

# Expanding Survey Response Options: Combining Dictation and/or Voice Recording with Text to Answer Narrative Open-ended Survey Questions

Melanie Revilla<sup>1</sup> · Mick P. Couper<sup>2</sup>

<sup>1</sup>RECSM-Universitat Pompeu Fabra

<sup>2</sup>University of Michigan, Institute for Social Research

While the advantages of voice input for answering open questions in web surveys seem clear, the challenge remains of maximizing voice inputs use while still giving respondents alternatives. This study experimentally explores three options for encouraging voice input in web surveys: a) *PushDictation*: respondents are asked to answer using dictation. If they try to skip the question, they are proposed to type in a textbox; b) *PushRecording*: respondents are asked to answer using voice recording. If they try to skip the question, they are proposed to type in a textbox; and c) *Choice*: respondents are offered three options to answer (dictation, voice recording or type in a textbox). These three options are compared to a *Control* group in which participants can only answer by typing in a textbox.

Using data from two open questions in a survey about nursing homes implemented in February/March 2023 ( $N = 1,001$ ) in an opt-in online panel in Spain (Netquest), we answer three research questions: (**RQ1**) What are the overall rates of response to the two open questions? (**RQ2**) What are the rates of use of voice input to these same questions? (**RQ3**) What is the overall quality of the data across the different conditions?

Overall, response to the open questions was significantly lower when voice inputs were proposed, especially in the *PushRecording* group (**RQ1**). Furthermore, significant differences emerged in voice input usage between the experimental groups (**RQ2**). Regarding data quality (**RQ3**), the *Control* group exhibited the lowest proportion of valid answers, while the average numbers of themes and characters were in general higher in the push groups.

Our results contribute to the growing but still limited literature about the use of voice input in web surveys, by adding new empirical evidence for several designs encouraging voice input to answer open questions.

**Keywords:** data quality; dictation; answers to open questions; voice recording; web surveys

---

**Supplementary Information** The online version of this article (<https://doi.org/10.18148/srm/2026.v20i1.8456>) contains supplementary material.

---

Corresponding author: Melanie Revilla, RECSM-Universitat Pompeu Fabra, Barcelona, Spain (Email: [melanie.revilla@hotmail.fr](mailto:melanie.revilla@hotmail.fr))

## 1 Introduction

Open questions have long been a part of the survey tradition (e.g., Lazarsfeld, 1935, 1944; Schuman, 1966). However, they had fallen out of favor, largely because of the cost of collecting, processing and analyzing responses to such questions. The move to web-based data collection further eroded the interest in using open questions. However, in recent years, the development of natural language processing (NLP) tools has made the analysis of textual data more amenable to researchers (e.g., Schonlau & Couper, 2016; Schonlau et al., 2021). Recent papers have enumer-

ated the benefits of open questions for both substantive and methodological purposes (e.g., Neuert et al., 2021; Singer & Couper, 2017). However, making it easier for respondents to provide answers to open questions in web surveys has remained a challenge. The increase in the use of mobile devices (mainly smartphones) for answering web surveys has exacerbated this problem: a number of studies have documented shorter responses to open questions on mobile devices than on PCs (e.g., Revilla & Couper, 2021).

The rise in smartphone use has also presented a potential solution: the last few years have seen a rapid increase in the use of voice assistants (e.g., Alexa, Siri, Google Voice) and voice functions. For example, Deloitte (2018) noted that “Nearly two-thirds (64%) of survey respondents now use a voice assistant on their smartphones—compared to 53% in 2017. What’s more, nearly half (46%) employed this capability on their smartphones within the ‘last week’—and almost a third (30%) did so within the ‘last day’” (p. 11, see also PWC, 2018). While the use of voice features has been particularly important on mobile devices, such features are also used on PCs (Wardini, 2023). Given these trends, several authors have enumerated the benefits of using voice input over text entry (typing) to elicit responses to open questions in (mobile) web surveys (Gavras et al., 2022; Höhne et al., 2024; Revilla, 2022). In particular, these authors argue that voice responses may yield longer and richer answers, while also capturing less cognitively-processed, more spontaneous/intuitive input—thereby offering new avenues for insight. Additionally, voice data provide paralinguistic information such as accent, tone, and intonation, which have been utilized in some studies, for example, to predict respondents’ interest in survey questions. From the respondents’ perspective, voice input may also be faster than typing their answers and thus perceived as less effortful, as well as more enjoyable.

While the advantages of voice input seem clear, its implementation for open questions has met with difficulties. As reviewed in Sect. 2, several studies have found respondents unwilling or unable to use voice input, and requiring them to use it to answer open questions has resulted in higher levels of break-off and non-compliance. Given that we do not expect all or even most respondents to use voice input, the challenge remains of maximizing the use of voice input while still giving respondents alternatives. This study experimentally explores three options for encouraging voice answers in the frame of web surveys (that can be completed through smartphones, tablets or PCs):

a) *PushDictation*: respondents are first proposed to answer through dictation (also called *Automatic Speech Recognition* or *ASR*), which means that while they speak, their voice is transcribed into text on their device’s screen. What is stored is a text file similar to what is normally

obtained with open narrative questions. If respondents continue without providing an answer, they are offered the option to type in a textbox.

- b) *PushRecording*: respondents are first proposed to answer by recording their voice. They can record several files, listen to them, and delete the ones they want before submitting them. If respondents continue without providing an answer, they are offered the option to type in a textbox.
- c) *Choice*: respondents are proposed three options to answer (dictation, voice recording or type in a textbox). All three options are presented on the page together with the question and respondents can try any of them and/or use several of them to answer.

The advantage of dictation for the respondent is that it displays the text in real time, allowing them to edit the text or dictate their answers again. In addition, dictation does not necessarily require the voice files to be saved (potentially reducing privacy concerns), although many voice input systems can also save the voice files if needed. The advantages for recording lie mostly with the researchers: 1) it is argued that recording will encourage more spontaneous and candid responding (Gavras et al., 2022) and 2) that the audio files can be used for detailed analysis after the fact (permitting, for example, detection of emotion or accents of respondents). Some respondents may also prefer recording their answers over dictating them (e.g., because this avoids seeing transcription errors which sometimes occur with dictation), especially if they are accustomed to doing so in daily life. Overall, differences can thus be expected between dictation and voice recording, and research testing both approaches is needed.

## 2 Background

Early studies investigated respondents’ willingness to use voice input to answer survey questions. In a survey of Netquest panelists in Spain in 2016, Revilla et al. (2018) found that around half the sample already used voice input at least sometimes and that a similar proportion was willing to use it to answer open narrative questions. Similarly, Höhne (2021) investigated respondents’ willingness to participate in online surveys with a smartphone using voice output and voice input in the German Internet Panel. He found that a “substantial minority of respondents is willing to participate in online surveys with a smartphone to have the survey questions read out loud and to give oral answers via voice input” (p. 6). Lenzner and Höhne (2022) asked respondents in a nonprobability online panel in Germany: “In general, are you willing to participate in upcoming surveys with your smartphone to give oral answers via voice input?” While this conflates smartphone use and voice input,

only 1% reported not having a smartphone, while 25% said definitely no and 31% said probably no (55% combined), 25% said probably yes and 15% definitely yes (40% combined). The remaining 4% chose the “Don’t know how it works” option. Those who were unwilling to use voice input mostly mentioned a preference for written communication as a reason. Lenzner and Höhne (2022) concluded: “Audio and voice channels in smartphone surveys appeal primarily to frequent and competent smartphone users as well as younger and tech-savvy respondents” (p. 604).

