methods, data, analyses | Vol. 18(2), 2024, pp. 263-280 DOI: 10.12758/mda.2024.06

Exploring Respondents' Problems and Evaluation in a Survey Proposing Voice Inputs

Melanie Revilla[1] & Mick P. Couper[2]

[1] *Institut Barcelona d'Estudis Internacionals (IBEI)*

[2] *Survey Research Center, Institute for Social Research, University of Michigan*

Abstract

Integrating voice inputs into web surveys holds the potential for various benefits, including eliciting more comprehensive and elaborate responses or extracting additional information from vocal tones and ambient sounds. Nevertheless, important challenges persist, including technical problems, privacy concerns, and low participation rates. Given the limited knowledge on this subject, this research note addresses four research questions, distinguishing between two voice input methods (dictation and voice recording) and two approaches to presenting them (providing a choice, or pushing respondents toward voice inputs, with a text alternative offered only in the absence of response): *RQ1***.** What reasons are provided for not opting for voice inputs when they are offered? *RQ2***.** Which variables are associated with the reported use of voice inputs? *RQ3***.** What challenges do individuals answering through voice inputs report? And *RQ4***.** How do respondents evaluate the different methods of answering they employed?

Drawing on data from a survey on nursing homes conducted in February/March 2023 within the Netquest opt-in online panel in Spain (1,001 completes), where participants were offered to respond to two experimental questions through voice methods, our analyses reveal that contextual factors and the perceived challenge of oral expression are key reasons for abstaining from voice input responses. Furthermore, individuals who exhibited complete trust in the confidentiality of their responses and those already using voice input in their daily lives were significantly more likely to opt for voice inputs. Among respondents utilizing voice inputs, recurring challenges included contextual constraints and difficulties in verbal expression, alongside technical problems. Despite these hurdles, a majority of participants found answering through voice easy, although a lower proportion reported liking it. These results contribute to the limited literature and can help enhance the effectiveness of voice input surveys.

Keywords: dictation, survey question evaluation, open questions, challenges, voice recording, web surveys

© The Author(s) 2024. This is an Open Access article distributed under the terms of the Creative Commons Attribution 3.0 License. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

In recent years, the integration of voice input technology into everyday activities has become increasingly common (Deloitte, 2018). Simultaneously, an increasing number of surveys have embraced this technology to collect responses for specific questions, typically open-ended narrative questions (see "Background section").

It has been argued that offering voice input in web surveys could present a variety of potential benefits, such as eliciting richer and longer answers or permitting the extraction of additional information from nuances in the tone of voice or from ambient noise (Höhne et al., 2023; Revilla, 2022; Singer & Couper, 2017). Nevertheless, persistent challenges, including technical issues, data protection and privacy concerns, and low participation rates (see e.g., Revilla & Couper, 2021), underscore the need for a comprehensive understanding of the factors influencing engagement with voice input for open questions. Gaining insights into how respondents perceive these novel methods of answering is essential for enhancing their overall effectiveness.

This research note presents the outcomes of a web survey (both mobile and PC devices were allowed) on opinions about nursing homes conducted in February/March 2023 within the Netquest opt-in online panel in Spain (N=1,001 completes). It focuses on the problems and challenges encountered by participants, as well as their evaluations, when two voice inputs are proposed as a response method for open narrative questions[1]:

- *Dictation* (also called *Automatic Speech Recognition* or *ASR*): Respondents speak, and their voice is instantly transcribed into text on their device's screen. Respondents can then edit the transcriptions using their keyboard.

1 Data from this same survey have also been used in a different paper focused on comparing participation and data quality across the different experimental groups presented in Table 1.

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).

Data availability
The anonymized dataset, R script, and supplementary online materials (SOM) are accessible at: https://osf.io/3crsg/?view_only=52bc495d5007463faa8a6e56bad9bf97

Acknowledgements
We are very grateful to Ixchel Perez Duran for all her work in designing the questionnaire, and to Maria Paula Acuña Pardo, Oriol Bosch, Patricia Iglesias, Ksenija Ivanovic, Carlos Ochoa and Tracy Silva for their assistance at different stages of this study's preparation. We also want to thank two anonymous reviewers and the journal editor who provided helpful feedback on an earlier draft of this manuscript.

