opinion, further research is necessary to disentangle the relation between memory-based processing and response burden when it comes to written answers.

This study has some limitations that hint at avenues for future research. First, we conducted our experiment in a nonprobability access panel with respondents that have a relatively high survey experience. Even though we drew a quota sample that matches the German population on age, gender, education and region (East and West Germany), the sample may limit the generalizability of our findings. It may thus be worthwhile to compare our results to open-ended questions with written and oral answers in a probability-based sample. Second, we did not randomize the order of the open-ended questions with requests for written and oral answers. We recommend that future research randomizes the question order to prevent any order effects. Third, future research may investigate the measurement quality of written and oral answers by using respondents' predicted sentiment scores to determine the association between these scores and criterion measures. Doing so, one would be able to estimate the criterion validity of open-ended questions with requests for written and oral answers (see Gavras & Höhne, 2020). Unfortunately, this analysis was beyond the scope of this paper. Fourth, there is a chance that third parties have listened to what respondents say, which, in turn, may have affected their oral answers. To put it differently, respondents may have addressed topics that potential over-hearers would agree with, introducing third party effects (see, for instance, Smith, 1997). We therefore recommend that future studies collect information on the presence of third parties during question answering by, for instance, employing self-reports on the survey environment. Fifth, in this study, we rely on automated content analysis of the answers to compare the content provided in the requests for written and oral answers. Future studies might make use of human coding with pre-defined schemes to compare it with the approach used in this study and to gain additional insights from the given answers. Finally, some research suggests that, for instance, the level of respondents' political interest and knowledge moderates the information processing applied when forming political attitudes (Zaller & Feldman, 1992). Therefore, it may be worthwhile to investigate information processing between respondents of different levels of political interest and knowledge in future research.

This study illustrates that open-ended questions with requests for oral answers represent a promising new way of measuring respondents' political attitudes. However, whereas the technology is finally at a stage that allows both the collection and analysis of oral answers, we still face two main challenges that need to be addressed in future research. First of all, requests for oral answers result in comparatively high dropout rates (45% in the oral condition compared to 13% in the written condition) and levels of item non-response (25% to 28% in the oral condition compared to 2% to 4% in the written condition). This is in line with findings reported by Revilla and Couper (2019) indicating that our study is no singular instance. Such an amount of missing data may impact survey outcomes (both from open-ended questions with requests for oral answers





and following questions) and decrease the generalizability of the results. In addition, it appears inevitable to take the amount of missing data into account when designing studies and determining appropriate sample sizes. Otherwise, this may have serious consequences for data utility. We highly recommend (experimentally) investigating the reasons for dropout and item non-response by, for instance, looking at the association with respondents' partisanship, political ideology and privacy concerns. Leaving aside gradual changes in day-to-day usage of technologies, potential ways for the mitigation of dropout and item non-response might also be an increase in incentives for survey participation or to let respondents choose the format in which they would like to answer (i.e. written or oral). One should note, however, that both studies, Revilla et al. (2020) and ours, were conducted in existing panels and panel members were used to the written answer format. The reservations that panel members seem to hold against voice interviewing may be a panel conditioning effect and disappear when working with a newly recruited sample. Second, written and oral answers lead to different conclusions about respondents' political attitudes. As concluded by Schober et al. (2015), written answers tell a different story than oral answers, which, in turn, may change the implications that we can infer for political actors and policies. The challenge for political scientists and survey researchers is now to discover which of the stories is more valid.

## ACKNOWLEDGEMENTS

The authors thank Daniel Qureshi (University of Frankfurt) and Stephan Schlosser (University of Göttingen) for their support during data collection. The data collection was funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) through the Collaborative Research Center 884 'Political Economy of Reforms' (SFB 884; project ID: 139943784). We thank the Board of the SFB 884 for supporting our research. Open access funding was enabled and organized by ProjektDEAL.

