mda

methods, data, analyses

Journal for Quantitative Methods and Survey Methodology

Volume 20, 2026 | 1

Video Interviewing for Survey Data Collection

Frederick G. Conrad, Andrew L. Hupp, Tim Hanson, and Gabriele B. Durrant (Editors)

Volume 20, 2026 | 1

methods, data, analyses is published by GESIS – Leibniz Institute for the Social Sciences.

Editors: Melanie Revilla (editor-in-chief), Ludwig Bothmann, Johannes Breuer, Jessica Daikeler, Edith D. de Leeuw, Gabriele Durrant, Barbara Felderer, Sabine Häder, Jan Karem Höhne, Anne Konrad, Peter Lugtig, Jochen Mayerl, Gerry Nicolaas, Melanie Partsch, Emanuela Sala, Matthias Schonlau, Norbert Schwarz, Carsten Schwemmer, Daniel Seddig, Mareike Wieland

Advisory board: Andreas Diekmann, Bärbel Knäuper, Dagmar Krebs, Frauke Kreuter, Christof Wolf

Managing editors: Barbara Felderer & Maximilian Linde
GESIS – Leibniz Institute for the Social Sciences
PO Box 12 21 55
68072 Mannheim
Germany

E-mail: mda@gesis.org
https://majournals.bib.uni-mannheim.de/mda

GESIS Leibniz Institute for the Social Sciences

Copyediting: Désirée Nießen, Barbara Felderer, Maximilian Linde, Laureen Sieber
Layout: Liebchen + Liebchen Kommunikation GmbH
ISSN 2190-4936 (Online)
© of the compilation GESIS, Mannheim, July 2026

All content is distributed under the terms of the Creative Commons Attribution 4.0 License. Any further distribution of this work must maintain attribution to the author(s), the title of the work, the journal citation, and the DOI.

Content

RESEARCH REPORTS

Editorial Introduction: Video Survey Interviewing Special Issue

Frederick G. Conrad[1], Andrew L. Hupp[1], Tim Hanson[2], Gabriele B. Durrant[3]

[1] *University of Michigan*
[2] *City St. George's, University of London*
[3] *University of Southampton*

© The Author(s) 2026. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Video communication was anticipated long before it was available. The idea was first proposed in the 1870s by Bell Labs with the first call placed on April 7, 1927 by then-U.S. Commerce Secretary Herbert Hoover to AT&T President Walter Gifford (Figure 1, left panel). Hoover spoke from Washington, D.C., and was seen by AT&T officials in New York City. The 1930s saw early prototypes of two-way calls in Germany with the Gegensehn-Fernsprechanlagen (visual telephone system).

Video communication began appearing in popular culture in the 1960s. In the early sixties, the animated series *The Jetsons* debuted in the U.S. in which the characters regularly communicated via videophone, and in the 1968 movie, *2001: A Space Odyssey*, Dr. Heywood Floyd, placed a video call from a space station to Earth. In 1964 Bell Labs debuted a “picture phone” prototype at the New York World’s Fair (Figure 1, right panel). In all of these cases a user placed the video call, much as one would a telephone call, i.e., without prior arrangement — a “cold call”. As it has turned out, video communication tends to be jointly arranged, i.e., scheduled, by the participants, more like a meeting (a video conference) than a phone call, with a minority of applications such as FaceTime following the phone call model.

Figure 1 Video communication evolution. Left Panel: AT&T President Walter Gifford seeing and speaking with then-U.S. Commerce Secretary Herbert Hoover in 1927. Right Panel: Picture Phone from the 1964 World’s Fair. Source: Courtesy of AT&T Archives and History Center

Direct correspondence to
Frederick G. Conrad, Institute of Social Research, University of Michigan, Ann Arbor, MI, USA
E-mail: fconrad@umich.edu

Both the meeting and phone call models pose challenges for recruiting sample members to participate in video survey interviews. The meeting model usually requires online contact information for sample members such as their email addresses to invite them to schedule an interview; there are generally no frames of email addresses for national populations[1] so researchers have had to improvise to make this approach work. The phone call model is also challenging because sample members would need to be users of the same video platform as the survey organization such as FaceTime, WhatsApp or WeChat, which cannot be assumed. Even if a sample uses the platform on which they are being contacted by the survey organization, it is unlikely they will answer the call from an unrecognized number/user. Nonetheless, video communication is now a standard way to converse for a growing segment of the world.

The Covid pandemic led to an urgent and unplanned increase in video interviewing because in-person interviewing was impossible during this time. Rather than deliberately and experimentally developing video interviewing methods, survey researchers innovated within the constraints imposed by the crisis, finding ways to enable face-to-face interaction between an interviewer and respondent without physical proximity: video interviewing is, in fact, a face-to-face mode, but mediated and remote, i.e., not in person. This innovation is reflected in the scope of papers for this special issue. While some are based on experimental research, several report on experiences and outcomes from "real life" production surveys, often carried out during or in the immediate aftermath of the pandemic. It is not yet clear what role video interviewing might play in future research but the pandemic-era experiences provide a useful evidence base for subsequent implementations.

Now that we are five years post-pandemic, survey researchers have some perspective on video interviewing. Thus, now is a good time to reflect on what has been learned about the topic and how video may be implemented in studies moving forward.

The special issue includes articles which report studies in different contexts, using a variety of research designs. More specifically, the special issue includes articles about studies conducted either:

1. before, during, or after the pandemic
2. in laboratory or in field settings
3. experimentally or observationally
4. in cross-sectional designs or in longitudinal designs

---

1 One indication that this is changing can be found in the increasing use of "digipost" services in which all residents of a country is provided an official email address for receiving digital correspondence from public authorities. For an example in Denmark see https://lifeindenmark.borger.dk/apps-and-digital-services/Digital-Post.

5. in comparative (i.e., cross-national) designs or within a single country
6. as comparisons of video to other interviewing modes (in-person, telephone, web, and pre-recorded video) or as an in-depth examination of only live video interviewing
7. using probability or non-probability sample sources
8. to measure willingness to participate in live or in prerecorded video interviews
9. with general populations or with specialized populations (e.g., children, older adults, those with mental health disorders)

This diversity of study contexts and designs suggests that video interviewing is flexible and can be implemented broadly, despite the need for specialized hardware and software. However, none of the studies were conducted in countries that might be considered primarily low or middle income, so we cannot conclude that video interviewing is universally applicable.

Survey researchers and methodologists have demonstrated interest in video interviewing for collecting quantitative data and its future as evidenced by sessions devoted to the mode at conferences of the American Association for Public Opinion Research (AAPOR) and the European Survey Research Association (ESRA) between 2021 and 2025. Several initiatives in the UK have explored and promoted video interviewing, including the Surveys Futures (https://surveyfutures.net/research-programme/research-strand-3/) collaboration (Research Strand 3: video interviewing) which hosts working papers and hosted a workshop on video interviewing in March 2026. In the US, AAPOR hosted a webinar on video interviewing in 2022 and at least two important surveys have continued to collect data via video interviewing in the post-pandemic era, the American National Election Studies (https://electionstudies.org/) and the Medical Expenditure Panel Survey (https://meps.ahrq.gov/mepsweb/). Internationally, a Special Interest Group on video interviewing has been established, and meets periodically to collaborate and disseminate information, including organizing conference sessions.

Born out of the pandemic-created need for data collection that is both remote and interviewer-administered, the deployment of video interviewing is no longer urgent. Although some video interviewing is currently a component of ongoing survey research, its future may be more likely as one of multiple modes offered to respondents than as a stand-alone mode. More likely, video interviewing will continue as one among many modes offered to respondents or particular subpopulations, as well as being deployed in longitudinal studies in which contact information – especially email addresses which are typically required to invite sample members to participate in video interviews – can be obtained in earlier waves of data collection, or in online panels whose clients might require video interviewing for particular studies. Whatever contexts in which video interviewing is used going forward, this special issue serves as a collection of researcher experience and evaluation of the approach that can inform decisions about its future use.

As for any new mode, it is important to establish basic principles and nomenclature used by those active in the field. In the special issue, we asked authors to use "video interviewing" to describe live interaction between an interviewer and a respondent mediated by video, although in the context of prerecorded video interviews – a web survey in which a videorecording of an interviewer asking questions has been embedded – authors differentiated the modes as "live video interviews" or "prerecorded video interviews." We do not use labels or acronyms that imply that interviews are *assisted* by video communication because video mediates but does not assist the parties in their interaction. Similarly, we asked authors to use the term "in-person" rather than "face-to-face" interviewing because video interviewing is also a face-to-face mode. Finally, we encouraged authors to describe legacy modes explicitly, e.g., phone, in-person, and web, rather than by acronyms indicating their implementation on a computer because computerization can be assumed unless something to the contrary is specified and so that readers unfamiliar with the conventional four-letter acronyms for computerized modes can make sense of what they are reading.

Beyond Feasibility of Video Interviewing

Before video interviewing was introduced in response to the pandemic, the first question asked by researchers and survey practitioners was whether it would work. Would survey respondents be willing and able to start and get through a video interview? Would video platforms work sufficiently well to allow the performance of survey tasks?