Direct correspondence to
Melanie Revilla, Institut Barcelona d'Estudis Internacionals (IBEI)
E-mail: mrevilla@ibei.org

- *Voice recording*: Respondents are asked to record their voice. They can create and review multiple audio files before submitting their responses.

Furthermore, two approaches to propose these voice inputs are compared:

- *Push*: Respondents are initially presented with only one of the voice input methods. If they skip the question without answering, the question is repeated with a message emphasizing the importance of their responses, and a text alternative.
- *Choice*: Respondents are offered three options: answering by typing in a textbox, by dictating, or by recording their voice. They can choose whichever they prefer, and can use multiple methods.

Background

Some studies have explored respondents' stated willingness to use voice input to answer survey questions (Höhne, 2021; Lenzner & Höhne, 2022; Revilla et al., 2018). Others have actually asked respondents to answer open-ended narrative questions through voice input, using experimental designs. For instance, studies by Lütters and colleagues (2018) and Meitinger and Schonlau (2022) randomly assigned participants to a voice-only group, a choice group (allowing selection between voice or text), and a text-only group. Other studies compared voice recording and text responses (Gavras, 2019; Gavras & Höhne, 2022; Gavras et al., 2022; Höhne & Gavras, 2022). Revilla et al. (2020) compared text with dictation for iOS respondents and text with voice recording for Android respondents.

The findings of these studies indicate that participation tends to be lower when respondents are offered voice input methods, even when given the option to choose between voice and text. For instance, in the study by Lütters et al. (2018), 49% of the participants answered in the voice-only group, and 54% in the choice group, compared with 94% in the text-only group. Further, in cases where a choice is available, a significant majority of participants opted for the text option (e.g., 93.9% in Meitinger et al., 2022).

Nevertheless, there are indications that voice answers could have higher quality, with significantly longer answers and a greater variety of words than text responses (e.g., Höhne & Gavras, 2022). Also, certain underrepresented groups (e.g., older or lower-educated individuals) may be encouraged to respond to open-ended questions when voice inputs are proposed (Gavras, 2019).

Some studies also explored respondents' evaluations and experiences, finding that participants are more positive about text than voice answers (Lütters et al., 2018; Revilla et al., 2020).

In addition, previous research suggests that participation, data quality and respondents' evaluation of voice input methods might be affected by partici-

pants' characteristics. For example, Revilla and Couper (2021) found that gender, education, mother tongue, using voice input in daily life, trust in anonymity, multitasking, and answering from home significantly affected at least one of their dependent variables related to nonresponse, data quality and evaluation of voice recording.

Finally, Revilla and Couper (2021) tried to improve the voice input option on Android devices. Providing different instructions to help respondents using the voice recording tool had minimal impact on uptake rates. A filter question to determine whether respondents were in a setting that permitted voice recording, directing others to text input, was more successful. However, technical issues and low participation persisted.

Overall, the available studies remain sparse, and in particular, little is known about possible differences between dictation and voice recording, and between different approaches to presenting the voice input options to participants.

Research Questions and Contribution

To fill these gaps, this research note addresses four research questions regarding the integration of voice inputs for responding to open-ended narrative questions:

RQ1) What reasons are provided for not using voice inputs when they are offered?

RQ2) Which variables are associated with the reported use of voice inputs?

RQ3) What challenges do individuals answering through voice inputs report?

RQ4) How do respondents evaluate different methods of answering open questions?

This study contributes to the limited literature on utilizing voice inputs in web surveys in several ways. Firstly, it provides fresh empirical evidence on two distinct voice input methods: dictation and voice recording.