## ORCID

Konstantin Gavras  <https://orcid.org/0000-0002-9222-0101>

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

Annelies G. Blom  <https://orcid.org/0000-0003-0377-301X>

## REFERENCES

- Achen, C.H. (1975) Mass political attitudes and the survey response. *American Political Science Review*, 69(4), 1218–1231.
- Benjamin, R.G. (2012) Reconstructing readability: recent developments and recommendations in the analysis of text difficulty. *Educational Psychology Review*, 24(1), 63–88.
- Campbell, A., Converse, P.E., Miller, W.E. & Stokes, D.E. (1960) *The American voter*. Chicago, [Ill.], London: University of Chicago Press.
- Claassen, C. (2020) Does public support help democracy survive? *American Journal of Political Science*, 64(1), 118–134.
- DeCastellarnau, A. (2018) A classification of response scale characteristics that affect data quality: a literature review. *Quality & Quantity*, 52(4), 1523–1559.
- Denscombe, M. (2008) The length of responses to open-ended questions. *Social Science Computer Review*, 26(3), 359–368.
- Eagly, A.H., Mladinic, A. & Otto, S. (1994) Cognitive and affective bases of attitudes toward social groups and social policies. *Journal of Experimental Social Psychology*, 30(2), 113–137.
- Esses, V.M. & Maio, G.R. (2002) Expanding the assessment of attitude components and structure: the benefits of open-ended measures. *European Review of Social Psychology*, 12(1), 71–101.
- FDZ (2020) Mikrozensus 2018, On-Site, Version 0. Available from: <https://doi.org/10.21242/12211.2018.00.00.1.1.0>
- Flesch, R. (1948) A new readability yardstick. *Journal of Applied Psychology*, 32(3), 221–233.

- Gavras, K. & Höhne, J.K. (2020) Evaluating political parties: criterion validity of open questions with requests for text and voice answers. *International Journal of Social Research Methodology*. Available from: <https://doi.org/10.1080/13645579.2020.1860279>
- Geer, J.G. (1988) What do open-ended questions measure? *Public Opinion Quarterly*, 52(3), 365–367.
- Geer, J.G. (1991) Do open-ended questions measure “Salient” issues? *Public Opinion Quarterly*, 55(3), 360–370.
- Google, (2020) Cloud Speech-to-Text API. Available from: <https://cloud.google.com/speech-to-text>
- Gummer, T., Quof, F. & Roßmann, J. (2019) Does increasing mobile device coverage reduce heterogeneity in completing web surveys on smartphones? *Social Science Computer Review*, 37(3), 371–384.
- Höhne, J.K. (2021) Are respondents ready for audio and voice communication channels in online surveys? *International Journal of Social Research Methodology*. Available from: <https://doi.org/10.1080/13645579.2021.1987121>
- Höhne, J.K., Cornesse, C., Schlosser, S., Couper, M.P. & Blom, A.G. (2020) Looking up answers to political knowledge questions in web surveys. *Public Opinion Quarterly*, 84(4), 986–999.
- Höhne, J.K., Gavras, K. & Qureshi, D. (2021) SurveyVoice (SVoice): a comprehensive guide for collecting voice answers in surveys. Zenodo. Available from: <https://doi.org/10.5281/zenodo.4644590>
- Höhne, J.K., Schlosser, S., Couper, M.P. & Blom, A.G. (2020) Switching away: exploring on-device media multitasking in web surveys. *Computers in Human Behavior*, 111, 106417.
- John, R. & Paul, N. (2002) An investigation of the lexical dimension of the IELTS Speaking Test. IELTS. Research Reports, Vol. 6.
- Kim, Y.M. & Garrett, K. (2012) On-line and memory-based: revisiting the relationship between candidate evaluation processing models. *Political Behavior*, 34, 345–368.
- Koizumi, R. (2012) Relationships between text length and lexical diversity measures: can we use short texts of less than 100 tokens? *Vocabulary Learning and Instruction*, 1(1), 60–69.
- 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.
- Laakso, M. & Taagepera, R. (1979) ‘Effective’ number of parties. A measure with application to West Europe. *Comparative Political Studies*, 12(1), 3–27.
- Lodge, M., McGraw, K.M. & Stroh, P. (1989) An impression-driven model of candidate evaluation. *American Political Science Review*, 83(2), 399–419.
- Lowe, W., Benoit, K., Mikhaylov, S. & Laver, M. (2011) Scaling policy preferences from coded political texts. *Legislative Studies Quarterly*, 36(1), 123–155.
- McGraw, K.M., Hasecke, E. & Conger, K. (2003) Ambivalence, uncertainty, and processes of candidate evaluation. *Political Psychology*, 24(3), 421–448.
- Nelson, K.N., Burk, D., Knudsen, M. & McCall, L. (2021) The future of coding: a comparison of hand-coding and three types of computer assisted text analysis methods. *Sociological Methods & Research*, 50(1), 202–237.
- Pang, B. & Lee, L. (2008) Opinion mining and sentiment analysis. *FNT in Information Retrieval (Foundations and Trends in Information Retrieval)*, 2(1–2), 1–135.
- Peterson, G., Griffin, J., LaFrance, J. & Li, J. (2017) Smartphone participation in web surveys. In: Biemer, P.P., de Leeuw, E., Eckman, S., Edwards, B., Kreuter, F., Lyberg, L.E., Tucker, N.C. & West, B.T. (Eds.) *Total survey error in practice*. Hoboken: John Wiley & Sons, pp. 203–233.
- Pietsch, A.-S. & Lessmann, S. (2018) Topic Modeling for analyzing open-ended survey responses. *Journal of Business Analytics*, 1(2), 93–116.
- Proksch, S.-O., Wratil, C. & Wäckerle, J. (2019) Testing the validity of automatic speech recognition for political text analysis. *Political Analysis*, 27(3), 339–359.
- Remus, R., Quasthoff, U. & Heyer, G. (2010) SentiWS - a Publicly Available German-Language Resource for Sentiment Analysis. Proceedings of the 7th International Language Resources and Evaluation, pp. 1168–1171.
- Revilla, M. & Couper, M.P. (2019) Improving the use of voice recording in a smartphone survey. *Social Science Computer Review*, 9, 1–20.
- Revilla, M., Couper, M.P. & Ochoa, C. (2018) Giving Respondents voice? The feasibility of voice input for mobile web surveys. *Survey Practice*, 11(2), 1–11.
- Revilla, M. & Ochoa, C. (2016) Open narrative questions in PC and smartphones: is the device playing a role? *Quality & Quantity*, 50(6), 2495–2513.

