SURVEYS · TECHNOLOGY AND THE FUTURE

Conversational and Chatbot Surveys: A New Mode of Data Collection

Chatbot surveys ask questions through a messaging-style conversation instead of a form. This article explains how conversational surveys work, what research shows about their effects on response quality and disclosure, their risks and how to design and evaluate one.

ModQu research article · 1,985 words · 20 references (APA 7th edition)

Most online surveys look like forms. A page shows one or more questions with boxes and buttons, and respondents click through to the end. A **conversational survey** works differently. Questions arrive one at a time as messages in a chat window, as they would from a person in a messaging app. Respondents type or tap their answers, and the survey replies. When the software behind the conversation can interpret answers and respond to them, it is often called a **chatbot survey**.

Conversational surveys combine features of two traditional modes. Like a web survey, they are self-administered and can be completed at any time. Like an interview, they proceed as a dialogue, with the possibility of follow-up questions and clarification. This article describes how they work, what the evidence shows about the data they produce and how to design them well.

How conversational surveys work

Conversational surveys range from simple to highly adaptive. At the simple end, the software presents a fixed questionnaire in chat form. Each question appears as a message, response options appear as buttons and the order follows the same skip rules as a web survey. The conversation is a new layout for a familiar instrument.

At the adaptive end, the software interprets free-text answers and decides what to say next. It might ask a respondent to say more about a short answer, acknowledge what they said, or ask for clarification when an answer is unclear. Earlier systems used rules and keyword matching to do this. Current systems increasingly use large language models, which can produce flexible, natural replies but are also less predictable (Ziems et al., 2024).

Conversational surveys can be delivered through a website, inside a mobile app or through messaging services and text messages. Text message surveys are an older and simpler relative: Schober et al. (2015) found that respondents who answered by text message on their smartphones gave more precise numerical answers and disclosed more sensitive information than respondents in voice interviews, partly because they could answer at their own pace and without a listener.

Figure 1 compares the main features of form-based web surveys, conversational surveys and interviews.