While both are voice input methods, voice recording has been studied more frequently than dictation. Besides, the methods exhibit some key differences that may affect respondents' experience with and evaluation of the methods. Notably, although respondents can review their answers in both methods, the process differs: editing the transcription versus recording a full answer again. Furthermore, privacy concerns can be less prevalent for dictation than for voice recording, since the voice file is not shared with the fieldwork company or researchers, fostering a sense of confidentiality. The cognitive load can also differ since in one case, visual support can be provided and answers can be reviewed by reading while in the other respondents can only listen to the audio

files. Thus, we expect that different reasons could be provided for not using the two kinds of voice inputs (e.g., more aspects related to privacy issues could be mentioned in voice recording) and that different variables could be associated with participation in questions proposing dictation versus recording. Similarly, differences are expected in the prevalence of the problems faced by the participants and in their evaluation of such methods.

Second, this study contributes by distinguishing between two approaches of offering the dictation and voice recording options (*Push* and *Choice*).

The way of offering the voice input options could affect the results to the different research questions: in particular, the "choice" method is expected to yield fewer reported problems/challenges and slightly more positive overall evaluations, since participants can select what they prefer.

Third, this is the first study to collect voice data through the *WebdataVoice* tool (Revilla et al., 2022), which allows for either dictation or voice recording on Android and iOS devices as well as PCs and has been designed to be user-friendly. Using this new tool could produce more favorable results compared to previous studies, especially fewer technical and understanding problems, which in turn could lead to more positive evaluations.

Overall, insights from this study can help researchers and survey designers tailor voice input surveys to mitigate reported problems/challenges and enhance participant evaluations.

Method and Data

Data Collection

Data were collected between February 22 and March 30, 2023, in the Netquest online opt-in panel in Spain. The objective was to obtain 1,000 participants completing the full survey. Quotas for gender and age, education, and autonomous community were defined to match the adult online population in Spain (under 75 years old) according to the National Statistics Institute.

Of the 4,789 panelists invited, 1,860 started the survey. Of those, 577 were excluded for various reasons (170 did not provide consent, 185 quotas full, 17 did not pass basic fraud checks and 205 reported unfamiliarity with nursing homes), leading to 1,170 panelists answering the first survey question after all the filter/quota questions. Another 169 panelists broke-off during the survey, meaning that 1,001 completed the full survey. The average age of those finishing the survey is 47 years old, 50.5% are female, and 35.0% have a higher education degree. On average they have been in the Netquest panel for six years (median=5.7), and have completed 157 surveys (median=141). Most participants used smartphones (73.6%) to respond. The average survey completion time was 9.1 minutes.

Questionnaire

The online questionnaire included more than 80 questions optimized for mobile devices but accessible from any device. None of the respondents got all questions, due to routing. The full questionnaire in Spanish and its English translation are available in the Supplementary Online Material (SOM) 1.

Respondents could continue without answering the questions, except those used to control quotas and filter/tailor other questions. Following the panel's usual practice, going back was not allowed.

The survey mainly dealt with perceptions of nursing homes in Spain (e.g., to what extent they trust them or consider that they are transparent) but also included a block of questions about political opinions (e.g., trust in the government), 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., how easy or difficult it was to answer open-ended questions using different methods).

The survey included the following two narrative open-ended questions asking respondents to explain why they selected a given answer in the previous question:

- ***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[2]] 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.
- ***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.

For these two questions, an experimental design was used: respondents were assigned to four groups, as presented in Table 1: a *Control* group, two "push" groups (*PushDictation* and *PushRecording*) and a *Choice* group where all three options were offered. Detailed instructions for both experimental questions can be found in SOM1. Screenshots of these questions (together with the question just before and the follow-up when relevant) for each of these groups are provided in SOM2.

2 This was tailored for each respondent depending on the previous answer.

Table 1 Experimental groups (same group for both WHYTRANSP and WHYTRUST)

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 - Recording - Text

In this research note, we are mainly interested in questions asking respondents a) their reasons for not using voice input methods to answer ***WHYTRANSP*** and ***WHYTRUST,*** b) which kinds of problems they faced to use these answering methods and c) how they evaluate these new ways of answering and the conventional (text) one[3]. We also use questions about the respondents' profile (socio-demographics and attitudinal variables) to answer ***RQ2*** (see below).