Several recent experimental studies have explored various forms of voice inputs for open questions in web surveys. Lütters et al. (2018) reported on an experiment in a German opt-in panel in 2017, in which panelists were randomized to three input conditions for answering three sensitive questions: 1) voice only, 2) voice plus text, and 3) text only. Fewer respondents completed the survey in the voice-only (49%) and choice (54%) groups than in the text-only group (94%). The average word count was similar across conditions for the first question, but significantly lower for text-only in the second and third questions, suggesting a fatigue effect for text. Respondents’ rating of the voice experience was significantly lower than for text. The authors were optimistic that the technical issues with voice inputs would be “sorted out soon” and concluded that speech as an answer type is an exciting new way of getting more open-ended answers, but not necessarily to the respondent.

In a pretest in the German Longitudinal Election Study (GLES) in 2018, Gavras (2019) randomly assigned respondents to a voice recording or text input version of the final “any comments” question. Details of the uptake or item missing data rates to the two versions were not provided. The text condition yielded more than twice as many comments as the voice condition. Gavras further reported that voice recording pushes under-covered groups (older, lower political interest and lower educated) to respond to open-ended questions.

Revilla et al. (2020) conducted an experiment in an opt-in panel in Spain in 2018, in which respondents using smartphones were randomly assigned to voice or text input. For iOS, the authors used the built-in dictation button. For Android, they used the MediaStream API to do voice recording. Each operating system had an experimental group and a control group answering by typing in a textbox. Among those assigned to the Android-Voice condition, 63% did not answer any of the six open questions (compared with 2% in Android-Control, 3% in iOS-Control, and 3% in iOS-Dictation). Similarly, 26% of Android-Voice respondents reported problems during the survey (compared with 5% in Android-Control, 6% in iOS-Control, and 9% in iOS-Dictation). Additional coder-detected technical problems were also higher in the Android-Voice condition. In the iOS-Dictation group, on average across the six experimental ques-

tions, 40% of respondents did not use the dictation button, instead typing their answer in the text box, whereas 42% provided answers only using the dictation button and 18% manually edited the text after using the dictation button. Among those who provided valid answers, the quality of answers was significantly lower in the iOS-Dictation group than the control for several indicators, while the results were more mixed for the Android-Voice group relative to the control. Android-Voice respondents elaborated more on their answers, but did not provide new information.

Given the difficulties of using the built-in features of the operating system to capture voice input and differences between operating system, Revilla and Couper (2021) conducted a follow-up study to test ways to improve the voice recording option on Android devices. They found that different instructions to help respondents use the voice recording tool had little effect on uptake rates. They also tested a filter question to determine whether respondents were in a setting that permitted voice recording, and routed those unable to use voice to text input. This approach was more successful. They concluded that despite these efforts to reduce problems, “a substantial proportion of respondents are still unwilling or unable to answer open questions using voice recording” (p. 1). They further noted that “providing another alternative than voice seems important” (p. 15).

Gavras and Höhne (2022) and Gavras et al. (2022) conducted an experiment among smartphone users in a German non-probability panel in 2019–2020. Respondents were randomly assigned to voice recording or text input for six open-ended political attitude questions. The voice condition had higher break-off rates (45% vs. 13%) and higher item non-response rates (25% to 28% vs. 2% to 4% across questions) than the text condition. They found that “written [text] answers are characterized by an intentional and conscious answering, whereas oral [voice] answers are characterized by an intuitive and spontaneous answering” (p. 873). They caution that missing data must be taken into account when designing studies using voice input, otherwise “this may have serious consequences for data utility” (p. 888).

Höhne et al. (2024) conducted a similar study among smartphone users in a German non-probability panel in 2021, focusing on four sensitive questions. Respondents were again randomized to a voice recording or text condition and were told that they would be asked to answer via voice (or text) before seeing any of the open questions. 51% of those in the voice condition broke off the survey, compared with 24% in the text condition. Overall, voice responses were significantly longer, had a larger variety of words, and were characterized by a more diverse set of vocabulary than text responses. They found no significant differences in sentiment or extremity of the responses. They noted that “voice answers (...) represent a promising extension of the existing methodological toolkit in web survey

research” (p. 18). They concluded: “For now, we recommend being open to voice answers and to investigate their merits and limits for web survey research in future studies” (p. 18). Lenzner et al. (2024) used data from the same study to investigate whether voice responses are effective in the context of web probing questions, finding higher nonresponse rates but also richer answers in the voice condition.

Finally, Meitinger et al. (2022) conducted an experiment in the LISS panel in the Netherlands in 2020. Respondents were randomly assigned to one of three conditions: 1) text only, 2) voice recording only, and 3) choice between text and voice recording. Of those assigned to the choice condition, 94% opted for responding by text. Those assigned to voice or opting for voice were then asked a series of questions about technical, situational, and legal issues regarding microphone usage. As a result of this sequential process, 78% of respondents in the voice-response only condition and 3% in the choice condition were screened out (almost half of those who chose voice input); none were screened out in the text-only condition. Among respondents who passed all of the screenout steps, the break-off rate was 20% in the voice-recording only condition, compared with 8% in the text condition and 8% in the choice condition. After answering the two open questions in a voice-recording format, respondents were asked whether they would have preferred to type their answer: more than half (54%) would rather have used text input. Meitinger et al. (2022) concluded that, it might be safer to let respondents choose between voice recording and written responses, in order to prevent potential bias in the sample due to participant break-off and item nonresponse.

A key conclusion from these studies is that technical and logistical hurdles pose a significant barrier to using voice input. While the results are generally positive regarding the benefits of voice input on the quality of open-ended responses (length, number of unique concepts, etc.), the key challenge remains getting people to use voice input to answer open questions. Giving those who are unwilling or unable to use it, an alternative is important. Further, the research above points to different ways of implementing voice inputs (e.g., dictation versus recording, or the use

of different tools to record the voice). With this in mind, we designed a study to test alternative ways of providing respondents a choice of voice or text inputs.

### 3 Experimental design and research questions

Our study focuses on ways to maximize the use of voice input to take advantages of the gains in data quality without increasing item-missing data by requiring respondents to use a voice tool. Further, as noted above, most studies have used voice recording for voice input while Revilla et al. (2020) also used dictation, depending on the operating system. No study (to our knowledge) has directly compared voice recording with dictation.

Regarding how to offer dictation and voice recording options to participants, we rely on the mixed-mode literature to inform our design. Early research offering respondents a choice of modes found lower response rates than the single-mode alternative (see Medway & Fulton, 2012, for a summary). These findings are consistent with the “paradox of choice” literature (Schwartz, 2004) that suggests that giving people more choices increases indecision. Similarly, the behavioral change literature suggests that “nudges” (Thaler & Sunstein, 2008) may more effectively guide people to choose the preferred option. More recent mixed-mode designs are employing a sequential push-to-web approach rather than a mode choice approach (Cornesse et al., 2021; Dillman, 2017; Patrick et al., 2018). We expected that a similar approach may work to “nudge” or encourage respondents to use the voice-input tool.