The early literature on video interviewing focused on testing its feasibility and assessing different design factors (Anderson, 2008; Centeno et al., 2024; Jeannis et al., 2013, Schober et al., 2020; Wood et al., 2020). The general conclusion from these papers, alongside presentations in video interviewing workshops and conference sessions, was that it is possible to carry out quantitative survey interviews by video, while making clear that there are several design choices that need to be considered. This is supported by papers in this special issue.[2] For example, **Hanson et al. (2026)** found that on the European Social Survey (ESS), most of those who participated by video got through the interview without technical issues, and respondent ratings of their experience of the video interviews

2 Throughout this editorial, references to articles included in this Special Issue are indicated in bold.

were in line with those reported for in-person interviews. **Durrant et al. (2026)** explored the feasibility of obtaining consent and conducting cognitive assessments in video interviews and concluded that these special tasks could be carried out in this mode much as in in-person interviews.

The feasibility of video interviewing as a practical data collection method is therefore now well established. Indeed, most papers in the special issue move beyond feasibility and address specific gaps in what is known about video interviewing and how it works in different contexts. **Hupp et al. (2026)**, and **Guggenheim et al. (2026)** examine participation and nonresponse, **Fomby et al. (2026)** and **Carr et al. (2026)** report on video interviews with special populations (children and those experiencing mental health symptoms, respectively), and **Feuer and Schober (2026)** investigate one design feature of video interviews, the presence or absence of a self-view window and how it affects interaction. **Hanson et al. (2026)**, **Zavala-Rojas et al. (2026)**, **Durrant et al. (2026)**, and **Cornesse et al. (2026)** compare data quality between video and other modes. **Sun and Conrad (2026)** explore interviewer-respondent dynamics in video versus in-person interviews, and **Claassen et al. (2026)** investigate hypothetical willingness to participate in prerecorded video interviews.

In the remaining sections, we discuss two key aspects of video interviewing: (1) participation, in particular recruiting respondents and completing the interview, and (2) data quality, in particular comparability of results in video and other modes, primarily in-person interviews but also web surveys and telephone interviews.

Participation

There are (at least) three ways that a sample member and survey organization can initiate a video interview. First, the survey organization can place an unsolicited (i.e., cold) video call to the sample member analogous to a telephone survey invitation/introduction. As indicated earlier, unsolicited video calls are not practical given the dominant meeting model in which video communication is largely scheduled ahead of time and allows users of various devices to join a call, unlike in cold call models.

Second, the sample member can contact the survey organization and ask to be interviewed. Such on-demand interviewing was possible in the 2020 American National Elections Studies (ANES) reported in **Guggenheim et al. (2026)**. Sample members were invited by mail to complete a web screener. Those who were eligible were redirected to a video waiting room where they could ask to be interviewed at that time; if no interviewers were available, they could wait until one was available, or they could schedule a future video interview (DeBell et al., 2022). Most participated in a video interview upon arrival. While **Guggenheim et**

al. (2026) reported that the on-demand model was successful, we note that it is inherently inefficient because interviewers can have periods of down time (for which they are paid) or more demand (sample members in the waiting room) than they can meet.

Scheduling, the third model, has become the default and was by far the most prevalent approach reported in the special issue articles. Two approaches to scheduling have been used: the sample member can self-schedule an interview[3] or can schedule the interview with a representative of the survey organization, such as an interviewer. The two approaches can, in principle, be used in tandem, for example a sample member can be both invited via email to self-schedule online and also provided with a link; if they do not follow-through, the organization phones the sample member and works with them to find a good time for the video interview. Once the interview is scheduled (and in some cases rescheduled), the sample member must show up at the appointed time to begin the interview. Both scheduling and showing up are obstacles to completing a video interview that are rarely factors in completing interviews in other modes.

This framework is laid out conceptually by **Hupp et al. (2026)** in the special issue and instantiated by their experimental study. They found that scheduling and showing up for a scheduled video interview are both impediments to completion. However, almost all of the sample members who made it to the appointment completed their interview. **Guggenheim et al. (2026)** reported a higher scheduling rate but this appears to be due to differences in the scheduling procedure between the two studies: in **Hupp et al. (2026)**, sample members received an unsolicited invitation to self-schedule an appointment for the interview, but in **Guggenheim et al. (2026)** an interviewer scheduled a follow-up video interview at the conclusion of an in-person interview.

Participation in video interviews seems to depend on what modes are made available to respondents. As **Durrant et al. (2026)**, report, in interviews where video was the only or primary data collection mode, response rates were notably lower than when in-person interviews were also an option. In the latter scenario, they report that video interviews attracted a higher proportion of younger and more educated sample members than did in-person interviews, with the opposite pattern (higher proportion of older and less educated respondents) observed for in-person interviews. It is possible that a mix of modes might jointly cover the target population more effectively than either mode alone as some respondents, e.g., those residing in high crime areas or who are younger and more familiar with video communication, might opt for video interviews but would decline to participate if in-person interviews were the only choice. Whether and how the inclusion of video interviews in the set of modes offered to sample members

---

3 See Conrad et al. (2023) for an example of self-scheduling video interviews; see McGonagle and Sastry (2021) for an example of self-scheduling a telephone interview.

results in a more balanced overall study sample than when video interviews are not an option is an important topic for further investigation.

Data quality

Prior to this special issue, the small published literature about video interviewing pointed toward considerable overlap in the quality of survey data collected in video and in-person interviews as well as in the respondents' experience. In particular, Endres et al. (2023) reported the same levels of satisficing and disclosure as well as respondents' satisfaction with the interview in both video and in-person interviews. On the one hand this is not surprising: the features comprising both modes are very similar, e.g., both are face-to-face in the sense that the interviewer and respondent see and hear each other in real time (see Schober et al., 2023 for a feature comparison of video, in-person, and telephone interviewing). On the other hand, the interviewer and respondent are not physically co-located in video interviews, i.e., the interaction is mediated by a screen, which could affect the experience and lead to different data quality outcomes.

Although Endres et al. (2023) found no differences between the two modes in the laboratory, it is possible that in production settings differences might emerge. For example, there may be distractions in respondents' immediate environment that a remote interviewer cannot mitigate. In addition, the respondent may appear to be looking at the camera, and thus engaged in the interview, when they are actually looking at a different on-screen window; this could be good for data quality if they have shifted their gaze to consult records needed to accurately answer a question but it could harm data quality if it reflects their unrelated behavior such as online shopping. The articles in the special issue that measure data quality were primarily conducted in production/field settings and while the story that emerges across these studies is more complicated than what has emerged from the laboratory, there is not compelling evidence of a systematic difference in quality or respondent experience between video and in-person interviewing.

Missing data, straightlining and disclosure. Hanson et al. (2026) compare missing data and straightlining in video and in-person interviews in the six ESS countries that conducted the largest number of video interviews. The authors report small but significant mode differences in the average number of missing answers per respondent in five of the six countries with smaller amounts in video than in-person interviews for four countries and the reverse for the fifth. They also report little difference in the amount of straightlining between the two modes.

Another indication that video interviewing is experienced much like in-person interviewing is that in both of these modes respondents exhibit more socially desirable responding (less disclosure) than they do when they self-administer the questionnaire, as in a web survey. **Cornesse et al. (2026)** report findings consistent with greater social desirability in video interviews than in a web survey for seven out of nine behaviors. Consistent with Endres et al. (2023) and Conrad et al. (2023), these findings suggest that the *presence of the interviewer* is experienced much the same way in video and in-person data collection to positive and negative effect: an interviewer presumably increases respondent engagement and conscientiousness compared to web data collection but inhibits candid responses when they are sensitive or otherwise stigmatized.

Interviewer effects. Hanson et al. (2026) report that interviewer effects (ICCs) were similar in size in both in-person and video interviews for most items, with somewhat larger differences for one country, suggesting that whether the impact of interviewers on variance is higher or lower for particular items, it is about the same in both modes. See West et al. (2022) for a comparison of interviewer effects in live and prerecorded video interviews.

Measurement invariance. Zavala-Rojas et al. (2026) report measurement invariance (i.e., equivalence) across the video and in-person interviews for the same six ESS-participating countries as in the **Hanson et al. (2026)** special issue article. For one concept, social trust, they report measurement invariance in all six countries; for another concept, attitudes toward immigration, they report equivalence in three of the six countries. Thus, in the majority of the comparisons, video and in-person interviews produced equivalent results.

Respondent experience. Hanson et al. (2026) report that respondents experience video and in-person interviewing in much the same way across the six ESS countries. They observed no differences between the modes on a post-interview questionnaire about their experience when results from all six countries were combined, although respondents rated in-person interviewing more positively than video interviewing in two countries, the reverse for one country, and no difference in the remaining three countries.

Video self-view. Video interviewing not only allows the interviewer and respondent to see each other but it also allows the respondent to see themselves via the self-view window typically displayed by default. This raises the question of how seeing oneself during the interview affects data quality, especially disclosure. **Feuer and Schober (2026)** tested the impact of the self-view window in a laboratory study in which they compared disclosure of sensitive information (socially undesirable responses) with and without the self-view window. They report no less disclosure when the self-view is present than when it is not and, if anything, more disclosure for some stigmatized behaviors.

Spoken communication in video interviews. Sun and Conrad (2026) report that the verbal communication – interaction between the interviewer and respondent – in in-person and video interviews – was largely indistinguishable. For example, the prevalence of respondents' disfluencies (e.g., *um*s and *uh*s) is the same in both modes. They do report more laughter in video than in-person interviewing, but propose this could be nervous laughter possibly due to technical problems in the video connection and/or display. In earlier work, Sun et al. (2021) reported that respondents establish rapport with interviewers to the same extent in both video and in-person interviews, presumably because the cues on which rapport is based are equally perceptible in the two modes. Taken together these findings suggest that verbal interaction in in-person and video-mediated interviews is largely indistinguishable.

