

Evaluating Political Parties: Criterion Validity of Open Questions with Requests for Text and Voice Answers

Konstantin Gavras

University of Mannheim (Germany)

Jan Karem Höhne

University of Mannheim (Germany)

Universitat Pompeu Fabra (Spain)

Abstract

The rise of smartphone surveys, coupled with technological advancements, provide new ways for measuring respondents' political attitudes. The use of open questions with requests for voice answers instead of text answers may simplify the answer process and provide nuanced information. So far, research comparing the measurement quality of text and voice answers is scarce. We therefore conducted an experiment in a smartphone survey (N = 2,402) to investigate the criterion validity of text and voice answers. Voice answers were collected using a JavaScript- and PHP-based voice recording tool that resembles the voice messaging function of Instant-Messaging Services. The results show that the open questions with requests for text and voice answers differ in terms of criterion validity. More specifically, the findings indicate that voice answers result in a somewhat higher criterion validity than their text counterparts. More refined research on the measurement quality of text and voice answers is required in order to draw robust conclusions.

Keywords: Criterion validity, Microphone, Open questions, Smartphone survey, Text answers, Voice answers

Introduction

Political science research and adjacent research fields require valid measurements of respondents' political attitudes to draw accurate and robust conclusions. However, obtaining valid measurements of respondents' political attitudes in surveys is a methodological challenge (Wuttke et al., 2020). Frequently, researchers employ closed questions (with rating scales consisting of pre-defined options). Although such questions allow researchers to employ a sizable body of questions and to obtain numerical values for statistical analyses, closed questions often provide only superficial information on respondents' political attitudes. Furthermore, closed questions are prone to measurement errors, such as extreme and middle response styles (van Vaerenbergh & Thomas, 2013).

Another long-standing method to measure respondents' political attitudes is the use of open questions (with text fields for entering answers) because they have the potential to collect in-depth and unfiltered information (Smyth et al., 2009). The majority of open questions

implemented in existing surveys is administered with requests for text answers. However, respondents frequently find it difficult to articulate themselves in writing and the answer entry process is burdensome.

The steady and rapid increase of web surveys – particularly smartphone surveys (Gummer et al., 2019) – and the recent technological advancements provide new avenues for measuring respondents' political attitudes using open questions. The microphones of smartphones allow researchers to administer open questions with requests for voice answers instead of text answers (Revilla & Couper, 2019; Revilla et al., 2020; Schober et al., 2015). The use of voice input instead of text input has the potential to facilitate the answer process, as in the former case, respondents only need to press a recording button and record their answers. This may result in higher measurement quality.

Although a few studies have compared open questions with text and voice answers, there is almost no research investigating measurement quality in terms of validity. For instance, Revilla et al. (2020) reported higher non-response rates for voice than text answers, but almost no differences in terms of valid (or meaningful) answers. In a follow-up study, Revilla and Couper (2019) investigated whether different instructions on how to record voice answers reduce non-response. The authors found no reducing effect across instructions and obtained a relatively high non-response rate. Finally, Schober et al. (2015) reported that text answers led to less rounding, less non-differentiation, and more disclosure of sensitive information than their voice counterparts. The authors concluded that text answers produce higher measurement quality than voice answers.

One methodological drawback of these studies is that the voice recording tools used were somewhat restricted. The study by Revilla et al. (2020) employed different voice recording functions for respondents using an Android device than those using an iOS device. The study by Revilla and Couper (2019) only considered Android respondents, whereas the study by Schober et al. (2015) focused on iOS respondents.

To our knowledge, no studies have investigated the measurement quality of text and voice answers in terms of validity, which is a pre-requisite for a proper evaluation of measurement quality. We fill this research gap by conducting an experiment in a smartphone survey (N = 2,402) that employs a voice recording tool that works across operating systems. We randomly assign respondents to a request condition (text or voice) and ask them to evaluate four German political parties using open questions. We also ask respondents one closed question on their party preference to evaluate the criterion validity of text and voice answers. We address the following research question: How do open questions with requests for text and voice answers affect measurement quality in terms of criterion validity?

Method

Experimental Design

One group (n = 1,694) received four open questions with a request for text answers (text condition). Another group (n = 1,679) received the same four open questions with a request for voice answers (voice condition).¹

¹ We also randomized respondents to different conditions (e.g., providing motivational messages) within the request conditions (text and voice), but these sub-conditions are not subject of this article.