Therefore, we experimentally test two sequential push-to-voice approaches, one using dictation and the other recording. We contrast this with a choice condition (proposing dictation, recording and/or text), and a text-only control condition (see Table 1).

Our primary research questions focus on the overall rates of response to the open questions (i.e., rates of panelists who provided an answer of any kind—voice or text—over all those who saw the question) and the rates of response through voice input.

**Table 1**

#### *Experimental Groups*

Control	PushDictation	PushRecording	Choice
Text answers only.	Propose dictation, if they do not answer, also offer text.	Propose recording, if they do not answer, also offer text.	Choice between: Dictation Voice Text

In the *Choice* group, the different tools were offered on the same page as the open question. This is different from Meitinger et al. (2022), where respondents were first asked to choose between several options, and then only got one way of responding.

**RQ1** What are the overall rates of response to open questions?

As reviewed above, the design challenge is to increase the use of voice input without threatening the overall rate of response to open questions. That is, can technical improvements and offering respondents alternatives minimize break-off and item nonresponse relative to the control group? In terms of overall question response rates, we expect the *Choice* group to have the next highest rate of response after the *Control* condition, followed by the two push conditions. Thus, our first hypothesis is:

**(H1)** Response rate *Control* > Response rate *Choice* > (Response rate *PushDictation* = Response rate *PushRecording*).

**RQ2** What are the rates of use of voice input to answer open questions?

We expect the superiority of the choice condition over the push conditions in terms of overall question response is likely to come at the expense of lower use of voice. That is, we expect fewer people will choose voice input when offered a choice than when first pushed to use voice. Regarding the differences between the two push conditions, on the one hand, we expect that the voice recording condition may require less effort on the part of respondents, and they may be more familiar with voice recording, given the widespread use of voice messaging apps like WhatsApp. On the other hand, dictation offers respondents greater autonomy and control, as well as increased privacy when the audio files are not retained (as is the case in this study). We thus expect the following:

**(H2)** (Use of voice input *PushRecording* = Use of voice input *PushDictation*) > Use of voice input *Choice*. By definition the *Control* group will have no voice input.

Our secondary research question focuses on data quality. As the literature has noted, there are a number of benefits for respondents and researchers of using voice input; however, these benefits only accrue if respondents actually use the voice feature. Besides, the benefits can vary depending on different factors, such as the exact question of interest, the target population or the tool used to collect the voice data. It is thus important to assess the overall quality in each experimental group of our study.

**RQ3** What is the overall quality of the data across the different conditions?

Consistent with prior literature, we assume that the quality of responses to the open questions, measured with different indicators (e.g., number of characters or number of themes, see Sect. 4.5.2), will be higher when respondents use voice than text. If this is the case, we expect the quality for responses to be highest in the two push conditions that are designed to maximize the use of voice input. Further, even if dictation should make it easier to edit or correct the responses using text input than doing a new voice recording, following the results of Revilla et al. (2020), we expect recording to achieve higher data quality than dictation. That is, our third hypothesis is:

**(H3)** Quality *PushRecording* > Quality *PushDictation* > Quality *Choice* > Quality *Control*.

## 4 Methods and data

### 4.1 Data collection

The data were collected between the 22<sup>nd</sup> of February and the 30<sup>th</sup> of March 2023 in the Netquest<sup>1</sup> opt-in online panel in Spain.

The objective was to get a sample of 1000 panelists completing the full survey. Quotas for gender and age (2 × 3 groups), education (3 groups), and autonomous community (17 groups) were defined to reproduce the proportions of the adult online population living in Spain (under 75 years old) according to the National Statistics Institute<sup>2</sup>.

From the 4789 panelists invited to the survey, 1860 started it, 113 abandoned on the initial survey page, 170 were excluded for not giving their explicit consent to participate, 17 for not passing basic fraud checks, 185 for exceeding the quotas, and 205 for stating being not at all familiar with the way nursing homes work in Spain. This left 1170 who started the survey and were eligible to participate. In addition, 25 broke-off before getting to our first experimental question. Thus, 1145 were assigned to one of the experimental groups and these are used in our analyses on rates of response. A further 144 panelists broke-off during or after the experimental questions. Thus, 1001 panelists completed the full survey and are used in our analyses on data quality.

The average age of these 1001 participants who completed the survey is 47 years old; 51% are female; and 35% have a higher education degree. On average, they have been registered in the Netquest panel for six years (median = 5.7) and have completed 157 surveys (median =

<sup>1</sup> [www.netquest.com](http://www.netquest.com)

<sup>2</sup> [www.ine.es](http://www.ine.es)

141), thus around one every two weeks. About a quarter (24%) responded on a PC, 3% from a tablet and 74% from a smartphone. The average survey duration is 9.1 min (median = 8.2). No significant differences between experimental groups in terms of gender, age or education level were found.

## 4.2 Implementation of the experiment

The experiment was implemented in two narrative open-ended questions that asked respondents to explain why they selected a given answer in the previous closed question:

- **OQ1.** [WHYTRANSP] Explain why you think that nursing homes provide [no information at all/very little information/some information/a lot of information/a huge amount of information] about the implementation of their services. Please give as much detail as you can. In your answer, mention if you think there is a difference among **public** and **private** nursing homes.
- **OQ2.** [WHYTRUST] Explain why you personally [not at all/very little/somewhat/very much/completely] **trust** nursing homes. Please give as much detail as you can. In your answer, mention if you think there is a difference among **public** and **private** nursing homes.

Respondents were assigned to one of the groups presented in Table 1. The assignment took place when respondents received **OQ1** and was maintained for **OQ2**. Respondents were assigned to the group with fewer completes at the time of the assignment to account for differences in break-off rates. The target sample size was 250 in each group. Appendix 1 provides the detailed wording (English translation) for the instructions that are specific to each group and the supplementary online material (SOM1) provides screenshots of these questions in all experimental groups (from a PC and from a smartphone).

The dictation and voice recording options were programmed using the *WebdataVoice* tool (Revilla et al., 2022). Thus, both options could be used in PCs, smartphones and tablets, as long as a microphone was available and the authorization to use it was provided. Respondents in the *Push-Dictation* and *PushRecording* groups were informed that if they did not have a microphone, they should click on the next button to continue. All respondents in these groups who skipped the question (because they did not have a microphone or for other reasons) were proposed again to answer but this time a textbox alternative was added. In both dictation and voice recording, the respondents had the option to check their answers before submitting them: for dictation, by reading the transcriptions; for voice recording, by listening to the audio files. If they were not satisfied, for dictation, respondents could use the keyboard to edit or

delete the transcription, or start another dictation. For voice recording, they could delete the voice file and record a new one. In the case of dictation, the fieldwork company only received the text files, as with any conventional open question, whereas for voice recording, the fieldwork company received the audio files. However, to minimize any data protection issue, those were immediately transformed by the fieldwork company into text files using the Vosk API<sup>3</sup> and the research team only received the transcriptions.

## 4.3 The questionnaire

The questionnaire included more than 80 questions that were asked in an online survey optimized for mobile devices but accessible also from PCs. However, none of the respondents got all questions, due to routing. The full questionnaire and its English translation are available in SOM2.