Impact of being able to see the respondent. How might the combination of video and spoken interaction affect outcomes of the interview compared to speech only, as in a telephone interview? **Carr et al. (2026)** reported greater prevalence of psychiatric disorders measured in video than telephone interviews and suggested this could be due, in part, to the clinical interviewer's ability to see respondents' faces and their physical movements in video but not phone interviews. One implication is that for some respondents, symptoms may have gone undetected and undiagnosed on the phone but detected and diagnosed in a video interview, suggesting higher quality data in clinical interviews conducted via video than telephone. The authors speculate that video interviewing may also have led to more self-reported symptoms because respondents felt greater rapport with interviewers they could see.

What have we learned so far?

Looking across the eleven articles in the special issue makes it possible to assess the different ways video interviewing has been used, to understand the strengths and weaknesses of different uses, and to consider where the method may be more or less effectively employed in the future. This allows us to draw some tentative conclusions and identify areas where more research is needed.

It seems clear that the optimal conditions for deploying video interviewing are not as a single mode of data collection. This reflects issues with population coverage (video interviewing appeals more to some sample members than others) and challenges of contact (inviting sample members typically requires a frame of email addresses). Moreover, making and showing up for an appointment involves effort not required for data collection in other modes. But with the current prevalence of mixed-mode designs, video interviewing may be a valuable complement to other modes.

Different uses of video as part of mixed-mode designs can be envisaged. It can be combined with in-person interviewing to bring in people who may have limited availability or be unwilling or unable to take part in an in-person interview. The evidence for near-equivalent data quality between video and in-person interviewing discussed in the previous section may make this combination of modes particularly strong. But there may also be potential to combine video with web to bring in people who may struggle with the text-based self-administered format of a web questionnaire and would be better able or more willing to complete the questionnaire with a remote – but human – interviewer. Or video could be used in combination with or in place of telephone interviewing in a mixed-mode design, which may help with building rapport and disclosure of certain sensitive information (as suggested in **Carr et al. (2026)**).

Even within mixed-mode designs, there may be different types of studies where video interviewing is a more feasible option than others. It is clear from evidence collected to date, including the papers by **Hupp et al. (2026)** and **Guggenheim et al. (2026)** in this special issue, that a significant challenge with video interviewing is the contact and scheduling process. Most people who start video interviews finish them – but getting them started in the first place can be far from straightforward. This relates to how contact is first made with target respondents. For many studies, email addresses or telephone numbers will not be available and so initial contact will rely on postal mailings or in-person visits. The evidence so far suggests that facilitating video interviews based on postal or in-person contacts is challenging. Video may therefore offer more potential for longitudinal studies, where more extensive contact information can be gathered in an earlier wave, or where rich sample frames include contact details beyond postal addresses. The latter is evidenced in the paper by **Hanson et al. (2026)** based on the ESS, where those countries that carried out a higher share of video interviews tended to be those in which interviewers could make first contact by telephone rather than needing in-person visits.

This special issue provides preliminary evidence on the conditions that may support effective use of video interviewing, which we hope will be helpful for survey practitioners. Yet questions remain. It is not fully clear, for example, to what extent video interviewing can bring in sample members who would not respond in other modes as opposed to enabling people who would likely have responded irrespective of mode. Relatedly, more research is needed on the extent to which including video as part of a mixed-mode strategy adds benefits versus extra costs and complexity. Similarly, more research is needed on how to lower the barriers to participating introduced by the need for self-scheduling. At minimum, however, the range of studies that have incorporated video interviewing in recent years, many of which are showcased in this special issue, ensure we now know far more about this mode. Video interviewing was hardly used as a survey data collection mode prior to the pandemic. It is now established and available for practitioners to use as part of a growing toolbox; we can expect that the conditions in which it can be most effectively used will be refined further over the coming years.

What's on the horizon?

Looking ahead, there are at least two directions in which we believe it is likely that video interviewing will evolve. The first direction consists of refinements of the current paradigm; the second direction extends the mode to technology-mediated self-administration.

The current paradigm was largely developed using existing video platforms that were not specialized for survey interviewing. This was an adequate way to proceed in the midst of a public health crisis but not optimal. Video interviewing platforms that integrate features available in current interviewing software with video capability would help interviewers administer video interviews more efficiently and facilitate adoption. In addition, specialized video interviewing platforms could allow respondents to switch to self-administration to collect sensitive responses, helping to overcome the social presence of a visible interviewer when that is undesirable.

One example of the second direction in which we see video interviewing evolving, i.e., its extension to self-administration is prerecorded video interviewing, i.e., web surveys in which embedded video recordings display interviewers reading questions. Preliminary evidence suggests respondents feel connected to video recorded interviewers, although the sense of connection is greater in live video interviews (Conrad et al., 2023) and this connection seems likely to increase engagement without increasing social desirability (Haan et al. 2017). It may also help support the participation of people who may be less able to complete text-based self-administered surveys (for example, those with low literacy).

The paper by **Claassen et al. (2026)** in the special issue assesses hypothetical willingness to participate in this type of data collection. They found that the majority of respondents were hypothetically willing to participate in "web surveys in which interviewers read questions to you via pre-recorded videos" especially with responses entered by clicking/tapping or typing. But fewer respondents said they were willing to participate if it required answering by voice or via video. The mode seems to be a hybrid of in-person and video interviewing on the one hand and self-administration as in conventional text-based web on the other, so may be the best of both. The technology required to deploy this mode is widely available.

A second type of video interviewing that does not involve a live, human interviewer relies on embodied (avatar) virtual interviewers, i.e., computer-animated agents that can speak questions and interpret respondents' spoken answers. Earlier tests of this approach either simulated interactivity (Conrad et al., 2015) or deployed non-interactive interviewing agents (e.g., Lind et al., 2013), but the arrival of large language models (LLMs) makes human-like interactivity readily available for driving interaction between virtual interviewers and respondents. We expect the combination of avatar and LLM technologies to result in interviewing agents whose social identities, personalities, and interviewing approach are under researchers' control and can be chosen by respondents (Conrad et al., 2020). The extent to which sample members will accept embodied virtual interviewers is currently unknown, but likely to increase as AI-driven agents – not just for interviewing – become more ubiquitous. As we have proposed for live video interviews conducted by human interviewers, virtual interviewers may be most effective as one of multiple ways to collect data, presumably more attractive to some groups of sample members than others.

Conclusion

While video interviewing may not make sense as a standalone mode in general, there are scenarios in which it is well suited as the one data collection mode: video interviewing might be a good fit if the data collection requires a live but remote interviewer who can see the respondent and can be seen by the respondent, may require the display of visual information, and is free of the usual challenges of scheduling and recruiting sample members. For example, studies of institutionalized sample members such as prisoners or deployed military personnel, and whose circumstances allow others (wardens, commanding officers) to schedule their interviews may **only** be feasible if those interviews are conducted by video.

But there are other research scenarios in which video interviewing may also make sense in conjunction with other modes: longitudinal studies in which contact information such as an email address for sample members can be obtained in an in-person interview and used to invite sample members to a future video interview; one-off interviews in which respondents are required to choose the mode in which they provide data, including video interviews; a dedicated video panel, perhaps a subset of a larger panel, whose members have agreed in advance to participate in video interviews. And if official email addresses become standard in some countries as in the Danish example mentioned earlier (https://lifeindenmark.borger.dk/apps-and-digital-services/Digital-Post), this will facilitate the deployment of video interviewing.

When the pandemic pushed researchers to explore video interviewing, the components were already in place, e.g., Skype and other video platforms were already being adopted and smartphones equipped with video capability were widespread. The technology has come a long way from what is depicted in Figure 1. So it was not a great stretch for researchers to consider video as a way to continue data collection despite the public health crisis. Some of the obstacles encountered then and reported in the special issue papers are still challenges but we have learned that video interviews can provide high quality data and the circumstances under which they might be deployed are coming into clearer focus.