Analyses

To answer ***RQ1,*** we report the answers to a question asking respondents[4] to select all that apply of the reasons for not using voice inputs in the following list: "I preferred another of the alternatives" (only in *Choice* group), "The device I am using to answer the survey does not have a microphone", "I tried, but I had technical problems", "I tried, but I had problems understanding the function", "I did not want to use it because of the context (e.g., I was around other people)", "I did not want to use it because I found it difficult to express myself orally", "Other reasons". The proportions of panelists selecting each option are reported for both dictation and voice recording, separating the push from the choice groups.

To assess whether there are differences between dictation and voice recording, we compare:

- *PushDictation* to *PushRecording*
- *ChoiceDictation* (i.e., respondents from the *Choice* group who have stated they used dictation – whether alone or in combination with other methods) to *ChoiceRecording* (i.e., respondents from the *Choice* group who have stated they used recording – whether alone or in combination with other methods).

3 Another narrative open-ended question asking about the perceived quality of the nursing homes was presented to the panelists. This question was placed before the two experimental ones, and all respondents were asked to answer it using a text-box.

4 This question was asked only to those who stated "No, I never used the dictation/voice recording tool" in the questions USEDDICTATION/ USEDVOICE (see SOM1 and Appendix A).

To assess whether there are differences between push and choice groups, we compare:

- *PushDictation* to *ChoiceDictation*
- *PushRecording* to *ChoiceRecording*.

We test whether differences in proportions across groups are significant at the 5% level using exact Fisher tests.

To answer ***RQ2***, logistic regressions analyses were conducted. The dependent variables are the use of dictation or voice recording reported in the questions USEDDICTATION and USEDVOICE (see Appendix A), grouping the two "yes" options to create indicators where 1 indicates that dictation or voice recording has been used, and 0 otherwise.

The key independent variable is the experimental group: push or choice (push being used as reference category). We control for the following sociodemographic characteristics: gender, age (two dummies for respondents having less than 30 and more than 60), and education level (two dummies for low and high education).

Additionally, based on previous research (Revilla & Couper, 2021) but also, since little is known yet, logical reasoning about which factors might influence the reported use of dictation and voice recording and data availability, we include the following set of independent variables:

- Having Spanish as a mother tongue (dummy): Non-native speakers might exhibit more reluctance to answer through voice options (e.g., because of concerns about their accent).
- Social trust (values ranging from "1-You can't be too careful" to "5-Most people can be trusted") and trust in the confidentiality of answers (dummy, 1 = complete trust and 0 = the rest): Higher levels of trust may be associated with lower privacy concerns, and, consequently, increased use of voice inputs.
- Comfort in using new technologies (dummy, 1 = "quite" to "completely comfortable", and 0 = "not at all" or "little comfortable"): Being comfortable with new technologies is expected to be associated with higher participation through voice inputs.
- Lack of awareness of voice inputs existence (one dummy for each type of voice input) and occasional use of voice inputs in daily life (one dummy for each type of voice input[5]): Distinguishing between these variables is essential, as individuals aware of voice inputs but not using them are likely to dislike such features, while those unaware might be positive about using them once they are informed about these possibilities. However, the lack of awareness regarding voice inputs suggests a potential lack of technological knowledge,

5 The four dummies for lack of awareness and use in daily life are created using FREQDICTATION and FREQVOICE.

which, in turn, may result in increased difficulties in utilizing the voice tools and subsequently lower voice participation. Overall, we expect that both individuals unaware of voice inputs and those aware but never using them are less likely to participate through voice.

- Device type (1 = smartphones/tablets, 0 = PCs): Since PCs are not always equipped with microphones, PC respondents might participate less using voice inputs.
- Place of completion (1 = home, 0 = other): Responding from home is expected to be associated with higher voice participation (e.g., lower privacy concerns at home).
- Presence of third parties (1 = people around, 0 = alone): The presence of third parties is expected to decrease voice participation, due to privacy concerns.

We report the odds ratios (OR) and 95% confidence intervals (CI) of these two logistic regressions (dummies based on USEDDICTATION and USEDRECORDING).