Respondents could continue without answering the questions, except those used to control quotas and filter/tailor other questions. A warning message was shown to 102 participants who tried to skip a question when multiple items were presented on the same page. Following the panel's usual practice, going back in the survey was not allowed.

The survey mainly dealt with citizens' perceptions of nursing homes in Spain (e.g., to what extent they trust them) but also included a block of questions about political opinions (e.g., left-right self-placement), as well as sociodemographic questions (e.g., mother tongue), questions about the context in which respondents answered the survey (e.g., presence of third parties) and about their evaluation of some questions (e.g., did they like answering open-ended questions by typing text).

In this paper, we are mainly interested in the two narrative open-ended questions detailed above (**OQ1** and **OQ2**). A third narrative open-ended question was also presented to the panelists: this question, asking about the perceived quality of the nursing homes, was placed before the two experimental ones, and all respondents were asked to answer it using a text-box format. It was used to compare the data quality across experimental groups when using the same response format for all of them. The results show that there are generally no significant differences for this question across experimental groups on the different indicators of quality used in this study (see SOM3).

## 4.4 Additional information

In addition to the responses to the survey questions, we gathered information from:

<sup>3</sup> <https://github.com/alphacep/vosk-api/>

- panel history: in particular information about participants' past engagement in surveys conducted by Netquest was used as a proxy for their experience and involvement in survey participation.
- paradata: we collected data about the devices used to complete the survey, as well as information about completion time, especially the focus time, i.e., the time that the survey page is active (similar to the SurveyFocus method proposed by Höhne et al., 2017). Additionally, we captured information generated by the *WebdataVoice* tool (Revilla et al., 2022) about the respondents' behaviors when dictation and/or voice recording were offered.

## 4.5 The analyses

All analyses were performed using R version 4.3.1 (R Core Team, 2023).

### 4.5.1 Analyses regarding rates of response

When individuals are presented with a question, they typically have several options at their disposal. They can choose to answer the question directly, skip the question by clicking the “next” button, or break off the survey. Furthermore, in the case of the two experimental questions, within the *PushDictation* and *PushRecording* groups, respondents who initially chose to skip a question were followed up with a prompt encouraging them to reconsider their response and offering a text-based alternative. Consequently, to analyze the rates of response, we first created a new variable for each experimental question encompassing the various scenarios, considering all the individuals who responded to the prior questions (i.e., 1145 respondents for *OQ1* and 1040 for *OQ2*).

Appendix 2 provides a description of these scenarios, together with the number of panelists in each. Based on this, to address *RQ1* (overall rates of response), we compute and report for each experimental question and group:

- *Break-off rate*: number of break-offs divided by number of people who saw the experimental question.
- *Item nonresponse rate*: number of item nonresponse divided by number of people who saw the experimental question.
- *Response rate*: number providing an answer (in any way, in the initial or follow-up question) divided by number of people who saw the experimental question. This is the complement of the other two.

Additionally, to address *RQ2* (rates of use of voice input), we compute and report for those who provided an answer, the rates of those who used voice inputs:

- *Response rate through voice*: number of responses provided using voice input divided by total number of responses provided (i.e., the number who saw the question minus break-off minus nonresponse).

For all four indicators, we test whether there are significant differences across experimental groups using Fisher's exact tests (Agresti, 1992). To test *H1* (Response rate *Control* > Response rate *Choice* > [Response rate *PushDictation* = Response rate *PushRecording*]), we compare *Control* versus *Choice*, *Choice* versus *PushDictation*, *Choice* versus *PushRecording*, and *PushDictation* versus *PushRecording* for the three first indicators. We also compare all treatment groups to the *Control* to see if there is any significant impact of proposing voice inputs.

To test *H2* ([Use of voice input *PushRecording* = Use of voice input *PushDictation*] > Use of voice input *Choice*), we compare *Choice* versus *PushDictation* and *PushRecording*, and *PushDictation* versus *PushRecording*, for the last indicator. Since there is no voice option in the *Control* group, we do not consider it in these analyses.

### 4.5.2 Analyses regarding data quality

To address *RQ3* regarding data quality, we consider different indicators of quality, focusing on all the respondents who completed the full survey (so they may have answered by voice recording, dictation or text). Full coding guidelines are available in SOM4.

First, we consider the conceptual validity (Billiet, 2016) of the answers, which assess whether the answers provide information to the researchers to measure the concepts of interest. To assess this, two researchers independently assessed the validity of answers to each question. Subsequently, we compared their assessments. Since significant discrepancies arose and these assessments determined whether further coding was warranted (with coding being performed only for “valid” answers), we involved a third coder, who was an expert in the substantive topic of the questions. For all cases where the initial two coders concurred, their assessment became the final code. In instances where disparities existed between Coders 1 and 2, or when one coder was uncertain about which code to apply, Coder 3 was tasked with reviewing and making the ultimate code decision.

We defined valid responses as those providing a substantive meaningful response to the question. Thus, invalid responses included nonsenses (e.g., “jjahng”), responses not in line with the question, those that said it depended on

other issues, those that simply referred back to the previous answer, and explicit “don’t know” responses or refusals to respond (see SOM4 for further details of the coding scheme).

Secondly, for respondents whose answers were categorized as “valid” based on the final codes, we further assessed whether they answered both parts of each open question (one inquiring about why they chose a specific answer in the preceding question and the other asking whether they perceived differences between public and private nursing homes). Moreover, coders were asked to indicate if they identified any problems with the answer (e.g., incomplete sentences) and to identify all the distinct themes mentioned in response to each question, employing a coding scheme developed by a subject matter expert. While the specific themes are not detailed in this paper, we used them to count how many different themes are mentioned in each response. The use of abbreviations or emojis were also coded. Finally, the length of each response (number of characters including spaces) was automatically calculated from the transcripts. Although each indicator has its limitations (e.g., longer responses do not necessarily reflect higher quality), the combination of multiple indicators offers a more robust assessment of response quality.

While notable discrepancies between coders were also observed for these indicators (please see the information about Inter-Rater Reliability in SOM5), Coder 3 did not participate in their coding, except for addressing specific uncertainties raised by Coders 1 and 2. This decision was made because the coding of these indicators did not impact other aspects of the analysis and due to resources limitations. Consequently, there is no single final code for these indicators. For the sake of simplicity, in the paper, we present the results based on the assessments made by Coder 1. Results based on the evaluations by Coder 2 can be found in SOM5. In general, the conclusions drawn from group comparisons remain consistent, with the primary difference lying in the absolute levels reported for different categories.

The next section presents the percentages of *Valid* answers (after Coder 3 assessment), among those who provided any form of response. Furthermore, among those providing valid answers, it presents the proportions of respondents answering both parts of the question, the proportions of answers where no problems were detected, and the average number of themes and of characters. Detailed results for other categories (e.g., “Nonsense”) as well as information about sample size and results about abbreviations (no significant differences across groups), are provided in SOM5.

Furthermore, to evaluate *H3* (Quality *PushRecording* > Quality *PushDictation* > Quality *Choice* > Quality *Control*), we conducted Fisher tests (for proportions) or T-tests

(for means) for the different indicators of quality to detect significant differences at a 5% significance level. Specifically, we compared *PushRecording* with *PushDictation*, *PushDictation* with *Choice*, and *Choice* and *Control*. We also compared *PushRecording* and *PushDictation* to *Control*. The results for the tests involving the other pairs of groups can be found in SOM5. We should note however that due to the sample size (less than 200 per group when focusing only on the valid answers), our statistical power is limited, which may prevent detecting small differences.