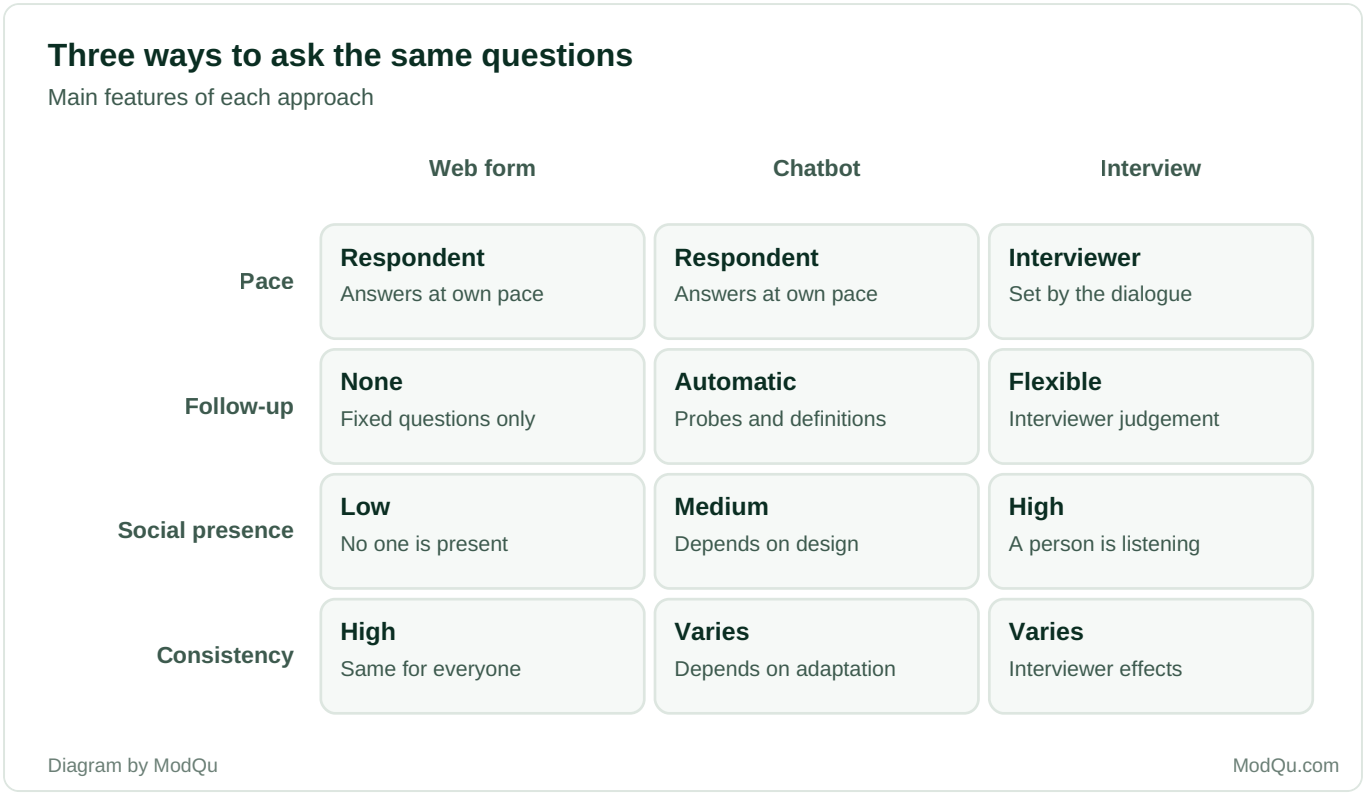


Figure 1. Conversational surveys compared with web forms and interviews. Diagram by ModQu.

Why conversation might help

The idea that a conversation could produce better answers than a form has roots in research on interviewing. Conrad and Schober (2000) showed that when interviewers were allowed to clarify the meaning of questions, respondents answered factual questions more accurately, especially when their circumstances did not fit the question neatly. A conversational survey can offer some of this flexibility, for example by providing a definition when a respondent asks what a term means.

Conversation may also increase engagement. Form-based web surveys can feel impersonal, and long sequences of similar questions encourage respondents to answer with minimal effort. A conversational format, with acknowledgements and varied replies, may keep respondents more attentive. It is particularly suited to open-ended questions, where a chatbot can ask for more detail when an answer is brief. Getting good answers to open-ended questions in web surveys has long been difficult. Smyth et al. (2009) showed that larger answer boxes and short instructions encouraging fuller answers produced longer and more informative responses, which shows that respondents react to cues about how much is expected of them. On smartphones, typing is harder, and Revilla and Ochoa (2016) found that open answers given on phones were somewhat shorter than those given on computers. A conversational follow-up is a further cue of this kind: it tells respondents directly that more detail would be welcome.

What the evidence shows

Response quality

Kim et al. (2019) compared a chatbot survey with a web survey using the same questions. They found that a chatbot using a casual conversational style produced less satisficing, such as fewer identical answers across a series of items, than the web survey. The effect depended on the conversational style, not only on the chat format.

Xiao et al. (2020) compared an AI-powered chatbot that responded to respondents' open-ended answers with a conventional online survey asking the same questions. The chatbot drew answers that were longer, more informative, more relevant and clearer, and respondents reported being more engaged. Celino and Re Calegari (2020) found that respondents generally accepted a conversational interface for answering a questionnaire and that answers were comparable in reliability to those from a traditional form.

These studies are encouraging but limited. Most used convenience samples, often of students or crowd workers, and compared short questionnaires. Effects may differ in general populations, for long surveys and as the novelty of chatbots fades.

Disclosure of sensitive information

A central question is whether people are more or less willing to report sensitive behaviour to a chatbot than on a form or to an interviewer. Research on survey modes shows that respondents report more sensitive behaviour when no interviewer is present (Tourangeau & Smith, 1996; Kreuter et al., 2008). The reason is social desirability: people present themselves favourably to others (Tourangeau & Yan, 2007).

A chatbot sits between these situations. It is a computer, but a conversation can give it a sense of **social presence**, the feeling of interacting with someone. Research on human responses to computers shows that people apply social rules to computers that display human-like cues, even when they know they are dealing with a machine (Nass & Moon, 2000). This suggests that a very human-like chatbot could bring back some of the social desirability pressure that self-administration removes.