References

Anderson, A. H. (2008). Video-mediated interactions and surveys. In F. G. Conrad & M. F. Schober (Eds.), *Envisioning the survey interview of the future* (pp. 95-118). Wiley. https://doi.org/10.1002/9780470183373.ch5

Carr, C., Guyer, H., McDaniel, K., Reed, W., Macuada, C., Geiger, P., Carpenter, L., Peytchev, A., Dever, J., Suresh, R., & Edlund, M. (2026). Video interviewing and observed differences in mental health outcomes. *Methods, Data, Analyses, 20*(1), 108–123. https://doi.org/10.12758/mda.2025.06

Centeno, L., Arrue, J., Good, C., Edwards, B., & Steiger, D. (2024). Practical considerations for developing and implementing a video mode for survey data collection. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique, 164*(1), 122-146. https://doi.org/10.1177/07591063241277065

Claassen, J., Lenzner, T., Höhne, J. K., & Ziller, C. (2026). A survey mode of the future? Investigating respondents' willingness to participate in self-administered video-based web surveys. *Methods, Data, Analyses, 20*(1), 317–346. https://doi.org/10.12758/mda.2025.10

Conrad, F. G., Schober, M. F., Hupp, A. L., West, B. T., Larsen, K. M., Ong, A. R., & Wang, T. (2023). Video in survey interviews: Effects on data quality and respondent experience. *Methods, Data, Analyses*, *17*(2) 135-170. https://doi.org/10.12758/mda.2022.13

Conrad, F. G., Schober, M. F., Nielsen, D., & Reichert, H. (2020). Social identities of virtual interviewers and their impact on survey responses. In K. Olson, J.D. Smyth, J. Dykema, A.L. Holbrook, F. Kreuter, & B.T. West (Eds.), Interviews effects from a Total Survey Error perspective (pp. 149-164). Boca Raton, FL; CRC Press https://doi.org/10.1201/9781003020219-15

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: Cognitive Science*, *6*, 1578. https://doi.org/10.3389/fpsyg.2015.01578

Cornesse, C., Phillips, B., Neiger, D., Farrell, E., & Carmo, P. (2026). Investigating live video interviewing in the Life in Australia™ panel survey: An experimental study. *Methods, Data, Analyses, 20*(1), 269–301. https://doi.org/10.25521/mda.674

DeBell, M., Amsbary, M., Brader, T., Brock, S., Good, C., Kamens, J., Maisel, N., and Pinto, S. 2022. Methodology Report for the ANES 2020 Time Series Study. Palo Alto, CA, and Ann Arbor, MI: Stanford University and the University of Michigan. DeBell, et al 2020 ANES Methodology Report https://electionstudies.org/wp-content/uploads/2022/08/anes_timeseries_2020_methodology_report.pdf

Durrant, G., Kocar, S., Brown, M., Hanson, T., Sanchez, C., Wood, M., Taylor, K., Tsantani, M., & Huskinson, T. (2026). Live Video Interviewing: Empirical Evidence of Opportunities and Challenges Across Seven Major Social Surveys. *Methods, Data, Analyses*, *20*(1), 237–268. https://doi.org/ 10.25521/mda.661

Endres, K., Hillygus, D. S., DeBell, M., & Iyengar, S. (2023). A randomized experiment evaluating survey mode effects for video interviewing. *Political Science Research and Methods*, *11*(1), 144-159. https://doi.org/10.1017/psrm.2022.30

Feuer, S., & Schober, M. F. (2026). Effects of the self-view window in live video survey interviews. *Methods, Data, Analyses, 2*0(1), 124–178. https://doi.org/10.12758/mda.2025.07

Fomby, P., Sastry, N., Johnson, H., Crane, S., & Connors-Burge, V. (2026). Video interviewing for standardized achievement assessments with children. *Methods, Data, Analyses, 20*(1), 87–107. https://doi.org/10.25521/mda.657

Guggenheim, L., Valentino, N. A., Hupp, A. L., DeBell, M., Hillygus, D. S., & Shaw, D. R. (2026). Understanding nonresponse in live video survey interviews: Evidence from two tests conducted as part of the American National Election Studies (ANES). *Methods, Data, Analyses, 20*(1), 46–86. https://doi.org/10.25521/mda.641

Haan, M., Ongena, Y. P., Vannieuwenhuyze, J. T. A., & De Glopper, K. (2017). Response behavior in a video-web survey: A mode comparison study. *Journal of Survey Statistics and Methodology, 5*(1), 48-69. https://doi.org/10.1093/jssam/smw023

Hanson, T., Briceno-Rosas, R., Ainsaar, M., & Kartau, H. K. (2026). Live online video interviewing as a complementary mode to in-person interviews: Evidence from the European Social Survey. *Methods, Data, Analyses, 20*(1), 179–208. https://doi.org/10.25521/mda.663

Hupp, A. L., Conrad, F. G., Larsen, K. M., Schober, M. F., West, B. T., & Harrison, M. (2026). Dynamics of Participating in Video Survey Interviews. *Methods, Data, Analyses, 20*(1), 21–45. https://doi.org/10.25521/mda.635

Jeannis, M., Terry, T. Heman-Ackah, R., & Price, M. (2013). Video interviewing: An exploration of the feasibility as a mode of survey application. *Survey Practice, 6*(1). https://doi.org/10.29115/SP-2013-0001

Lind, L. H., Schober, M. F., Conrad, F. G., & Reichert, H. (2013). Why do survey respondents disclose more when computers ask the questions? *Public Opinion Quarterly, 77*(4), 888-935. https://doi.org/10.1093/poq/nft038

McGonagle, K., & Sastry, N., (2021). An experimental evaluation of an online interview scheduler: Effects on fieldwork outcomes. *Journal of Survey Statistics and Methodology, 9*(3), 412-428. https://doi.org/10.1093/jssam/smaa031

Schober, M. F., Conrad, F. G., Hupp, A. L., Larsen, K. M., Ong, A. R., & West, B. T. (2020). Design considerations for live video survey interviews. *Survey Practice, 13*(1). https://doi.org/10.29115/SP-2020-0014

Schober, M. F., Okon, S., Conrad, F. G., Hupp, A. L., Ong, A. R., & Larsen, K. M. (2023). Predictors of willingness to participate in survey interviews conducted by live video. *Technology, Mind, and Behavior, 4*(2), 215-229. https://doi.org/10.1037//tmb0000100

Sun, H., & Conrad, F. G. (2026). How does video interviewing affect the interviewers' and respondents' paralinguistic behaviors? A first exploration. *Methods, Data, Analyses, 20*(1), 302–316. https://doi.org/10.25521/mda.581

Sun, H., Conrad, F. G., & Kreuter, F. (2021). The relationship between interviewer-respondent rapport and data quality. *Journal of Survey Statistics and Methodology, 9*(3), 429-448. https://doi.org/10.1093/jssam/smz043

West, B. T., Ong, A. R., Conrad, F. G., Schober, M. F., Larsen, K. M., & Hupp, A. L. (2022). Interviewer effects in live video and prerecorded video interviewing. *Journal of Survey Statistics and Methodology, 10*(2), 317-336. https://doi.org/10.1093/jssam/smab040

Wood, M., Cole, K., Dangerfield, P., Stockman, A., & Spencer, S. (2020). BCS70 Video call Pilot: Report of Pilot stage. NatCen Social Research. https://cls.ucl.ac.uk/wp-content/uploads/2017/02/BCS70-Video-Call-Pilot.pdf

Zavala-Rojas, D., Keles, O. K., & Schwarz, H. (2026). Measurement invariance of survey data across face-to-face and video interview modes. *Methods, Data, Analyses, 20*(1), 209–236. https://doi.org/10.25521/mda.662

methods, data, analyses | Vol. 20(1), 2026, pp. 21-45 DOI: 10.25521/mda.635

Dynamics of Participating in Video Survey Interviews

Andrew L. Hupp[1], Frederick G. Conrad[1], Kallan M. Larsen[2], Michael F. Schober[3], Brady T. West[1] & Makenna Harrison[1]

[1] *University of Michigan*
[2] *University of North Carolina at Chapel Hill*
[3] *New School for Social Research*

Abstract

Survey researchers have begun to conduct survey interviews by video. This study investigates participation, i.e., starting and completing, two types of video survey interviews, those conducted live by a human interviewer and those in which a video-recorded interviewer asks questions in a web survey; the results are compared to outcomes from a conventional text-based web survey. We invited sample members to participate in either (a) a live video interview, (b) a prerecorded video interview, or (c) a conventional text-based web survey, recruiting them from two different opt-in (non-probability) sample sources. To participate in a live video interview, sample members were required to schedule the interview and show up at the appointed time. The need to schedule and show up created an obstacle to starting the interview, especially for those used to text-based web surveys, which can be started by clicking a link when convenient. The results indicate that, of those invited to a live video interview, about 10% scheduled an appointment, although this differed substantially by sample source. Of those who scheduled an interview, about half showed up to start it. Once these sample members started the live video interview, almost all finished. In contrast, sample members who started prerecorded video interviews broke off at much higher rates than did those in the two other modes. This was especially true for respondents familiar with conventional text-based web surveys. This article discusses implications and recommendations for how researchers might increase participation for both types of video interviews and adapt them to the population being studied.

Keywords: live video survey interviews, prerecorded video survey interviews, scheduling video interviews, recruitment, breakoffs

© The Author(s) 2026. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

As more people routinely communicate by two-way video, survey researchers have begun to explore live video interviewing as a survey mode (Anderson, 2008) in which respondents and interviewers see and hear each other in real time. This mode may be a good fit for research that requires face-to-face interviewing when physical co-presence is challenging due to safety concerns (e.g., dangerous public health conditions or the sample member lives in a high-crime neighborhood), or the geographical dispersion of the target population. One advantage is that live video interviewing does not require any travel by the interviewer which, depending on the study design, may substantially reduce survey costs. And both data quality and respondents' sense of rapport with video interviewers appears, so far, to be at least as good as in in-person interviews (e.g., Endres et al., 2022; Sun et al., 2021).

One-way video is arguably even more common in everyday life than two-way video—for example, on YouTube and other social media platforms; on learning sites that offer on-demand instructional videos; in recordings from video doorbells and police body cameras; and in live streams from baby monitors and surveillance cameras. In this approach, video may be recorded in advance (as on YouTube) or streamed live (as with baby monitors). The technology may also be interactive in the limited sense that viewers can respond by clicking or typing (e.g., liking a video, posting a comment, or subscribing to a channel), whereas surveillance video generally does not support this kind of interaction.