To answer ***RQ3***, we first report the proportion of respondents (within those who stated having used the voice input methods, see Appendix A) who reported having faced the following problems: "Technical problems (e.g., microphone not working)", "Problems understanding the function", "I could not speak freely because of the context (e.g., I was around other people)", "I found it difficult to express my answers orally", or "None of these". The proportions of panelists selecting each option are presented for both dictation and voice recording, separating the push from the choice groups. Tests of significance are implemented, in a similar way as for ***RQ1***.

Finally, to answer ***RQ4***, we report the proportions of respondents who found it easy/difficult and who dis/liked using the voice input methods and answering by text. While these questions[6] were all asked using a five-point bipolar scale, for the analyses we combined the two positive (e.g., extremely and quite easy) and the two negative (e.g., extremely and quite difficult) answer categories, thus presenting three categories (positive, neutral, negative). Again, tests of significance are implemented as in previous analyses, although this time we additionally test for significance of the differences between text and the four other groups.

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

6 See questions EASYDICTATION, EASYVOICE, EASYTEXT, LIKEDICTATION, LIKEVOICE and LIKETEXT in the questionnaire (SOM1).

Results

Stated Reasons for not Using Voice Inputs (RQ1)

Table 2 presents the proportions of respondents who selected each of the reasons for not using voice inputs when offered, distinguishing dictation and voice recording, and push and choice groups.

Table 2 Reasons for not using dictation or voice recording for those who stated not having used them (% of those answering the question)

Reasons for not using voice input	Dictation		Recording	
Group	Push (n=130)	Choice (n=186)	Push (n=107)	Choice (n=169)
Prefer another alternative	NA	57.5	NA	51.5
Concerns about context	24.6	16.7	**30.8**	**17.2**
Hard to express orally	21.5	13.4	22.4	17.2
No microphone	**20.0**	**7.0**	**17.8**	**6.5**
Technical problems	**18.5***	**1.6**	7.5*	3.5
Problems understanding the function	**6.9**	**1.1**	4.7	1.2
Other reason	**17.7**	**9.7**	**22.4**	**9.5**

Note. The sum is not 100 because respondents could select several reasons. Bold numbers indicate significant differences between push and choice groups (5% level) within methods. Stars (*) indicate significant differences between dictation and recording (*PushDictation* vs *PushRecording* or *ChoiceDictation* vs *ChoiceRecording*). P-values of all tests are provided in SOM3.

First, focusing on the reasons offered to all groups and excluding the "other" option, the ranking is similar for all four groups: concerns about the context is the main reason for not using voice inputs, followed by the difficulty of expressing one's ideas orally. Technical and understanding problems, in contrast, are reported less often.

However, important differences exist across groups. In particular, technical problems are reported as a reason for not using voice input by a much larger proportion of respondents in the *PushDictation* group, compared to the others.

Furthermore, in the choice groups, more than half of the respondents mentioned that they "preferred another alternative". Since this option was not offered for the push groups, this creates important differences in the reported levels of other reasons between push and choice groups.

Variables Associated with the Use of Voice Inputs (RQ2)

Moving to ***RQ2***, Table 3 presents the OR and 95% CI of the two logistic regressions, with reported use of dictation or voice recording to answer at least one experimental question as dependent variables.