Questions

Target questions: We used four political attitude questions on the evaluation of German political parties (CDU/CSU [Christian Democratic Union/Christian Social Union], SPD [Social Democratic Party], Greens [Alliance 90/The Greens], and AfD [Alternative for Germany]). These questions were presented on separate web survey pages. The questions were developed in German, which was the mother tongue of approx. 98% of the respondents. To maximize comparability between the text and voice requests for an answer, we employed an optimized survey layout, which generally prevents horizontal scrolling.

To record respondents' voice answers, we implemented a JavaScript- and PHP-based voice recording tool in the browser-based smartphone survey that records voice answers via the microphone of smartphones, regardless of the operating system. Figure 1 shows the design of the target questions with requests for text and voice answers.

The target questions with requests for text and voice answers were preceded by instructions. The instructions were tailored to the type of answer request (see the Appendix for English translations of the target questions and instructions).

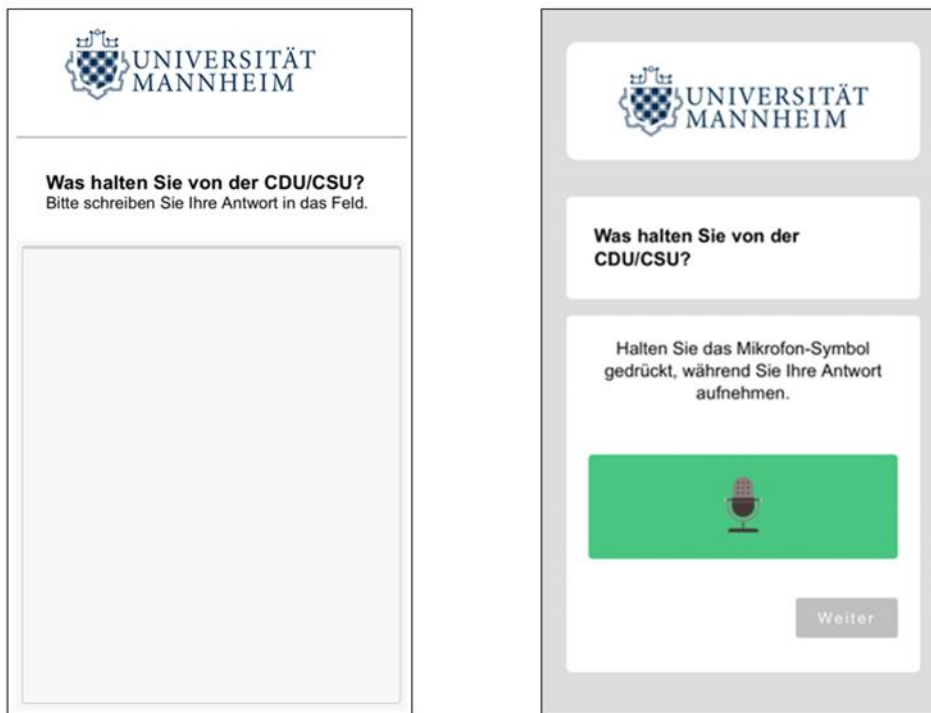


Figure 1. Example of the target questions on the German party CDU/CSU with requests for text (on the left side) and voice answers (on the right side)

Note. The 'Next' ('Weiter') button of the target question with a request for a text answer is not displayed because of space limitations. We did not limit the number of characters in the open answer field or the recording time in the voice tool.

Criterion question: One question on respondents' party preferences (see the Appendix for the English translation) was used as the criterion measure to evaluate criterion validity. This question was chosen as the criterion question because it was shown to be conceptually relevant

to the topic of the target questions (Lachat & Wagner, 2018). Additionally, it correlated significantly with all four experimentally manipulated target questions in the full sample. To determine criterion validity, we investigated which one of the two conditions (text and voice) produced higher correlations between the target questions and the criterion question. The criterion question was also written in German and had the following response categories: CDU/CSU (20%), SPD (12%), Greens (30%), AfD (11%), FDP (8%), the Left (8%), and other party (11%).

Data Source

Data were collected by Forsa (Germany) in December 2019 and January 2020. Forsa drew a quota sample from their access panel based on age, gender, education, and region (East and West Germany). The quotas were calculated using the German Microcensus.

The email invitation to the survey included information on the survey duration (about 15 minutes), the device (i.e., smartphone) to be used for survey completion, and a link to the survey. The first survey page outlined the topic and procedure of the survey and included a statement of confidentiality.

To restrict survey completion to smartphone respondents, we detected respondents' device at the beginning of the survey. Respondents who attempted to access the survey using a non-smartphone device were prevented from proceeding the survey and were asked to use a smartphone.