One-way video has been studied as a survey mode in which prerecorded interviewers (both human and computer-generated) ask questions and respondents enter their answers directly in a web survey (e.g., Conrad et al., 2023; Conrad et al., 2020; Fuchs, 2009; Fuchs & Funke, 2007; Gerich, 2008; Haan et al., 2017; Höhne et al., 2024; Krysan & Couper, 2003; Lind et al., 2013; Strojil & Cígler, 2025). Evaluations of one-way video survey data collection suggest that respondents can feel some social connection to the recorded interviewer (e.g., Conrad et al., 2023), while showing little to no social desirability bias or time pressure. In this article, we refer to two-way video interviewing as "live video" and to one-way video survey interviewing as "prerecorded video."

The aim of this article is to identify and contrast the obstacles to (a) recruiting sample members and (b) to their completing (i.e., not breaking off during) an interview in either of the two video modes (live and prerecorded). Starting and completing a conventional text-based web survey serves as a standard of comparison for the two video modes. The data were derived from a study (Conrad et al., 2023) exploring response quality in the two video modes that also compared

Direct correspondence to
Andrew L. Hupp, Institute for Social Research, University of Michigan, Ann Arbor, MI
E-mail: ahupp@umich.edu

responses to those collected in a text-based web survey. Thus, the current article allows the comparison of participation across these three modes.

Recruitment for Live and Prerecorded Video Survey Interviews

Recruiting participants into live video interviews is fundamentally different from recruiting them into prerecorded video "interviews." To participate in a live video interview, a sample member generally needs to be invited in another mode, usually by email or postal mail, and then must schedule an interview time, making the process more like planning a meeting than receiving an invitation phone call.[1] In contrast, sample members who have been invited to a prerecorded video interview via email, text message, or postal mail follow a link in the invitation to start the interview, as in a web survey. Thus, recruiting for live video interviews has at least one more stage, i.e., scheduling, than is the case for prerecorded video interviews.

Scheduling a Live Video Interview

If the study collects data from the same respondents on multiple occasions (e.g., Survey of Health, Aging and Retirement in Europe or the US Current Population Survey), it may be possible for an interviewer to schedule a video interview as part of an earlier interview session or when the survey organization visits the sample member for reasons related to the study (e.g., to collect biospecimens). By virtue of being in person, the interviewer can also make sure the video connection and technology are functioning and that the sample member is sufficiently familiar with the technology to participate in a later video interview[2].

It is also possible for a sample member to self-schedule the interview (see Conrad et al., 2023 for an example of self-scheduling a live video interview and McGonagle and Sastry, 2021 for a discussion of self-scheduling a telephone interview). This can be offered to sample members in the previous scenario who are not able or willing to schedule while the interviewer is present or in a one-time survey via another mode, e.g., email or postal mail. In either case, the sample member is likely provided a link to an online scheduling portal, where available interview times are listed and the sample member is instructed to select one, much like one might self-schedule a medical appointment, car repair, or haircut.

---

1 An exception to the meeting analogy is being invited via a video platform such as FaceTime, which would be like a cold call or phone invitation. However, FaceTime or similar platforms have not been successfully used for live video survey interviews to our knowledge.

2 In principle an interviewer can schedule a video interview when connected to a sample member for a phone interview, but because they are not on site, they will not be able to confirm that the video technology and sample member's ability to use it are adequate for subsequent data collections.

There is also some precedent for an approach that does not involve scheduling: the interview can be conducted "on-demand," i.e., the sample members are provided with a link to a video waiting room, which they enter, and when an on-demand interviewer is available, the data collection can begin (DeBell et al., 2022; Guggenheim et al., 2026). On-demand interviewing requires having interviewers on call, although probably not around the clock (unlike prerecorded video interviews, which can be started anytime).

Keeping the Appointment for a Live Video Interview

Assuming the sample member has scheduled a video interview, there remains one obstacle to beginning data collection: the sample member must show up at the appointed time. There is little published research on the extent to which survey interview appointments are kept. In the absence of literature about showing up for scheduled survey interviews, a reasonable proxy is the extensive body of research carried out about showing up for medical appointments where patients missing appointments has a substantial impact on cost. A key take-away from this literature is that as more effort is required to keep an appointment, e.g., greater distance to the clinic or lack of transportation, the rate of broken appointments increases (e.g., Dantas et al., 2018; Deyo & Inui, 1980). In fact, appointments are kept at a substantially higher rate for telehealth (i.e., video) visits than in-person clinic visits (Alkilany et al., 2021; Drerup et al., 2021) presumably because no travel is required. Not surprisingly, medical appointments are kept at a higher rate when patients are reminded that they have an upcoming appointment (e.g., Almong et al., 2003; Opon et al., 2020). Presumably the amount of effort required to show up and whether one remembers they have scheduled an appointment for a survey interview both affect the "show rate" in much the same way. Note that the advantage in show rates observed for telehealth compared to in-person clinic visits is unlikely to directly correspond to show rates for live video versus in-person survey interviews. Scheduling an appointment to be interviewed is unlikely to be as urgent (and may be more burdensome) to a respondent than scheduling an appointment to be examined by a doctor is to a medical patient: one's health almost surely motivates a patient to schedule and show up for a video medical meeting more than a survey researcher's research question motivates a sample member to schedule and show up for a video survey interview.

The analogy between telemedicine appointments and video interview appointments is not perfect: in the former it is travel by the patient that is made unnecessary by video, while in the latter it is travel by the interviewer – not the respondent – that is eliminated. But, despite these differences, the two types of scheduled video interaction place some additional burden on either the patient or respondent: both require the patient/respondent to navigate to the video platform with which they may be unfamiliar, possibly use a headset, resolve technical problems, etc. just to get started. None of those are the case for in-person clinic visits or in-person survey interviews.

Figure 1 displays the hypothetical loss of cases, i.e., the growth in nonresponse or "drop-off," at each stage in the recruitment process for live video survey interviews. It seems likely that there will be substantial loss (i.e., nonresponse) between invitation and scheduling. In addition to finding live video interviewing potentially more intrusive or unsettling (as one can imagine sample members might perceive video interviewing) than in-person or telephone interviews, the extra effort required to schedule an appointment may be a substantial deterrent to participation. As the medical appointment literature makes clear, not everyone who has scheduled an appointment shows up, even when one's health is involved, which is almost surely more motivating than the promise of having one's voice heard in a survey. Of those who do show up for their appointment, a small subset will not start the interview because, for example, technical problems prevent going forward. Thus, the proportion of invited sample members who complete the data collection process (i.e., the respondents) may be smaller than in modes that do not require scheduling and showing up, as these are decision points at which the video sample member may choose not to participate.

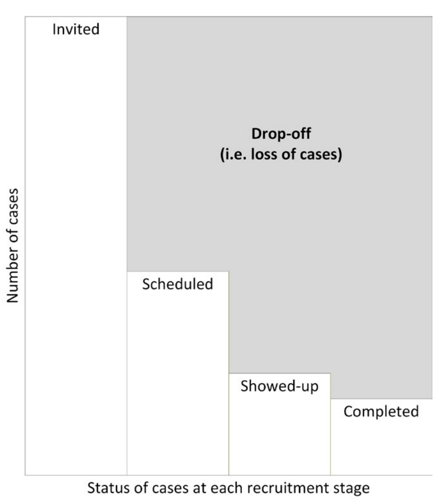


Figure 1 Conceptualization of live video interview drop-off (loss of cases) from invitation to completion

Note that in Figure 1, the white bars to the right of "Invited," such as "Scheduled," represent sample members who are not yet respondents: they have not started or completed the interview and could drop off at the next stage, but they are not yet nonrespondents. They are akin to sample members in a telephone survey who have neither declined nor accepted a telephone invitation to participate but have agreed to be called back to further discuss their participation: they are still "live prospects" (see Conrad et al., 2013). This leads to our first research question:

RQ1: *What percent of sample members drop-off at each recruitment stage (including scheduling, showing up, completion) for live* video interviews?

Breakoffs

Nonresponse can also take the form of starting but not finishing an interview, i.e., breaking off. One factor that might affect breakoffs (or its complement, completion) is the extent to which a participant is familiar with and prefers interacting in the mode to which they have been assigned. Olson et al., (2012) demonstrated that completion is higher when the mode of data collection matches participants' articulated mode preference (collected in a survey conducted in the previous year). The authors tested the effect of mode preference on completion of web, mail, and phone surveys; we assume that the mechanism that they tapped into is in effect present in other modes as well, including live video interviews. If a sample member has started a live video interview, they are, in principle, more likely to complete it if they prefer interacting with a live, visible person than self-administering a textual questionnaire.