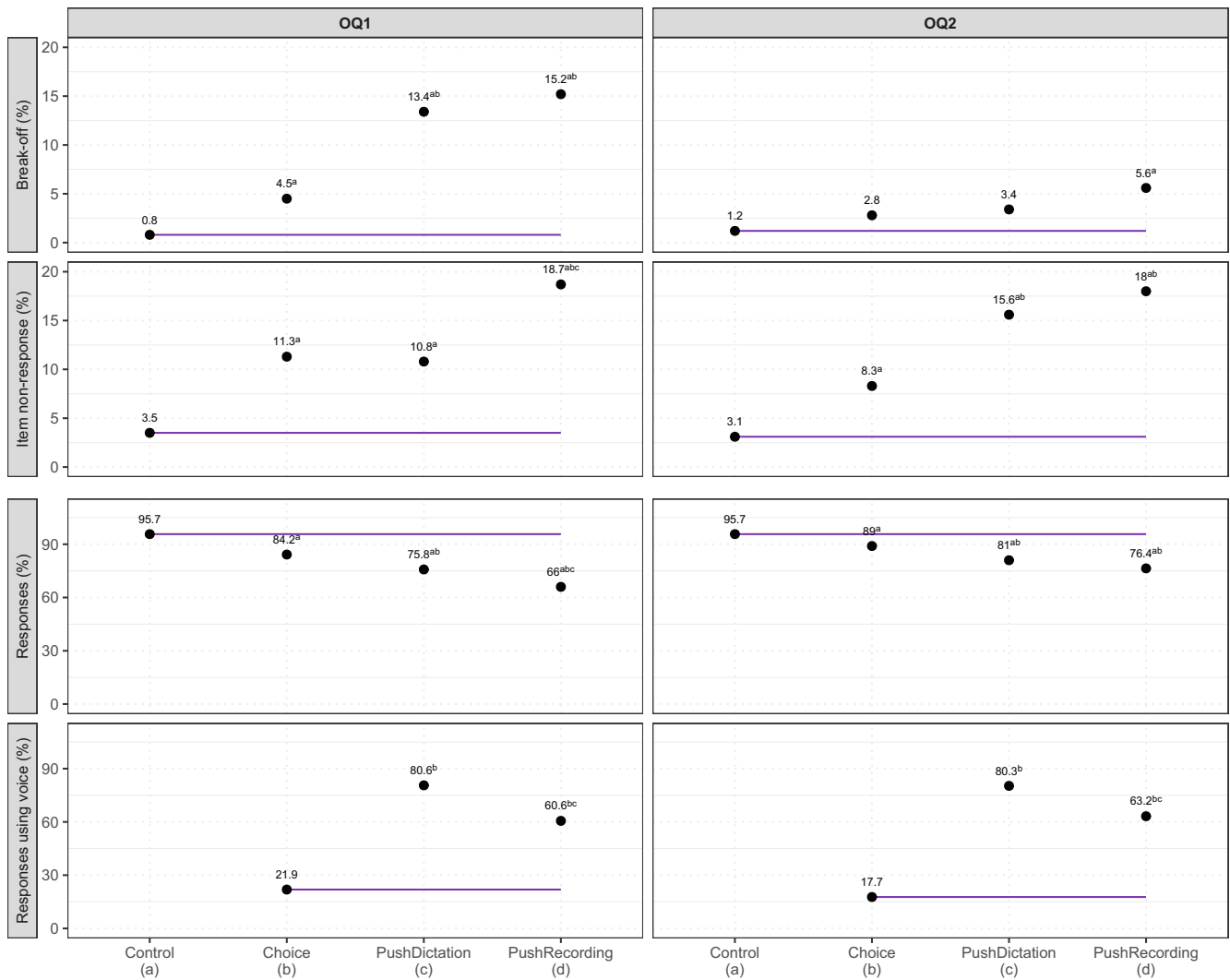
To account for potential selection bias in the participation of different experimental groups and investigate their impact on the primary quality indicators, we also conducted logistic and linear regression analyses. The dependent variables for these regressions are the main quality indicators of interest, namely: Valid Answer, Answered Both Parts, Abbreviation(s), No Problem Detected (of any kind), Number of Themes and Number of Characters. The main independent variables are dummy variables representing the three treatment groups, with the *Control* group serving as the reference category. Additionally, we incorporated a set of control variables expected to influence the quality of responses to the experimental questions. Appendix 3 presents our expectations for each independent variable, along with descriptive statistics for a comprehensive understanding of each considered variable. Overall, regressions results are in line with the descriptive analyses. Thus, we present only the descriptive results in the main text (full results available in SOM6).

## 5 Results

### 5.1 Overall rates of response to open questions and of response using voice inputs (*RQ1* & *RQ2*)

The results related to the overall response to the open questions (*RQ1*) and use of voice inputs (*RQ2*) are presented in Fig. 1 (see also Appendix 2 for details about the number of panelists in each situation).

Regarding *RQ1*, the break-off rates are significantly higher in the experimental groups for the first experimental question relative to the control (as indicated by the superscript <sup>a</sup>), particularly in the push groups (peaking at 15% in the *PushRecording* group). The break-off rates for *PushDictation* and *PushRecording* are also significantly higher than for *Choice* for *OQ1* (as indicated by the superscript <sup>b</sup>). For *OQ2*, break-off rates are substantially lower in the experimental groups (as expected given survival to this question), but remain significantly higher in the *PushRecording* than in the *Control* group. Item non-response is also significantly higher in all experimental



**Fig. 1**

*Response per Experimental Question and Group Superscripts indicate statistically significant differences at the 5% level with the group indicated by the letter. To facilitate comparison, the horizontal line represents the Control group (or Choice group for Responses using voice).*

groups, for both experimental questions. Overall, the rates of response are significantly higher in the *Control* group compared to all experimental groups, aligning with *H1*. Moreover, the *Choice* group has significantly higher response rates than the push groups, which also supports *H1*. However, the *PushDictation* group yields higher rates of response than the *PushRecording* group (not supporting *H1*), although this difference only reaches statistical significance for *OQ1*.

Regarding *OQ2*, among those who provided a response, approximately 20% used voice in the *Choice* group (6% used only recording, 8% only dictation and 8% used two or more methods for *OQ1*; 10% used only recording, 5% only dictation and 2% used two or more methods for *OQ2*), in

contrast to roughly 80% in the *PushDictation* and around 60% in the *PushRecording* group. As expected (*H2*), both the *PushRecording* and *PushDictation* groups had substantially higher rates of voice input than the *Choice* group. However, we also found significantly higher use of voice input in the *PushDictation* than the *PushRecording* group, while we expected similar use in these groups, hence providing partial support for *H2*. Nevertheless, it is important to note that participants are counted as having used dictation upon clicking the dictation button (paradata). However, it is possible that some participants may have chosen to type all or part of their responses instead of using dictation, since a text box appeared on the screen when the dictation button was clicked. If we consider the self-reported usage of dic-

tation and voice recording (questions USEDICTATION and USEVOICE) for respondents who completed the full survey, a greater proportion of respondents indicate having used voice recording compared to dictation, as detailed in Appendix 4. We should thus be cautious about the result regarding differences in voice use between *PushRecording* and *PushDictation*.