Overall, 19,754 panelists were invited to participate, of which 2,538 panelists were screened out because the quotas were achieved, the panelists declined to provide text and voice answers, or the panelists tried to access the survey with a non-smartphone device. In total, 3,373 panelists started the survey, of which 971 panelists broke-off before being asked any study-relevant questions (217 in the text condition and 754 in the voice condition).² This leaves us with 2,402 panelists (1,477 in the text condition and 925 in the voice condition) available for analysis.

Sample Characteristics

In total, 2,402 respondents participated in the study, which corresponds to a participation rate of approx. 12% among all invitees. On average, these respondents were born between 1975 and 1979, and about 51% of them were male. Approx. 23% had completed lower secondary school, approx. 33% intermediate secondary school, and approx. 44% college preparatory secondary school or university-level education.

As we faced break-offs, we conducted chi-square tests to evaluate the sample composition between the request conditions (text and voice). The results showed no significant differences regarding age, gender, and education.

Results

Initially, we automatically transcribed respondents voice answers into text using Google's Transcribe API "Speech-to-Text" (Google, 2020). As shown by Proksch et al. (2019, p. 342),

² The break-off rate is calculated independently from panelists that were screened out because of quotas, refusals, or non-smartphone devices. Therefore, the break-off numbers refer to those who started the survey.

the performance of the API does not substantially differ from human transcription. They found a cosine similarity of $r > 0.9$ on average between Google-transcribed and human-transcribed political speeches in German.

We then conducted sentiment analyses to extract sentiment scores for each respondent and target question on the evaluation of German political parties. This is done to obtain adequate measures that can be correlated with the measures of the criterion question on party preference. We used the German sentiment dictionary SentiWS (Remus et al., 2010) in which words are assigned sentiment scores that range from -1 (negative) to 1 (positive). The scores indicate the strength of the sentiment-afflicted words. We used the following formula to estimate sentiment scores (Lowe et al., 2011):

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

Equation 1. Estimating sentiment scores

Note. 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. We take the log of the results.

To evaluate the criterion validity of text and voice answers we ran four separate linear OLS regressions using the extracted sentiment scores of the target questions as independent variables. We used the criterion question as the dependent variable, which was recoded into four dummies: CDU/CSU (1 = yes), SPD (1 = yes), Greens (1 = yes), and AfD (1 = yes). To facilitate the interpretation of the results, we normalized sentiment scores to a scale running from 0 (negative) to 1 (positive).

Table 1. Linear OLS regressions predicting party preference with party evaluation (unstandardized coefficients)

| Independent variables | Text Coefficients (SE) | R ² | Voice Coefficients (SE) | R ² | Difference |
|------------------------------|---------------------------|----------------|----------------------------|----------------|------------------|
| Party evaluation: CDU/CSU | 0.24*** (0.05) | 0.01 | 0.28*** (0.07) | 0.02 | -0.04 (0.09) |
| Party evaluation: SPD | 0.04 (0.04) | 0.00 | 0.22*** (0.06) | 0.02 | -0.18* (0.07) |
| Party evaluation: Greens | 0.32*** (0.06) | 0.02 | 0.31*** (0.08) | 0.02 | 0.01 (0.10) |
| Party evaluation: AfD | 0.11** (0.04) | 0.00 | 0.15** (0.05) | 0.01 | -0.04 (0.06) |

Note. *p < 0.05, **p < 0.01, ***p < 0.001. Dependent variable: party preference. We normalized the sentiment scores to a scale running from 0 (negative) to 1 (positive).

As shown in Table 1, the correlations between the target questions on the evaluation of the German political parties and the criterion question on party preference differed across the two request conditions (text and voice). More specifically, in three out of four cases (CDU/CSU, SPD, and AfD), the open questions with a request for voice answers produce higher correlations than their text counterparts. This is suggested by the negative differences between the unstandardized regression coefficients. However, only the difference for the open question on

the SPD is statistically significant. Regarding the open question on the Greens, no substantial differences between the unstandardized regression coefficients were observed.

Discussion and Conclusion

The goal of this study was to investigate the criterion validity of text and voice answers to open questions dealing with the evaluation of German political parties. We conducted a smartphone survey and randomized respondents to different request conditions (text or voice). The results show that open questions with requests for text and voice answers differed regarding criterion validity. Overall, voice answers resulted in somewhat higher criterion validity than text answers. This is an important finding that indicates that open questions with requests for voice answers are a promising new method for measuring respondents' political attitudes in smartphone surveys.