RQ2: *Do sample members who are likely to prefer interacting with a live human interviewer complete live video interviews at a higher rate (i.e., break off less often) than do sample members who are likely to prefer responding to self-administered, text-based web questionnaires?*

Because web surveys – which are typically text-based – are ubiquitous, many sample members invited to prerecorded video interviews may expect an experience similar to what they are familiar with in text-based web surveys when agreeing to participate and may expect to read and respond to textual questions, not to watch and listen to a video-recorded interviewer ask questions.[3] Prerecorded video interviewing is fundamentally multimodal (spoken question and visible, moving interviewer), which may be aversive to respondents who are used to a text-only interface, and perhaps more importantly, silence and stillness in the interface. Similarly, respondents may find the slow, spoken delivery

[3] We note it is possible to implement pre-recorded video interviews that include text of the question as well as the video-recorded presentation of that question so the experience may not be entirely different from a text-based web survey, but will include audio and video.

of questions – slower than the speed at which they may be able to read them – unpleasant, especially if they cannot "barge in" to circumvent the spoken question. It is possible that this combination of factors, i.e., the disconnect between properties of the mode and what respondents expect, slow question delivery, and inability to barge in, could increase breakoffs in prerecorded video interviews compared to traditional text-based web surveys. Moreover, despite the presence of a recorded interviewer, those who have started the interview are sure to recognize that ending the interaction will have no social costs, e.g., they will not offend the interviewer whose recordings they have been responding to. On the other hand, the recorded interviewer's presence could reduce breakoffs relative to a textual web survey: respondents seem to react socially to prerecorded interviewers, almost as if they feel a modest level of "rapport" with them even though they recognize the recorded interviewer's inanimacy (e.g., Conrad et al., 2023).

RQ3: *Will sample members who are primarily familiar with text-based web surveys complete a text-based web survey at a higher rate (i.e., break off less often) than they would complete a prerecorded video interview?*

The pattern of breakoff across the questionnaire could differ depending on the mode. In Live Video interviews, respondents could feel obliged to continue through the questionnaire, having established a relationship with the interviewer, albeit mediated. But the relationship may take time to develop, suggesting that if there are breakoffs in live video interviews, they will occur early. Thus, breakoffs could be rare across the entire interview. Moreover, those sample members who have scheduled and shown up for the appointment may be predisposed to complete the interview. Similarly, in web surveys, respondents have been shown to break off most frequently on the early pages (Mittereder & West, 2022; Peytchev, 2009) so if they "survive" beyond that point, they are likely to complete the questionnaire. For prerecorded video interviews, the location of breakoffs is harder to anticipate as the mode has not been widely evaluated. On the one hand, it is possible that the respondents will feel motivated by recorded interviewers to complete the study, but on the other, they may find the movement and sound of the video recording to be distracting and annoying, a reaction that could become more prevalent as the interview progresses. In this case, breakoffs would increase over the questionnaire.

RQ4: *What is the pattern of breakoffs across the questionnaire when (a) a live interviewer asks the questions, (b) a prerecorded interviewer asks the questions, and (c) there is no interviewer, (i.e., a textual web survey)?*

Breakoffs may be related not only to the interview mode but also to the device the respondent uses. It is well known that breakoffs in web surveys are more prevalent when respondents use a mobile device than a computer (desktop or laptop) (see Couper et al., 2017 for a review) due, among other reasons, to the small screen size and its effect on entering responses. On the one hand, we might expect that breakoffs will be equally likely for live video interviews conducted on mobile and desktop devices, because respondents register their answers via speech in both cases, i.e., they do not enter responses by typing or tapping, minimizing the effect of screen size. Moreover, they are likely to have established a comparable relationship with the interviewer on both devices similarly affecting breakoffs[4]. On the other hand, to the extent the interviewer presents textual or graphical materials, breakoffs could be higher on mobile devices because these materials may well be harder to read or view on a small screen. Moreover, in a live video interview on a mobile device the likelihood is greater that the respondent will be in an environment that is not conducive to sustained attention, e.g., while in transit, than if the interview is conducted on a computer.

Because prerecorded video interviews share at least some features with both textual web surveys and live video interviews, it is natural to ask how devices affect breakoffs in prerecorded video interviews. It is possible that respondents will feel connected with the video-recorded interviewers to the same extent on both mobile devices and computers, leading to similar breakoff rates across the interview. Alternatively, if the mode is implemented so that respondents must select answers from on-screen options (as opposed to speaking their answers), the small screen in mobile prerecorded video interviews could inflate breakoffs relative to breakoffs on a computer. In addition, designers may have less control over how video recordings behave on mobile devices than computers, which could lead to less usable interfaces and thus increased breakoffs.

RQ5: *How do breakoffs compare between mobile devices versus computers, for (a) live video interviews, (b) prerecorded video interviews, and (c) conventional text-based web surveys?*

Data and Methods

The data that this paper is based on were collected between November 2019 and March 2020, in an experimental study (reported in Conrad et al., 2023; see also West et al., 2022) that compared data quality and respondent experience in three data collection modes: (a) live video[5] (i.e., two-way, video-mediated exchanges), in which nine professional interviewers read 36 questions out loud and respondents spoke their answers that the interviewer then recorded; (b) prerecorded

[4] In fact, the proportion of the screen devoted to the interviewer's face could well be higher on a mobile than desktop device, potentially hyperfocusing the respondent on the interviewer.

[5] See https://www.mivideo.it.umich.edu/media/t/1_1zoid4cu for an example of the live video interview.

video,[6] in which separate video recordings of the same nine interviewers each asking the same 36 questions (i.e., one video recording of each interviewer asking each question) were embedded in a web survey, and respondents answered by selecting or typing in their answers; and (c) a conventional text-based web survey[7] in which respondents read on-screen, textual questions and directly entered their answers by selecting or typing them in. It was conducted to learn about the two video modes and not to produce population estimates. In the current article, we examine factors affecting participation, i.e., recruitment, starting, and completing (or its complement, breakoffs) in the three modes.

The live video interviews were conducted in a web browser so that respondents did not need to download and install an app. The prerecorded video interviews were designed so that the question was only presented orally, and response options were only displayed after the video finished playing; this was done to make these interviews more comparable to live video interviews in which interviewers were trained to read all the response options before accepting answers. On desktop and laptop computers the videos auto-played after the page loaded; on mobile devices videos did not auto-play due to a software limitation, requiring respondents to manually start each video. Prerecorded video and the text-based web survey were optimized for smaller screens by implementing wide response buttons (see Figures 2 and 3).

6 See https://www.mivideo.it.umich.edu/media/t/1_vjhtigaf for an example of the prerecorded video interview.

7 See https://www.mivideo.it.umich.edu/media/t/1_82z2zs7y for an example of the web survey.

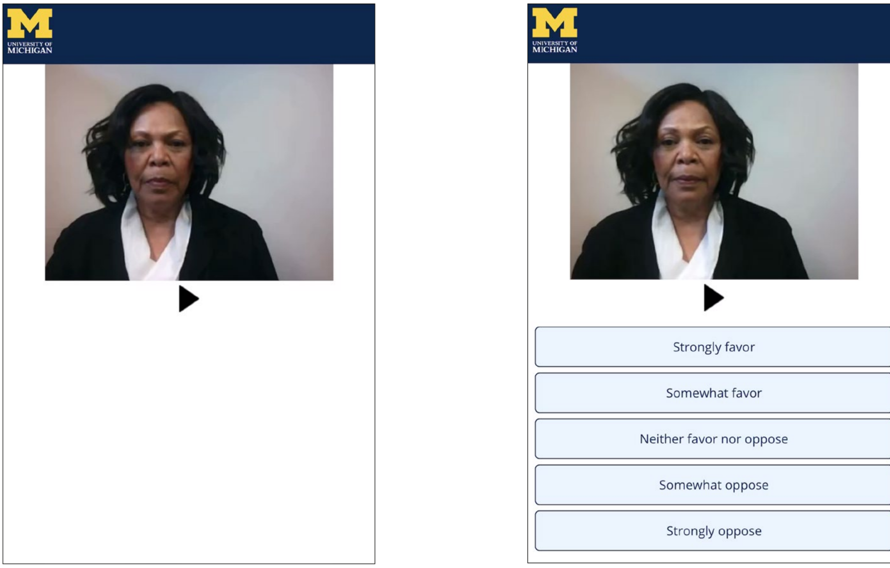


Figure 2 Respondent's mobile screen before (left) and after (right) the video has played in Prerecorded Video

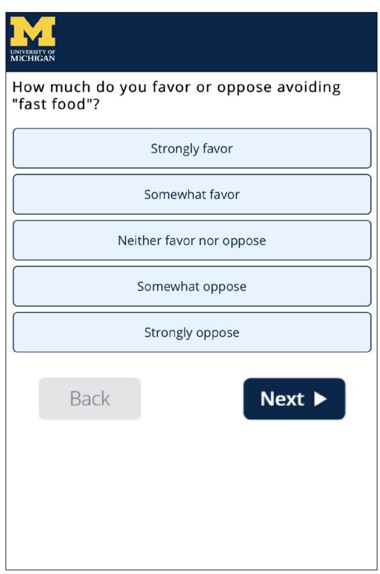


Figure 3 Respondent's mobile screen in text-based Web Survey

Sample Source

Initially, we attempted to recruit participants via Address-Based Sampling (ABS), but ABS yielded completed interviews from less than 1% of invited live video households[8] (see Conrad et al. 2023), although more reasonable outcomes (~8-10%) for prerecorded video and the text-based web survey were obtained, so we do not include the few completed cases recruited via ABS in any analyses. Due to the low yield of interviews with ABS, we explored two opt-in (i.e., non-probability) sample sources, one a conventional web panel (WebPanel) and one a clinical research panel (ClinicalPanel), because (a) they facilitated recruiting (being created for participation in research studies), and (b) their non-probability character was not a concern because the goal of the study was to understand the viability of live and prerecorded video interviewing for collecting survey data, not to produce population estimates. Sample members were recruited randomly from both sample sources into one of the three survey modes.