## 5.2 Data quality (RQ3)

Fig. 2 presents the percentages of *Valid* answers (Coder 3 assessment), among those who provided any form of response, as well as the percentages of responses answering both parts of the question, where no problems were detected, and the average number of themes (using Coder 1 results) and characters (computed automatically), among those providing valid answers. It also presents how these indicators vary between participants who answered using text (red dots) or voice (blue dots), within each of the main groups of interest (*Choice*, *PushDictation*, and *PushRecording*). Full results (including sample sizes) are available in SOM5.

The percentages of valid responses fall within the range of 80–90%. For both experimental questions, there were only small, non-significant differences in the percentages of valid responses. Only for *OQ1*, the *Choice* and *PushRecording* groups showed significantly higher percentages of valid responses compared to the *Control*.

Turning to respondents who provided valid answers, fewer than 20% answered both parts of a given experimental question, with generally even lower levels for *OQ2*. However, no significant differences are found across groups. It is worth noting that if considering the codes of Coder 2, two significant differences emerge: between the *Control* and *PushRecording* groups (in favor of *PushRecording*), and between the *Choice* and *PushDictation* groups (in favor of *PushDictation*). Despite these differences between coders, we generally observe few reliable differences between groups in the proportion who answered both parts of the questions.

Regarding the problems detected, we observed (for both coders) that while few problems were detected in the *Control* and *Choice* groups, more problems were identified in the *PushDictation* and especially the *PushRecording* groups for both questions. However, most of these problems resulted from parts of the answers being repeated (for example, in the first experimental question, 13% of the answers in the *PushRecording* group and 8% in the *PushDictation* group exhibited repetitions; see SOM5). Importantly, these repetitions do not really reduce data quality; they primarily make the answers more cluttered, i.e., harder to read or understand because they contain unnecessary duplication,

but this does not imply that the information is incorrect or of lower quality and this does not result in any loss of information.

Furthermore, respondents in the *PushRecording* group tend to include slightly more themes in their answers compared to those in the other groups, particularly for *OQ2*. Indeed, for *OQ1*, we only observe one significant difference between the *PushRecording* and *Control* groups. Besides, this difference is not statistically significant when considering the results from Coder 2. In contrast, for *OQ2*, we detect significantly higher mean numbers of themes in the *PushDictation* and *PushRecording* groups when compared to the *Control* group, but also in the *PushRecording* group when compared to the *PushDictation* group. When taking into account the results provided by Coder 2, similar results are found.

Lastly, when considering the average number of characters, we observe a significant increase in all treatment groups compared to the *Control* group. In addition, the *PushRecording* group has significantly longer answers than the other experimental groups, particularly for the second question. Indeed, in *OQ2*, the average number of characters decreases compared to the average number of characters in *OQ1* in all groups except the *PushRecording* one.

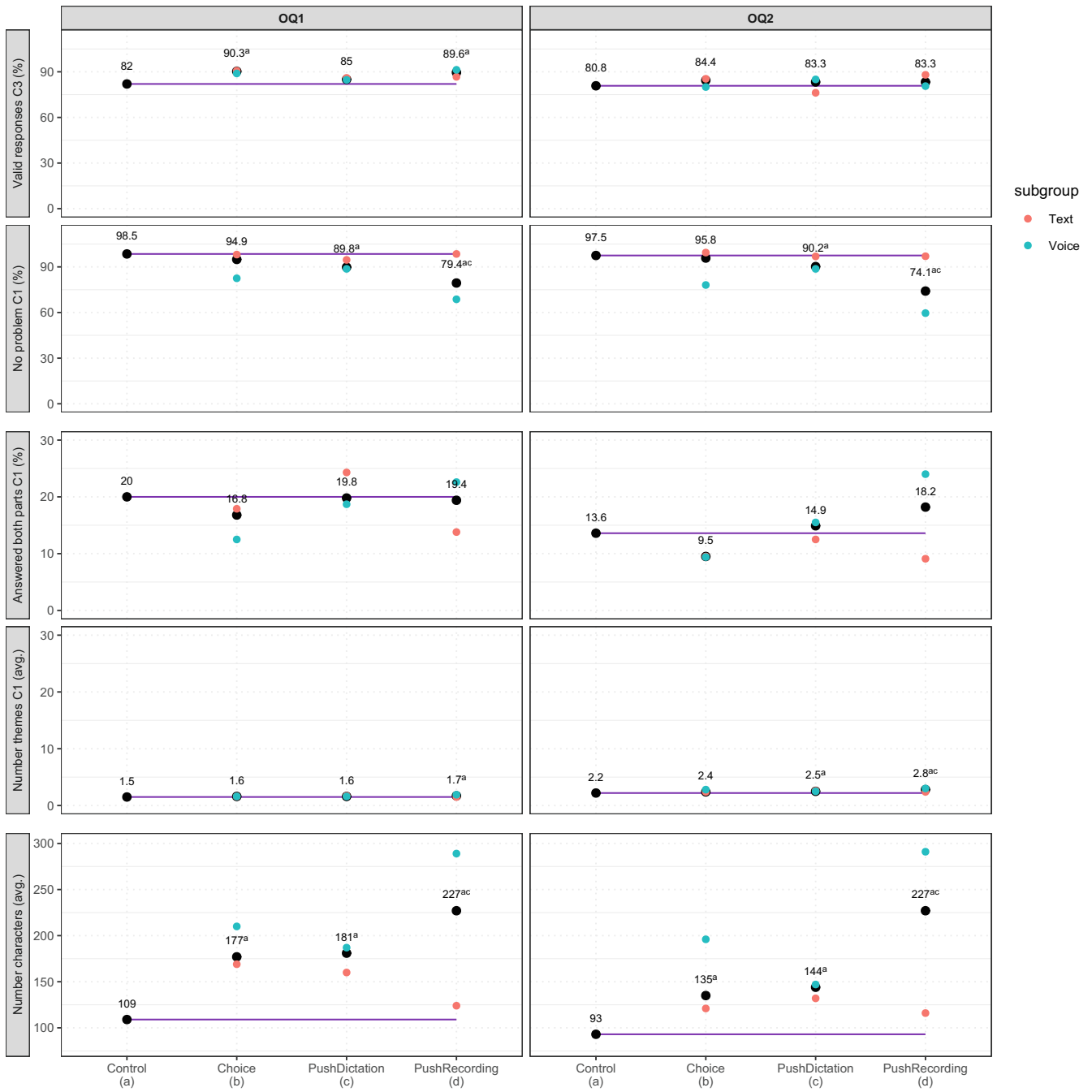
Taken together, these findings show partial support for *H3* (Quality *PushRecording* > Quality *PushDictation* > Quality *Choice* > Quality *Control*).

## 6 Discussion

This study explored how to enhance the use of two voice inputs (dictation and voice recording) to answer open-ended narrative questions in web surveys when offered in a sequential (push) or concurrent (choice) approach. To do so, we compared *PushDictation*, *PushRecording*, *Choice* and *Control* groups, in terms of overall response rates to two experimental open questions (*RQ1*), rates of voice input usage (*RQ2*), and data quality (*RQ3*), using data from a survey about nursing homes conducted in Spain in 2023 in the Netquest opt-in online panel.