Table 3 OR and 95% CI of the logistic regressions

DV: reported using…	Dictation			Recording		
	OR	2.5%	97.5%	OR	2.5%	97.5%
Choice group	**0.31**	0.20	0.47	**0.22**	0.14	0.33
Female	1.44	0.95	2.21	1.14	0.76	1.73
Age (Chi^2=5.7 for Dictation and 0.4 for Recording, d.f.=2, p>.05 in both cases)						
30 or less	0.54	0.27	1.02	0.84	0.46	1.50
31 to 59	-	-	-	-	-	-
60 or more	1.39	0.80	2.38	0.92	0.54	1.54
Education (Chi^2=3.7 for Dictation and 2.0 for Recording, d.f.=2, p>.05 in both cases)						
Low	0.64	0.38	1.07	1.36	0.80	2.31
Middle	-	-	-	-	-	-
High	0.63	0.36	1.08	1.00	0.58	1.74
Spanish native language	1.14	0.52	2.64	1.44	0.71	2.96
Social trust	1.08	0.90	1.30	0.98	0.81	1.18
Complete trust in confidentiality	**1.83**	1.13	2.98	**2.13**	1.32	3.47
Comfortable with technology	1.36	0.74	2.56	1.19	0.70	2.03
Not aware dictation/recording	1.14	0.57	2.24	0.98	0.24	3.35
Use dictation/recording in daily life	**4.21**	2.68	6.71	**2.57**	1.57	4.27
Answer from mobile	**2.01**	1.20	3.42	1.29	0.76	2.21
Answer from home	1.11	0.65	1.89	0.75	0.46	1.25
People around	0.68	0.41	1.10	0.86	0.52	1.40
Intercept	**0.18**	0.05	0.69	0.35	0.10	1.18
AIC		*577.34*			*585.93*	
N		*490*			*473*	

Note: Bold numbers indicate statistically significant odds ratios.

The use of both voice inputs is influenced by several factors. Firstly, the method employed to offer the voice inputs plays an important role. As expected, individuals provided with a choice are less inclined to use voice inputs compared to those in the push groups. Secondly, individuals who completely trust that their

answers are treated confidentially are more likely to use voice inputs. Moreover, respondents who already use voice inputs in their daily lives are also more likely to employ them within the survey context. Additionally, answering from a mobile device also increases the likelihood of using dictation.

Notably, only a few variables exhibit significant effects. In particular, despite the survey context being the most frequently cited reason for not using voice inputs (excluding the "prefer another alternative," which was exclusively proposed in the *Choice* group; see Table 2), factors such as being at home and the presence of third parties do not yield significant effects. Similarly, variables that one might expect to be correlated with difficulties in articulating oral responses (e.g., non-native Spanish speakers or lower education levels) do not demonstrate significant effects. Lastly, the comfort level in using new technologies, which could be associated with understanding problems, also does not show any significant effects.

Stated Problems (RQ3)

Panelists who stated they used dictation and/or voice recording to answer to at least one of the experimental questions were asked whether they faced various problems when using these tools. Table 4 reports the proportion of respondents reporting having encountered each issue, distinguishing dictation/voice recording and push/choice groups.

Table 4 Reported problems for those who stated having used dictation or voice recording (in % of those answering the question)

Reported problems	Dictation		Recording	
Group	Push (n=120)	Choice (n=60)	Push (n=145)	Choice (n=56)
None	45.8*	50.0	64.1*	58.9
Technical problems	21.7*	16.7	6.9*	8.9
Hard to express orally	20.8*	11.7	11.0*	7.1
Could not speak freely given context	10.0	20.0	15.9	19.6
Problems understanding the function	6.7	8.3	2.8	5.4

Note. The sum is not 100 because respondents could select several reasons (except if they selected "none"). There are no significant differences between push and choice groups (5% level). Stars (*) indicate significant differences between dictation and recording (*PushDictation* vs *PushRecording* or *ChoiceDictation* vs *ChoiceRecording*). P-values of all tests are provided in SOM3.

First, a majority of respondents in the voice recording groups did not report experiencing any of the difficulties we asked about. In the dictation groups, slightly fewer than half (46% and 50%) reported encountering no issues.

In particular, the *PushDictation* group exhibited a significantly higher incidence of technical problems and greater difficulty in articulating responses orally compared to the *PushRecording* group. Furthermore, 10% to 20% of respondents (contingent on the group) reported constraints in expressing themselves freely due to contextual factors. Conversely, challenges pertaining to comprehension of tool functionality were the least frequently reported.

Evaluations (RQ4)

Finally, to answer ***RQ4***, Table 5 presents the evaluations of respondents of three ways of answering: by text (used by all respondents to answer at least one open-ended narrative question), dictation and voice recording (for those reporting using them to answer at least one question).