The evidence is mixed. Tourangeau et al. (2003) added human-like features, such as pictures of a researcher and personalised messages, to web and telephone voice surveys and found little effect on reports of sensitive behaviour. Lucas et al. (2014) found that people disclosed more to a virtual human interviewer when they believed it was controlled by a computer rather than a person. Conrad et al. (2015) found that animated virtual interviewers with more facial movement increased engagement but could also affect disclosure. The balance between engagement and privacy appears to depend on design details.

Figure 2 illustrates how results from a hypothetical comparison might look.

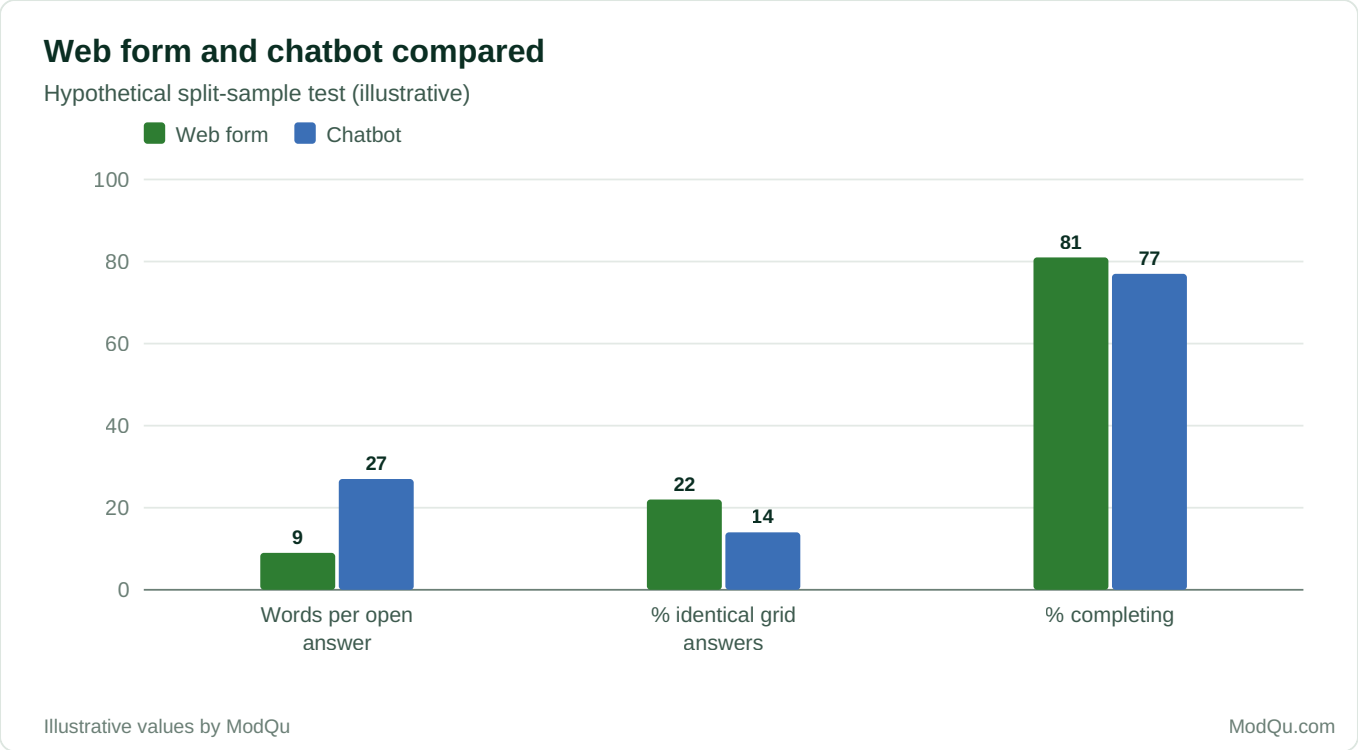


Figure 2. Hypothetical comparison of a web form and a chatbot survey (illustrative values). Chart by ModQu.

Time and completion

Conversational surveys usually take longer than web forms with the same questions, because each question is shown separately and the conversation adds messages. That is a cost for respondents and may increase breakoff in long surveys. On smartphones, where web forms can be awkward, the chat format may be easier to use, but evidence on completion rates in general population samples is limited (Couper, 2017).