This study was a first attempt to investigate the measurement quality of text and voice answers. Therefore, the study has some limitations that represent opportunities for future research. We only used four target questions on the evaluation of German political parties and one criterion question on party preference. We encourage future research to employ further test and criterion questions dealing with different topics, such as political efficacy and political interest. This would allow us to draw more robust conclusions about the measurement quality of text and voice answers.

Another limitation is that our data were collected using a non-probability access panel. Respondents from such panels are frequently characterized by a high survey experience. Previous research has shown that experienced respondents differ from less experienced respondents. For instance, Toepoel et al. (2008) showed that experienced respondents are more prone to satisficing. Thus, it would be worthwhile in future research to investigate open questions with requests for text and voice answers using data from a probability-based panel with less experienced respondents. This would also increase the generalizability of the findings.

Finally, there is a key challenge that must be tackled to establish open questions with requests for voice answers. Although Revilla et al. (2018) show that many respondents are generally willing to give voice answers, requests for voice answers result in comparatively high break-off rates. In this study, we observed a break-off rate of about 45% in the voice condition, compared to a break-off rate of about 13% in the text condition. This implies that open questions with requests for voice answers are not ready to entirely replace open questions with requests for text answers. However, they might be a promising supplement for specific respondent groups, such as young respondents or respondents with literacy issues. Given the comparatively high break-off rate it seems wise to leave respondents the choice which type of answer request to pick. In addition, it seems worthwhile developing appropriate strategies and techniques to reduce break-off rates when it comes to requests for voice answers in smartphone surveys.

References

Google. (2020). Cloud Speech-to-Text API. <https://cloud.google.com/speech-to-text>

- Gummer, T., Quoß, 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.
- Lachat, R., & Wagner, A. (2018). How Party Characteristics Drive Voters' Evaluation Criteria. *Electoral Studies*, 55, 11–20.
- Lowe, W., Benoit, K., Mikhaylov, S., & Laver, M. (2011). Scaling Policy Preferences from Coded Political Texts. *Legislative Studies Quarterly*, 36(1), 123–155.
- 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*, 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., Bosch, O. J., & Asensio, M. (2020). Testing the Use of Voice Input in a Smartphone Web Survey. *Social Science Computer Review*, 38(2), 207–224.
- 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.
- Schober, M. F., Conrad, F. G., Antoun, C., Ehlen, P., Fail, S., Hupp, A. L., Johnston, M., Vickers, L., Yan, H. Y., & Zhang, C. (2015). Precision and Disclosure in Text and Voice Interviews on Smartphones. *PloS One*, 10(6), 1-20.
- 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.
- Toepoel, V., Das, M., & van Soest, A. (2008). Effects of Design in Web Surveys. *Public Opinion Quarterly*, 72(5), 985–1007.
- van Vaerenbergh, Y., & Thomas, T. 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.
- Wuttke, A., Schimpf, C., & Schoen, H. (2020). When the Whole Is Greater than the Sum of Its Parts: On the Conceptualization and Measurement of Populist Attitudes and Other Multidimensional Constructs. *American Political Science Review*, 114(2), 356–374.

Appendix

English translations of the instructions, target questions with a request for text and voice answers, and the criterion question.

Instruction for the text condition:

Next, we would like to ask you some questions on political issues and parties. You will be asked to provide the answers in your own words. You can enter your answers in the open answer field via the keypad of your smartphone.

After successful entry, click on “Next” to continue with the survey as usual.

Instruction for the voice condition:

Next, we would like to ask you some questions on political issues and parties. You will be asked to provide the answers in your own words. You can record your answers via the microphone of your smartphone.

Press and hold the microphone icon while recording your answer.

Once you have recorded your answer, you can stop pressing the microphone icon. A tick will indicate that you have successfully recorded your answer.

After successful recording, click on “Next” to continue with the survey as usual.

Target questions:

What do you think about the CDU/CSU?

What do you think about the SPD?

What do you think about the Greens?

What do you think about the AfD?

Additional instruction (text condition): Please write your answer in the open answer field.

Additional instruction (voice condition): Please press the microphone icon while recording your answer.

Criterion question:

If there were a federal election next Sunday, for which party would you vote?

Response categories: CDU/CSU, SPD, Greens, AfD, FDP, the Left, and other party

Note. The four target questions on the evaluation of German political parties were preceded by two open questions, which are not subject of this article. The first one dealt with the most important political problem in Germany and the second one dealt with the performance of the German chancellor (Angela Merkel). The original German wordings are available from the first author on request.