### 6.1 Summary of results

Regarding overall response to the open questions (*RQ1*), the *Control* group demonstrated significantly higher response rates compared to the experimental groups, aligning with *H1*. Thus, even if a text option was always available, proposing voice inputs decreased the overall rates of response. Moreover, *PushDictation* yielded higher response rates than *PushRecording* for *OQ1*, contrary to expectations. Revilla and Couper (2024), using the same dataset,



**Fig. 2**

*Main Indicators of Data Quality, per Experimental Question and Group Superscripts indicate statistically significant differences at the 5% level with the group indicated by the letter. C1 stands for Coder 1 and C3 for Coder 3. To facilitate comparison, the horizontal line represents the Control group.*

investigated the reasons participants gave for not using voice input. The most frequently cited reason was concern about the context in which the survey was completed (e.g., the presence of other people), followed by difficulties in expressing one’s ideas orally. In contrast, technical issues

and problems understanding the task were mentioned less often. They also found that individuals who had complete trust in the confidentiality of their responses, as well as those who already used voice input in their daily lives, were significantly more likely to choose voice input. However, no

significant effects were observed for key sociodemographic variables, including gender, age, and education.

Regarding the use of voice input (*RQ2*), significant differences emerged between *Choice* and both push groups, but also between the two push groups: the *PushDictation* group exhibited higher rates of voice input usage compared to the *PushRecording* one, providing partial support for *H2*. However, caution is warranted in interpreting these results, as there are differences in the use of dictation as measured by the paradata and self-reports.

As for data quality (*RQ3*), the *Control* group exhibited the lowest proportion of valid answers, with significant differences for *OQ1* compared to the *Choice* and *PushRecording* groups. However, fewer than 20% of participants answered both parts of a question, with no significant differences between groups. Problems detected were more prevalent in both push groups. Importantly, these problems were mainly repetitions, which did not compromise the interpretation of the answers, but made answers more cluttered. Furthermore, the average number of themes mentioned was significantly higher in the *PushRecording* group. Finally, the average number of characters in the responses was significantly higher in all experimental groups, especially in the *PushRecording* one. All in all, partial support is found for *H3*.

## 6.2 Limitations

Caution is advised when interpreting these findings due to certain study limitations. Factors such as the specific topic (opinions about nursing homes), question type (probes that request an explanation for prior answers) and complexity (high), the study location (Spain), and the sample source (opt-in panel) can all influence the results. Thus, we should be cautious about generalizing these results. Further research is needed to assess their robustness in diverse contexts. Moreover, both paradata and self-reports may contain errors, which can affect the results, especially in the case of dictation. Revilla and Couper (2024) show that respondents who reported using dictation or voice recording also experienced different problems. In particular, 22% of participants in the *PushDictation* group reported technical issues, compared to only 7% in the *PushRecording* group. These difficulties may have affected the quality of the responses. The transcription tool used for voice recordings (Vosk) may also have introduced errors, potentially influencing the findings related to data quality (Meitinger et al., 2024; Höhne et al., 2025). In addition, given that each of the indicators used to assess data quality has its limitations, it is important to consider them jointly to obtain a more comprehensive evaluation. Moreover, variations in data quality assessment among coders exist. However, these differences tend to im-

pact absolute values more than group comparisons. Automatic coding (e.g., using machine learning techniques) could be tested in future research.

## 6.3 Practical implications

Despite the acknowledged limitations, our findings provide valuable insights into the use of voice input to answer open questions in web surveys, both in terms of response to open questions and data quality.

Firstly, the results indicate that, despite always offering a text option, introducing voice input decreases the overall rates of response to open questions. This aligns with prior research and suggests that all strategies implemented so far to maximize response using voice, such as enhanced instructions or the development of more user-friendly tools, remain insufficient. This may be due to the fact that the main obstacles to adopting voice input—contextual constraints and difficulties with oral expression—are beyond the control of researchers (Revilla and Couper, 2024). However, there remains a need to explore innovative approaches to improve participant engagement and completion rates. One potential avenue involves testing the effectiveness of additional incentives for respondents providing voice answers.

Secondly, the study reveals that the lower level of response is somewhat offset by slightly higher data quality, especially in the *PushRecording* group. It is worth noting also that the benefits of using voice input in terms of improved data quality might be more pronounced when multiple open-ended questions need to be asked (since slightly more positive results are found for *OQ2* than *OQ1*). Completion times (measured using the focus time, i.e., the duration the page containing the question was active) are also significantly shorter for the experimental groups compared to the *Control* group, even taking into account the potential follow-up for those skipping the question in a first place: for instance, for *OQ1*, the median completion time is 27 s in the *PushRecording*, 31 s in the *PushDictation* and the *Choice* groups, and 61 s in the *Control* group (see SOM5 for full results). Proposing voice inputs can thus help reduce survey duration, which in turn could positively affect respondents' experiences.

Thirdly, our results show that how voice input is offered (push or choice) and the type of voice input (dictation or voice recording) matter. Thus, understanding nuanced differences in voice input use among different approaches is crucial. Recognizing the strengths and limitations of each approach will help researchers develop more effective surveys.

Overall, our results are consistent with the literature on voice answers in web surveys. They suggest that voice input features, especially voice recording proposed in a push-to-

voice design, appear promising for enhancing data quality, encouraging longer and more detailed responses, and thus allowing researchers to get richer insights to open narrative questions in web surveys. However, persistent challenges, especially in terms of response, remain, and all the efforts done so far have not been able to fully address this issue. Thus, as Lenzner et al. (2024) aptly state, “we must conclude that there is yet little evidence that justifies a shift from the written to the oral communication mode” (p. 1314) in the context of web survey questions. Nevertheless, increases in the use of voice input in everyday life and technical improvements in voice input technologies offer hope for the future use of voice input for eliciting responses to open questions in web surveys.

**Acknowledgements** The authors are very grateful to Ixchel Perez Duran for all her work in designing the questionnaire for this study and preparing the coding guidelines, and to Maria Paula Acuña Pardo and Ksenija Ivanovic for their hard work coding the answers to the open-ended questions. Moreover, the authors thank Carlos Ochoa for all his support especially for preparing the R scripts, as well as Oriol Bosch and Patricia Iglesias for their helpful comments during preparation of the study. Finally, we are grateful to the reviewers for their constructive feedback, which has helped us improve the manuscript.

**Funding** This project received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (grant agreement No. 849165). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Data availability** The anonymized datasets and all the scripts used for their analysis are accessible in OSF: <https://osf.io/9ypgk/>. The Appendices and all the supplementary online materials (SOM) for the paper are also available in the same folder.