Members of these two sample sources typically engage in quite different types of research studies. Members of the WebPanel typically participate in conventional self-administered, text-based surveys in which the interface is static and silent, i.e., no human presence. Thus, panel members rarely, if ever, interact with a human interviewer, potentially finding such an interaction more or differently demanding than they would like, which might in turn depress their participation in live video interviews. And they certainly do not expect participation to require scheduling a time in which to complete the interview, as they expect to participate at a time of their choosing. They also are unlikely to expect the questions in prerecorded video interviews, in effect, a web survey to be presented by recorded interviewers that move and speak and, thus, may find this unappealing, potentially depressing completion of prerecorded video interviews.

In contrast, members of the ClinicalPanel typically participate in health and medical research, often in person, e.g., to undergo blood draws, imaging, blood pressure measurement, etc., so a remote but live interaction is probably more similar to what they're familiar with than it would be to WebPanel members; yet it is unlikely that, at the time of our study, they had ever participated in a live video interview. The similarity of live video interviewing to the in-person research that ClinicalPanel members were familiar with seems likely to promote participation in live video interviews. And the requirement to schedule an appointment is a component of clinical research – one cannot usually visit the medical facility at a time of their choosing – so scheduling an appointment for

[8] In the ABS pilot we invited 300 sample members to participate in a live video interview, which resulted in 2 completed interviews; in prerecorded video we invited 150 sample members, and this resulted in 12 completed interviews; in web we invited 150 sample members, and this resulted in 14 completed interviews.

a live video interview is unlikely to create an insurmountable obstacle. Taken together, willingness to schedule a data collection session and to interact with live interviewers should promote completion of live video interviews for ClinicalPanel members compared to WebPanel members.

Similarly, we would expect ClinicalPanel members to complete prerecorded video interviews at a higher rate than WebPanel members because the former group is likely to be unfazed by the (recorded) presence of a human interviewer who asks questions compared to the latter group. Moreover, if the ClinicalPanel members are willing to schedule an appointment for a live video interview, they should be at least as willing to start a prerecorded video interview, which requires only clicking a link when they are ready to begin (as in textual web surveys). Thus, we expect these factors to promote participation in the web survey and prerecorded video interviews for members of the ClinicalPanel relative to WebPanel members. Taken together, these considerations suggest it is likely that members of the two panels differ substantially in their preference of survey modes, based on the modes they are most familiar with and the types of research tasks they regularly carry out and are comfortable with.

Survey Invitation

Sample members in each mode only received one invitation to the study (i.e., no follow-up invitations). Those assigned to prerecorded video and the web survey were sent customized invitations containing a link to the survey. Those assigned to live video interviews were given a link to a scheduling application in an invitation (email for ClinicalPanel members, internal online invitation for WebPanel members), listing available interview times within the next two weeks to select from, and were asked to provide an email address. Once a sample member had selected an interview time, they were sent a confirmation email with the day and time of the appointment and the link to use for the interview. They were reminded of the appointment via email and, if they provided a phone number (which most people did), a text message. These were sent the day before the scheduled appointment, a few hours before the scheduled time on the day of the appointment, and five minutes after the start of the appointment if the sample member had not shown up for the interview.

All live video invitees were offered $20 to complete the interview; prerecorded video and web survey invitees were offered $5, except ClinicalPanel invitees, who were offered $20 for comparability with typical clinical research (see Conrad et al. 2023, Online Appendix G, for details).

Results

RQ1: What percent of sample members drop-off at each recruitment stage (including scheduling, showing up, completion) for live video interviews?
The need to schedule a live video interview clearly increased nonresponse (see Table 1). Of those who were invited to participate, only 10.3% scheduled an appointment, i.e., 89.7% dropped off. However, the scheduling rate differed substantially between the two sample sources, with 100% of the invited ClinicalPanel members scheduling an appointment, but only 5.6% of the invited WebPanel members scheduling an appointment. We return to this in the Discussion.

Of those who did schedule an appointment, less than half (49.1%) showed up for the interview. But of those who showed up, most (98.3%) completed the interview. Although those who showed up were likely to complete the interview, only 4.9% of the 5,783 sample members who were invited completed a live video interview. The patterns of scheduling, showing up for, and completing live video interviews are generally consistent with the conceptualization presented in Figure 1, although for the ClinicalPanel, scheduling was not an obstacle. Clearly, scheduling and showing up for an appointment – which are critical to live video interviewing – can be obstacles to participation for at least some sample members.

Table 1 Count and percentage of cases that remain at each recruitment stage for live video interviews

Invited	Scheduled an appointment	Showed up for appointment	Showed + Completed interview
5,783	593 (10.3%)	291 (49.1%)	286 (98.3%)

RQ2: Do sample members who are likely to prefer interacting with a live human interviewer complete live video interviews at a higher rate (i.e., break off less often) than do sample members who are likely to prefer responding to self-administered, text-based web questionnaires?
As Table 2 shows, those who started live video interviews completed them at high rates. Sample source (which we consider a proxy for mode preference) had little impact: 95.0% of WebPanel members who started, completed it; 99.5% of those ClinicalPanel members who started completed it. While breakoffs were not sensitive to sample members' mode preferences, response rates were. Looking at those who are likely to prefer responding to a self-administered, text-based questionnaire (i.e., WebPanel), only 1.4% of the invited sample members completed the interview. In comparison, of those who were likely to prefer interacting with a live human interviewer (i.e., ClinicalPanel), 74.2% of invited sample members

completed the interview. Clearly, the kind of interaction that sample members were used to had a major impact on response.

Table 2 Completion and response rates (AAPOR RR1) by mode by sample source

Mode	Sample Source	Invited	Showed up / Started	Completed	Completion Rate	Response Rate
Live Video	Total	5,783	291	286	98.3%	4.9%
	WebPanel	5,500	80	76	95.0%	1.4%
	ClinicalPanel	283	211	210	99.5%	74.2%
Prerecorded Video	Total	1,230	563	414	73.5%	33.7%
	WebPanel	1,120	478	330	69.0%	29.5%
	ClinicalPanel	110	85	84	98.8%	76.4%
Web Survey	Total	552	445	422	94.8%	76.4%
	WebPanel	445	379	356	93.9%	80.0%
	ClinicalPanel	107	66	66	100%	61.7%

RQ3: Will sample members who are primarily familiar with text-based web surveys complete a text-based web survey at a higher rate (i.e., break off less often) than they would complete a prerecorded video interview?
For sample members who were familiar with text-based web surveys (WebPanel members), most who started the web survey completed it (93.9%), while far fewer of those who started a prerecorded video interview completed it (69.0%). Despite prerecorded video being similar in many ways to a conventional text-based web survey, these respondents clearly experienced the two modes differently. We cannot disentangle whether it was the presence of the video-recorded interviewer, including their voice, the implementation of the mode (e.g., it was not possible to respond until the video recording had finished playing), or an interaction of the two that led to the relatively low completion rate for these sample members, but clearly they were less willing to persevere in completing this type of web survey, i.e., with an embedded recording of an interviewer asking questions, than a conventional text-based web survey. This phenomenon is even more evident when looking at response rates (80.0% vs. 29.5%, AAPOR RR1).

RQ4: What is the pattern of breakoffs across the questionnaire when (a) a live interviewer asks the questions, (b) a prerecorded interviewer asks the questions, and (c) there is no interviewer (i.e., a textual web survey)?
There were very few breakoffs in live video interviews. The five sample members who started but did not finish the interview in this mode (see Table 3, row

1) broke off at different points, i.e., it does not seem that landing in a live video interview by itself led them to abandon the interview, nor was there one question that seemed to lead them to break off (see Figure 4), but the main point is that breakoffs were of negligible frequency for live video interviews.

Conversely, breakoffs were very frequent in prerecorded video: more than a quarter of respondents broke off (see Table 3, row 2). In a logistic regression model[9] breakoffs are 7.24 times more likely than in the web survey (Table 4). Breakoffs occurred on almost every question (see Figure 4). The breakoffs in this mode were driven largely by the sample of WebPanel members (see Table 2).

Table 3 Breakoffs across the three modes, sample source, and device

		Breakoffs	Proportion
Mode	Live Video (291)	5	0.02
	Prerecorded Video (563)	149	0.26
	Web Survey (445)	23	0.05
Sample Source	WebPanel (937)	175	0.19
	ClinicalPanel (362)	2	0.01
Device	Mobile (646)	122	0.19
	Computer (651)	55	0.08

Table 4 Breakoff predicted by mode, device, and sample source

Term	Estimate	Standard error	Odds ratio	Statistic	*p*
(intercept)	−6.35	0.76	< 0.01	−8.35	< .01
Mode: Live Video (ref=Web)	0.12	0.52	1.13	0.23	.82
Mode: Prerecorded Video (ref=Web)	1.98	0.24	7.24	8.32	< .01
Sample Source: WebPanel	3.16	0.73	23.57	4.35	< .01
Device: Mobile	0.65	0.19	1.92	3.49	< .01

9 Implemented in the glm() function in R. The model predicts breakoffs based on mode (Live Video, Prerecorded Video, or Web Survey), sample source (WebPanel or ClinicalPanel), and device (mobile or computer).