Table 5 Evaluation of the way of answering questions

			Dictation		Recording	
Var.	Answer categories	Text (n=1,001)	Push (n=120)	Choice (n=60)	Push (n=145)	Choice (n=56)
EASY	Easy	72.9	*63.3**	*51.7*	**77.2***	***51.8***
	Neither easy nor difficult	20.8	**15.0**	***41.7***	**13.8**	***44.6***
	Difficult	6.3	***21.7****	**6.7**	9.0*	3.6
LIKE	Liked	48.3	40.0	46.7	*38.6*	*33.9*
	Neither liked nor disliked	46.6	44.2	46.7	48.3	*62.5*
	Disliked	5.2	*15.8*	6.7	*13.1*	3.6

Note. Bold numbers indicate significant differences between push and choice groups (5% level). Stars (*) indicate significant differences between dictation and recording (*PushDictation* vs *PushRecording* or *ChoiceDictation* vs *ChoiceRecording*). Numbers in italics indicate significant differences compared to Text. P-values of all tests are provided in SOM3.

Overall, most respondents found it easy to answer (51.7% to 77.2%), in the case of text as well as in the case of voice inputs. In contrast, a minority of respondents reported liking answering in each of the ways (33.9% to 48.3%).

However, while there are no significant differences between experimental groups in how much respondents dis/liked answering in different ways, in the case of easiness, differences are observed. In particular, significantly more respondents found it difficult to answer through dictation than through voice recording. Also, respondents given a choice reported significantly more that it

was “neither easy nor difficult” to use the voice tools, compared to those in the push groups.

Conclusions

Voice input surveys offer exciting opportunities, but several challenges persist. This study provides new empirical evidence, comparing two voice input methods (dictation and voice recording) and two ways of proposing them to participants (push and choice).

Summary of Results

The results show, first, that in the *Choice* groups, the primary reason stated for not using voice input (***RQ1***) is that respondents prefer text input. Then, in all groups, follow concerns related to the context (e.g., the presence of others) and the difficulty of orally expressing one’s ideas. Although technical and understanding problems are still present, especially in the *PushDictation* group, they are reported by smaller proportions of respondents compared to other issues. Notably, the prevalence of technical and understanding issues is lower than in the study by Revilla and Couper (2021), where technical problems were reported by 12% to 25% of the respondents and understanding problems by 14% to 17% (depending on the groups; all groups used voice recording). This reduced reporting of technical and understanding issues relative to previous studies may be attributed to the use of a new tool, *WebdataVoice*, and/or to the increasing proficiency of panelists in using their devices.

Moving on to ***RQ2***, employing logistic regression analyses, we found that only a few of the tested variables exhibit significant associations with the reported use of voice inputs to answer the experimental questions: providing a choice (as opposed to pushing to voice), having complete trust in the confidentiality of the answers, already using voice inputs in daily life, and, in the case of dictation, answering through a mobile device. In contrast, other variables, such as being at home or having people around, do not show significant effects, despite the context being cited as a key reason for not using voice inputs (see ***RQ1***).

As for the challenges posed by the use of voice input tools (***RQ3***), a majority of respondents who reported using these tools did not report experiencing any of the challenges we asked about. However, in the choice groups, around 20% of respondents reported constraints associated with the context, while in the *PushDictation* group, similar proportions reported both technical problems and difficulty of expressing answers orally.

Turning to the evaluation of different answering methods (***RQ4***), namely text, dictation and voice recording, the majority of respondents found it easy to

answer in all three methods, although the specific levels varied across groups. Nevertheless, fewer participants reported liking the voice input methods. However, compared to the study by Revilla and Couper (2021), we found higher levels of liking of the tools (33.9% to 46.7% versus 22.6% to 30.8%).

Limitations and Practical Implications

This study is subject to certain limitations. Firstly, the sample size disparity among groups, particularly notable in the choice group where a small proportion opted for voice tools, might account for the limited statistical significance observed in some instances. Secondly, reliance on self-reports introduces the possibility of errors. Thirdly, we do not have detailed information on the specific nature of problems encountered, such as the specifics of "technical problems". Finally, factors such as the topic (opinions about nursing homes), question type (probes), country (Spain), and sample source (opt-in panel) could influence the results. Therefore, further research is required to evaluate their robustness in different contexts.