Risks and limitations

Inconsistent stimulus

A basic principle of standardised survey measurement is that every respondent receives the same questions (Tourangeau et al., 2000). An adaptive chatbot breaks this principle. If the chatbot rephrases a question, offers different examples or asks different follow-ups, respondents are answering different questions. That makes answers harder to compare and can introduce bias if the chatbot behaves differently for different types of respondent. A practical approach is to keep the wording of closed questions fixed and allow adaptation only in follow-up probes for open-ended questions.

Errors and inappropriate replies

Chatbots based on large language models can produce replies that are wrong, off-topic or inappropriate, and their output cannot be fully predicted in advance (Ji et al., 2023). A chatbot might misinterpret an answer, give an incorrect definition or reply insensitively to a respondent who discloses distress. Bender et al. (2021) noted that language models can reproduce biases found in their training text. Every reply that the system might generate cannot be checked, so designers must limit what the chatbot is allowed to do and test it extensively.

Privacy and transparency

Respondents should know whether they are talking to a person or a machine, and how their answers will be processed and stored. If a third-party AI service is used to interpret answers, data may leave the research organisation's control. Ethical and legal obligations to respondents apply as they would to any survey, and the conversational format should not be used to make respondents feel they are talking to a person when they are not.

Mode effects

Because conversational surveys differ from both forms and interviews, results may not be comparable with earlier data collected in another mode. Organisations that switch an ongoing survey to a chatbot should run the two modes side by side for a period to measure any change.

Designing a conversational survey

Several practices follow from research on survey design and on the early chatbot studies:

- **Decide what should adapt.** Keep closed question wording and response options fixed. Use adaptive replies mainly to probe open-ended answers and to give standard definitions on request.
- **Set the tone deliberately.** A friendly, casual style may increase engagement (Kim et al., 2019), but human-like features may affect answers to sensitive questions. Test the style with the target population.
- **Write probes in advance.** Prepare a set of approved follow-up questions, so that the chatbot chooses among tested probes rather than inventing new ones.
- **Keep the conversation short.** Each extra message adds time. Combine acknowledgements with the next question rather than sending them separately.
- **Make answering easy on phones.** Use buttons for closed questions and allow respondents to scroll back and change answers.
- **Record the full transcript.** Store every message the chatbot sent, so that analysts can see exactly what each respondent was asked.
- **Be open about the machine.** Tell respondents at the start that they are answering an automated survey and explain how answers will be used.

Analysing conversational data

Conversational surveys produce more text than forms, and the text varies in length and structure. Analysts need methods for coding open-ended answers at scale, from manual coding to machine learning and language models, with checks against human coders. Answers to closed questions can be analysed as in any survey, but analysts should check whether follow-up probes changed later answers.

The transcript is also a source of paradata. The number of follow-ups each respondent received, the time between messages and requests for clarification all show where respondents had difficulty. As with web surveys, such process data can reveal problems with particular questions that are not visible in the answers alone (Couper, 2008).

Because adaptive chatbots treat respondents differently, analysts should also check whether follow-up rates differed between groups. If a chatbot probed younger respondents more often than older ones, for example, differences in the richness of their answers may reflect the chatbot's behaviour rather than real differences between the groups.

An example

A public library service wants to understand why people stopped visiting. Its previous web survey produced short open-ended answers such as "busy" or "no time". The service tests a chatbot version on half of a sample of former members. Closed questions are identical in both versions. For the open question, the chatbot uses three approved follow-up probes, such as "Could you tell me a bit more about that?", when an answer has fewer than ten words.

The chatbot version takes two minutes longer on average and has a slightly lower completion rate. Its open-ended answers are three times as long and more often mention specific reasons, such as opening hours and travel. Answers to the closed questions are similar in both versions. The service adopts the chatbot for the open questions only and keeps the form for the rest (illustrative example).