## References

- Agresti, A. (1992). A survey of exact inference for contingency tables. *Statistical Science*, 7(1), 131–153. <https://doi.org/10.1214/ss/1177011454>.
- Billiet, J. (2016). What does measurement mean in a survey context? In C. Wolf, D. Joye, T.W. Smith & Y. Fu (Eds.), *The sage handbook of survey methodology* (pp. 193–209). SAGE.
- Cornesse, C., Felderer, B., Fikel, M., Krieger, U., & Blom, A.G. (2021). Recruiting a probability-based Online panel via postal mail: experimental evidence. *Social Science Computer Review*, 40(5), 1259–1284. <https://doi.org/10.1177/08944393211006059>.
- Deloitte (2018). *2018 Global mobile consumer survey: US edition. A new era in mobile continues*. <https://www2.deloitte.com/tr/en/pages/technology-media-and-telecommunications/articles/global-mobile-consumer-survey-us-edition.html>
- Dillman, D. A. (2017). The promise and challenge of pushing respondents to the web in mixed-mode surveys. *Survey Methodology*, 43(1), 3–30.
- Gavras, K. (2019). *Voice recording in mobile web surveys—evidence from an experiment on open-ended responses to the ‘final comment*. Paper presented at the General Online Research conference, Cologne.
- Gavras, K., & Höhne, J.K. (2022). Evaluating political parties: criterion validity of open questions with requests for text and voice answers. *International Journal of Social Research Methodology*, 25(1), 135–141. <https://doi.org/10.1080/13645579.2020.1860279>.
- 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*, 185(3), 872–890. <https://doi.org/10.1111/rssa.12807>.
- Höhne, J.K. (2021). Are respondents ready for audio and voice communication channels in online surveys? *International Journal of Social Research Methodology*, 26(3), 335–342. <https://doi.org/10.1080/13645579.2021.1987121>.
- Höhne, J.K., Schlosser, S., & Krebs, D. (2017). Investigating cognitive effort and response quality of question formats in web surveys using paradata. *Field Methods*, 29(4), 365–382. <https://doi.org/10.1177/1525822x17710640>.
- Höhne, J. K., Gavras, K., & Claassen, J. (2024). Typing or speaking? Comparing text and voice answers to open questions on sensitive topics in smartphone surveys. *Social Science Computer Review*, 42(4), 1066–1085. <https://doi.org/10.1177/08944393231160961>.
- Höhne, J.K., Lenzner, T., & Claassen, J. (2025). Automatic speech-to-text transcription: evidence from a smartphone survey with voice answers. *International Journal of Social Research Methodology*. <https://doi.org/10.1080/13645579.2024.2443633>.
- Lazarsfeld, P.F. (1935). The art of asking WHY in marketing research: three principles underlying the formulation of questionnaires. *National Marketing Review*, 1(1), 32–43. <http://www.jstor.org/stable/4291274>.
- Lazarsfeld, P.F. (1944). The controversy over detailed interviews—an offer for negotiation. *Public Opinion Quarterly*, 8(1), 38–60. <https://doi.org/10.1086/265666>.
- Lenzner, T., & Höhne, J.K. (2022). Who is willing to use audio and voice inputs in Smartphone surveys, and why? *International Journal of Market Research*, 64(5), 594–610. <https://doi.org/10.1177/14707853221084213>.

- Lenzner, T., Höhne, J.K., & Gavras, K. (2024). Innovating web probing: comparing written and oral answers to open-ended probing questions in a Smartphone survey. *Journal of Survey Statistics and Methodology*, 12(5), 1295–1317. <https://doi.org/10.1093/jssam/smae031>.
- Lütters, H., Friedrich-Frekxa, M., & Egger, M. (2018). *Effects of speech assistance in Online questionnaires*. Paper presented at GOR Conference, Cologne.
- Medway, R., & Fulton, J. (2012). When more gets you less: a meta-analysis of the effect of concurrent web options on mail survey response rates. *Public Opinion Quarterly*, 76(4), 733–746. <https://doi.org/10.1093/poq/nfs047>.
- Meitinger, K., van der Sluis, S., & Schonlau, M. (2022). *Implementing voice-recordings in a probability-based panel: what we learnt so far*. Paper presented at the CIPHER Virtual Conference. [https://cesr.usc.edu/cipher\\_2022](https://cesr.usc.edu/cipher_2022)
- Meitinger, K., van der Sluis, S., & Schonlau, M. (2024). Keep the noise down: on the performance of automatic speech recognition of voice-recordings in web surveys. *Survey Practice*. <https://doi.org/10.29115/SP-2023-0022>.
- Neuert, C., Meitinger, K., Behr, D., & Schonlau, M. (2021). Editorial: the use of open-ended questions in surveys. *methods, data, analyses*, 15(1), 3–6. <https://nbnresolving.org/urn:nbn:de:0168-ssoar-73172-3>.
- Patrick, M.E., Couper, M.P., Laetz, V.B., Schulenberg, J.E., O'Malley, P.M., Johnston, L., & Miech, R.A. (2018). A sequential mixed mode experiment in the U.S. National monitoring the future study. *Journal of Survey Statistics and Methodology*, 6(1), 72–97. <https://doi.org/10.1093/jssam/smx011>.
- PWC (2018). *onsumer intelligence series: prepare for the voice revolution*. <https://www.pwc.com/us/en/services/consulting/library/consumer-intelligence-series/voice-assistants.html>
- R Core Team (2023). *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org/>
- Revilla, M. (2022). How to enhance web survey data using metered, geolocation, visual and voice data? *Survey Research Methods*, 16(1), 1–12. <https://doi.org/10.18148/srm/2022.v16i1.8013>.
- Revilla, M., & Couper, M.P. (2021). Improving the use of voice recording in a Smartphone survey. *Social Science Computer Review*, 39(6), 1159–1178. <https://doi.org/10.1177/0894439319888708>.
- Revilla, M., & Couper, M.P. (2024). Exploring respondents' problems and evaluation in a survey proposing voice inputs. *methods, data, analyses*, 18(2), 263–280. <https://doi.org/10.12758/mda.2024.06>.
- Revilla, M., Couper, M.P., & Ochoa, C. (2018). Giving respondents voice? The feasibility of voice input for mobile web surveys. *Survey Practice*. <https://doi.org/10.29115/SP-2018-0007>.
- 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. <https://doi.org/10.1177/0894439318810715>.
- Revilla, M., Iglesias, P., Ochoa, C., & Antón, D. (2022). *WebdataVoice: a tool for dictation or recording of voice answers in the frame of web surveys*. OSF. <https://doi.org/10.17605/OSF.IO/B2WYZ>.
- Schonlau, M., & Couper, M.P. (2016). Semi-automated categorization of open-ended questions. *Survey Research Methods*, 10, 143–152. <https://doi.org/10.18148/srm/2016.v10i2.6213>.
- Schonlau, M., Gweon, H., & Wenemark, M. (2021). Automatic classification of open-ended questions: check-all-that-apply questions. *Social Science Computer Review*, 39(4), 562–572. <https://doi.org/10.1177/0894439319869210>.
- Schuman, H. (1966). The random probe: a technique for evaluating the validity of closed questions. *American Sociological Review*, 31(2), 218–222. <https://doi.org/10.2307/2090907>.
- Schwartz, B. (2004). *The paradox of choice: why more is less*. Harper Perennial.
- Singer, E., & Couper, M.P. (2017). Some methodological uses of responses to open question and other verbatim comments in quantitative surveys. *methods, data, analyses*, 11(2), 115–134. <https://doi.org/10.12758/mda.2017.01>.
- Thaler, R.H., & Sunstein, C.R. (2008). *Nudge: improving decisions about health, wealth, and happiness*. Penguin Press.
- Wardini, J. (2023). *Voice search statistics 2023: smart speakers, VA, and users: Serpwatch*. Serpwatch.io. <https://serpwatch.io/blog/voice-search-statistics/> (Created 14.03.).