Despite these limitations, this research contributes to the existing literature by shedding light on the differences between dictation and voice recording, as well as between push and choice designs. Importantly, it underscores that many obstacles to the adoption of voice input extend beyond the researcher's control. The primary impediments, contextual constraints and difficulty in oral expression, are inherently beyond the purview of researchers conducting web surveys.

Nevertheless, some of the results can help enhance the effectiveness of voice input surveys. For instance, our analyses suggest that trust in the confidentiality of the answers is one of the few variables which significantly affects the use of voice inputs, in line with Revilla and Couper's (2021) results. These levels of trust could be improved by joint efforts between researchers and fieldwork companies to guarantee data protection, for example by improving transparency and security measures. Also, we found that answering through mobile devices increases the likelihood of using dictation. Thus, researchers planning to propose dictation could encourage participants to answer through mobile devices. Finally, technical and understanding problems remain, even if these do not seem to be the main obstacles to the use of voice inputs to answer survey questions. Strategies to reduce them are therefore still needed. More generally, further research in this area is warranted to uncover additional insights and refine best practices for voice-based surveys.

References

Deloitte (2018). *2018 Global mobile consumer survey: US edition. A new era in mobile continues*. Retrieved from https://www2.deloitte.com/tr/en/pages/technology-media-and-telecommunications/articles/global-mobile-consumer-survey-us-edition.html

Gavras, K. (2019, March 6–8). *Voice Recording in Mobile Web Surveys – Evidence from an Experiment on Open-Ended Responses to the 'Final Comment'*. [Paper presentation] General Online Research conference, Cologne, Germany.

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 Series A: Statistics in 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., & Gavras, K. (2022). Typing or speaking? comparing text and voice answers to open questions on sensitive topics in smartphone surveys. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4239015

Höhne, J. K., Kern, C., Gavras, K., & Schlosser, S. (2023). The sound of respondents: Predicting respondents' level of interest in questions with voice data in smartphone surveys. *Quality & Quantity*. https://doi.org/10.1007/s11135-023-01776-8

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

Lütters, H., Friedrich-Freksa, M., & Egger, M. (2018, February 28–March 2). *Effects of Speech Assistance in Online Questionnaires*. [Paper presentation] General Online Research conference, Cologne, Germany.

Meitinger, K., van der Sluis, S., & Schonlau, M. (2022, March 3–4), *Implementing Voice-Recordings in a Probability-based Panel: What We Learnt So Far.* [Paper presentation] CIPHER virtual conference https://cesr.usc.edu/cipher_2022

R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. 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., 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., Couper, M. P., & Ochoa, C. (2018). Giving Respondents Voice? The Feasibility of Voice Input for Mobile Web Surveys. *Survey Practice, 11*. https://doi.org/10.29115/SP-2018-0007

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.* [Computer software]. OSF. https://doi.org/10.17605/OSF.IO/B2WYZ

Singer, E. & Couper, M.P. (2017). Some Methodological Uses of Responses to Open Questions and Other Verbatim Comments in Quantitative Surveys. methods, data, analyses 11(2), 115-134. https://doi.org/10.12758/mda.2017.01

Appendix A

Reported Use of Dictation and Voice Recording

Table A1 presents the answers to the questions USEDDICTATION and USEDVOICE, asking respondents to report whether they used dictation (respectively, voice recording) to answer at least one of the experimental questions. Three response options were proposed: "Yes, I used only the dictation tool whenever I had this option", "Yes, I used the dictation tool, but also other options (e.g., the keyboard)", and "No, I never used the dictation tool" (same with voice recording).

Table A1 Reported use of voice inputs per group (in %)

Reported use of...	Dictation		Voice recording	
Group	Push (n=250)	Choice (n=246)	Push (n=252)	Choice (n=225)
Yes, only this	20.8	8.5	36.9	8.4
Yes, but not only	27.2	15.9	20.6	16.4
No	52.0	75.6	42.5	75.1