- 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.
- Roberts, M.E., Stewart, B.M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S.K. et al. (2014) Structural topic models for open-ended survey responses. *American Journal of Political Science*, 58(4), 1064–1082.
- Schlosser, S. & Höhne, J.K. (2018) ECSP – Embedded Client Side Paradata. Available from: <https://doi.org/10.5281/zenodo.1218941>
- Schlosser, S. & Höhne, J.K. (2020) ECSP – Embedded Client Side Paradata. Available from: <https://doi.org/10.5281/zenodo.3782592>
- Schober, M.F., Conrad, F.G., Antoun, C., Ehlen, P., Fail, S., Hupp, A.L. et al. (2015) Precision and disclosure in text and voice interviews on smartphones. *PLoS One*, 10(6), 1–20.
- Shapiro, M.J. (1970) Discovering interviewer bias in open-ended survey responses. *The Public Opinion Quarterly*, 34(3), 412–415.
- Smith, T.W. (1997) The impact of the presence of others on a respondent's answers to questions. *International Journal of Public Opinion Research*, 9(1), 33–47.
- 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.
- Stanga, J.E. & Sheffield, J.F. (1987) The Myth of zero partisanship: attitudes toward american political parties, 1964–84. *American Journal of Political Science*, 31(4), 829–855.
- Taber, C. & Young, E. (2013) Political information processing. In: Huddy, L., Sears, D.O. & Levy, J.S. (Eds.) *The Oxford handbook of political psychology*, 2nd edn. Oxford: Oxford University Press, pp. 525–558.
- Templin, M. (1957) *Certain language skills in children: their development and interrelationships*. Chicago: University of Minnesota Press.
- Torney-Purta, J.V. & Valsiner, J. (2005) *The development of political attitudes in children*. Somerset: Taylor and Francis.
- Tourangeau, R. & Bradburn, N.M. (2010) The psychology of survey response. In: Marsden, P.V. & Wright, J.D. (Eds.) *Handbook of survey research*. Emerald: Bingley, pp. 315–346.
- Tourangeau, R., Rips, L.J. & Rasinski, K.A. (2000) *The psychology of survey response*. Cambridge, New York, Melbourne: Cambridge University Press.
- Wallach, H.M., Murray, I., Salakhutdinov, R. & Mimno, D. (2009) Evaluation Methods for Topic Models. Proceedings of the 26th International Conference on Machine Learning, pp. 1105–1112.
- West, B.T. & Blom, A.G. (2017) Explaining interviewer effects: a research synthesis. *Journal of Survey Statistics and Methodology*, 5(2), 175–211.
- Yule, G.U. (1944) *The statistical study of literary vocabulary*. Cambridge: Cambridge University Press.
- Zaller, J. & Feldman, S. (1992) A simple theory of the survey response: answering questions versus revealing preferences. *American Journal of Political Science*, 36(3), 579–616.
- Zanna, M.P. & Rempel, J.K. (1988) Attitudes: A new look at an old concept. In: Bar-Tal, D. & Kruglanski, A.W. (Eds.) *The social psychology of knowledge*. Cambridge: Cambridge University Press, pp. 315–334.
- 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.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Gavras, K., Höhne, J.K., Blom, A.G. & 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 (Statistics in Society)*, 185(3), 872–890. Available from: <https://doi.org/10.1111/rssa.12807>