Conclusion

Conversational surveys are a genuinely new mode of data collection, combining the convenience of self-administration with some of the flexibility of interviews. Early evidence suggests that they can improve engagement and the quality of open-ended answers, but the evidence comes mostly from small and convenience samples. The same adaptability that makes chatbots appealing also threatens standardisation, consistency and privacy. Researchers who use them should decide carefully what the chatbot may change, test it with real respondents, keep full transcripts and compare its results with established modes before relying on it.

References

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922

Celino, I., & Re Calegari, G. (2020). Submitting surveys via a conversational interface: An evaluation of user acceptance and approach effectiveness. *International Journal of Human-Computer Studies, 139*, Article 102410. https://doi.org/10.1016/j.ijhcs.2020.102410

Conrad, F. G., & Schober, M. F. (2000). Clarifying question meaning in a household telephone survey. *Public Opinion Quarterly, 64*(1), 1–28. https://doi.org/10.1086/316757

Conrad, F. G., Schober, M. F., Jans, M., Orlowski, R. A., Nielsen, D., & Levenstein, R. (2015). Comprehension and engagement in survey interviews with virtual agents. *Frontiers in Psychology, 6*, Article 1578. https://doi.org/10.3389/fpsyg.2015.01578

Couper, M. P. (2008). *Designing effective web surveys*. Cambridge University Press. https://doi.org/10.1017/CBO9780511499371

Couper, M. P. (2017). New developments in survey data collection. *Annual Review of Sociology, 43*, 121–145. https://doi.org/10.1146/annurev-soc-060116-053613

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys, 55*(12), Article 248. https://doi.org/10.1145/3571730

Kim, S., Lee, J., & Gweon, G. (2019). Comparing data from chatbot and web surveys: Effects of platform and conversational style on survey response quality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Paper 86). Association for Computing Machinery. https://doi.org/10.1145/3290605.3300316

Kreuter, F., Presser, S., & Tourangeau, R. (2008). Social desirability bias in CATI, IVR, and web surveys: The effects of mode and question sensitivity. *Public Opinion Quarterly, 72*(5), 847–865. https://doi.org/10.1093/poq/nfn063

Lucas, G. M., Gratch, J., King, A., & Morency, L.-P. (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior, 37*, 94–100. https://doi.org/10.1016/j.chb.2014.04.043

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues, 56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Revilla, M., & Ochoa, C. (2016). Open narrative questions in PC and smartphones: Is the device playing a role? *Quality & Quantity, 50*(6), 2495–2513. https://doi.org/10.1007/s11135-015-0273-2

Schober, M. F., Conrad, F. G., Antoun, C., Ehlen, P., Fail, S., Hupp, A. L., Johnston, M., Vickers, L., Yan, H. Y., & Zhang, C. (2015). Precision and disclosure in text and voice interviews on smartphones. *PLOS ONE, 10*(6), Article e0128337. https://doi.org/10.1371/journal.pone.0128337

Smyth, J. D., Dillman, D. A., Christian, L. M., & McBride, M. (2009). Open-ended questions in web surveys: Can
increasing the size of answer boxes and providing extra verbal instructions improve response quality? *Public Opinion Quarterly, 73*(2), 325–337. https://doi.org/10.1093/poq/nfp029

Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly, 60*(2), 275–304. https://doi.org/10.1086/297751

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133*(5), 859–883. https://doi.org/10.1037/0033-2909.133.5.859

Tourangeau, R., Couper, M. P., & Steiger, D. M. (2003). Humanizing self-administered surveys: Experiments on social presence in web and IVR surveys. *Computers in Human Behavior, 19*(1), 1–24. https://doi.org/10.1016/S0747-5632(02)00032-8

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge University Press. https://doi.org/10.1017/CBO9780511819322

Xiao, Z., Zhou, M. X., Liao, Q. V., Mark, G., Chi, C., Chen, W., & Yang, H. (2020). Tell me about yourself: Using an AI-powered chatbot to conduct conversational surveys with open-ended questions. *ACM Transactions on Computer-Human Interaction, 27*(3), Article 15. https://doi.org/10.1145/3381804

Ziems, C., Held, W., Shaikh, O., Chen, J., Zhang, Z., & Yang, D. (2024). Can large language models transform computational social science? *Computational Linguistics, 50*(1), 237–291. https://doi.org/10.1162/coli_a_00502