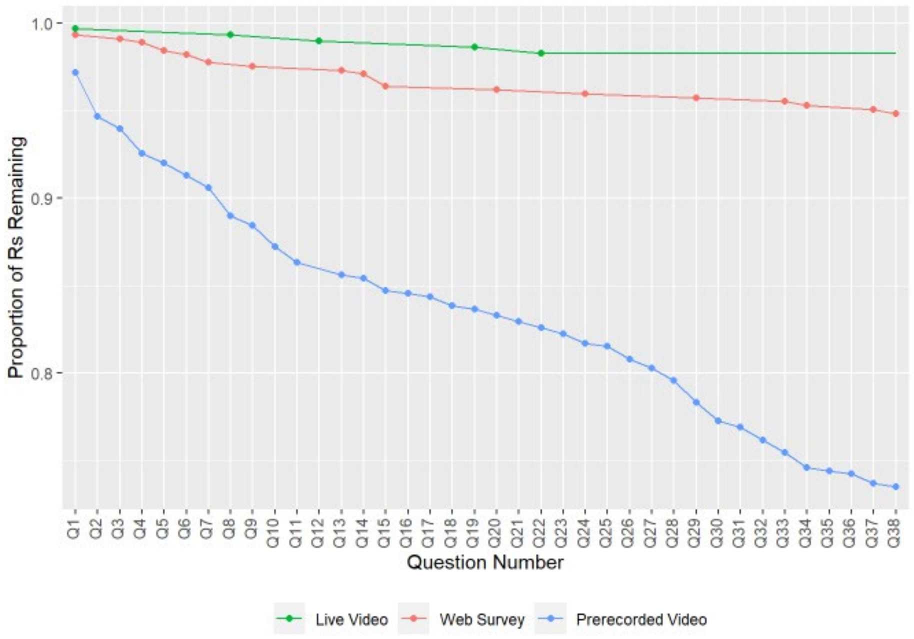


Figure 4 Proportion of respondents (i.e., sample members who started the questionnaire) remaining at each question in each mode. Each point represents one or more breakoffs; the proportion of respondents remaining at each question is represented on the Y-axis; individual questions are represented on the X-axis.

Similar to live video, web survey respondents did not break off in high numbers. The 23 respondents who did break off did so on different questions. All of the 23 respondents who broke off were from the WebPanel. We attribute these outcomes to the ubiquity of web surveys in general and most sample members' familiarity with the mode. A textual web survey was apparently far more appealing – and presumably less surprising – to the WebPanel sample than was a web survey in which a video-recorded interviewer was embedded.

Finally, the vast majority of breakoffs across all modes involved respondents from the WebPanel (see Sample Source rows in Table 3): these respondents were 24 times more likely to break off than were respondents from the ClinicalPanel (Odds Ratio, Table 4). This difference is almost surely driven by breakoffs in prerecorded video: 148 of the 149 breakoffs were from the WebPanel, whereas the breakoffs in the other modes were lower in the two sample sources: in live video, there were only five breakoffs (four of which were from the WebPanel), and in the web survey, there were only 23 breakoffs (100% of which were from the WebPanel).

RQ5: How do breakoffs compare between mobile devices versus computers for (a) live video interviews, (b) prerecorded video interviews, and (c) conventional text-based web surveys?

Breakoffs were higher for respondents using mobile devices (19%) than computers (8%) as shown in Table 3, and this is highly significant in the logistic regression model (bottom row of Table 4). The effect of the device was almost entirely evident in the two self-administered modes: mobile users accounted for 65.2% of the breakoffs in web surveys and 71.5% of the breakoffs in prerecorded video (there were too few breakoffs to analyze in live video interviews).

The higher breakoff rate for mobile devices in prerecorded video than web surveys is likely due to the smaller screens on mobile devices than computers, even though wider-than-usual response buttons were implemented in the mobile user interfaces for web and prerecorded video. Breakoffs on mobile devices were even higher in prerecorded video than the web survey, which we attribute to implementation: due to a constraint of the interviewing software at the time, videos on mobile devices did not autoplay, requiring respondents to tap/touch “Play” to start the video; each additional action may well have pushed them to break off (Conrad et al. 2006; Gray & Fu, 2004).

Discussion

The current study provides initial evidence about the challenges researchers face promoting participation in both live and prerecorded video interviews. Because nonparticipation takes different forms in both video modes, minimizing it will require different strategies for the two modes.

As we proposed, participating in a live video interview requires sample members to overcome multiple obstacles and opportunities to drop off: they need to *schedule* an interview, *show up* at the appointed time and start the interview, and they need to *complete* the interview once they start it.

Scheduling a live video interview. In the current study, 10% of invited sample members scheduled an appointment on average, but the scheduling rate differed substantially between sample sources, with 100% of the invited ClinicalPanel members scheduling an appointment! In contrast, only 5.6% of the invited WebPanel members scheduled an appointment. This difference in scheduling rate likely reflects differences in the sample’s familiarity with and tolerance for scheduling an interview appointment, i.e., those whose research experience involves participating when they choose, prefer not to schedule. This might suggest that live video interviewing is well suited for studies where the same person is interviewed multiple times. The initial survey can be done in person, with the interviewer helping to schedule follow-up video interviews if they are in the not-too-distant future. Moreover, because the interviewer can collect the sample member's contact information, including email, in that initial visit, a sample member could be invited to schedule a video interview closer to the next data collection, or be reminded via email about any previously scheduled appointment. In addition, by virtue of their familiarity with and trust in the survey organization established over multiple interviews, respondents may be more open to use an interviewing mode with which they have no prior experience.

While there could be many reasons that sample members do not schedule a live video interview when they are invited to, one reason might be that they simply are not sufficiently motivated to do so. This might be addressed by incentivizing (e.g., paying) them to both schedule and show up for the interview as well as to complete it. If reluctance to schedule an interview is related to concerns about privacy (as suggested in Schober et al., 2023), recruitment could benefit from informing respondents what measures have been taken to protect their data and maintain confidentiality. Similarly, scheduling rates may be low because members of the public lack trust in science, government, or technology, and video communication – especially for collecting survey data – may exacerbate this lack of trust. Building a relationship with the sample member, as in a longitudinal panel study, seems likely to help reassure sample members' responses and identities are secure.

Showing up for a scheduled live video interview. Once the sample members scheduled an interview, the percentage who showed up at the appointed time was a fraction of those who were scheduled to participate. Similarly, Guggenheim et al. (2026) reported that in a pilot study of live video interviewing only about half showed up for the scheduled interview. So far, the evidence suggests that there can be substantial loss of cases because sample members do not honor their appointment. It is unclear if the rate of scheduling and honoring an appointment for a live video interview is better or worse than in phone or in-person interviews, but it is essential in live video interviews in way that it's not for phone on in-person interviews. Showing up in live-video involves honoring a commitment to show up at a certain time; showing up in a prerecorded video interview or web survey is a matter of clicking when it is convenient.

This raises the question of how researchers can increase the show rate in live video interviews. Among the possible approaches to reducing no-shows in this mode are:

- **Remind sample members about the upcoming appointment.** If an in-person interview was conducted initially, the interviewer can leave an appointment card. In addition, the organization can vary the modes[10] in which they remind the sample member about the upcoming appointment.

[10] Email and text message reminders can include a calendar invitation.

- **Require eCheck-in to confirm an appointment.** Asking sample members to confirm their appointment may increase their commitment to the study, analogous to asking respondents to commit to answer questions conscientiously (e.g., Cannell et al., 1981; Cibelli et al., 2020). This approach is widely used for self-scheduling in domains from medical appointments to restaurant reservations.
- **Make it as easy as possible to join.** Ideally, respondents will be able to join the scheduled interview with a single click in a browser, i.e., without it being necessary to download an app. For comparability across respondents, the video platform should perform similarly across operating systems and devices, which should in turn lead to easier procedures for joining.

We note that none of these are likely to be a panacea. Confirmation of appointments and ease of joining were implemented to a certain degree in this study, and no-show rates were nonetheless high. It is possible that no-show rates would have been higher had we not implemented these measures, but without an experimental comparison we cannot say for sure.

Failure to show up for a live video interview at the appointed time can create monetary costs that are not at issue in other modes: live video interviewers will likely be paid to wait for a respondent who does not show up on time and ultimately does not start the interview. Depending on the protocol, e.g., requiring interviewers to wait 10 minutes before reclassifying the case as a no-show, the cost of these unproductive intervals can add up. Costs could be defrayed if interviewers are given other low-priority tasks to carry out while waiting that could be terminated if a respondent joins the video session, but any waiting is costly. One way to reduce interviewer waiting due to no-shows is to intentionally overbook appointments, much as airlines overbook flights (see Zacharias & Pinedo, 2014 for an example of overbooking medical appointments to reduce costs).

Completing a live video interview. Almost all respondents (98%) who started a Live Video interview completed the interview. This high level of completion could be a selection effect, i.e., the requirements of scheduling and showing up for a scheduled interview could have filtered out sample members at higher risk of breaking off[11] had they started the live video interview. Alternatively, it could be a function of the interviewer's social presence, i.e., just as terminating an in-person interview midstream is something most respondents would consider

11 Partial completion (breakoffs) and unit nonresponse in interviews may similarly be affected by burden (Bradburn, 1978). One subgroup that is at an increased risk of unit nonresponse and, by extension breakoff, is sample members with the lowest levels of education (Spitzer, 2020). Those with the lowest levels of education (High School or less) were underrepresented in our sample (see Conrad et al., 2023), so it is possible that had they been represented more accurately, breakoffs in live video interviews could have been higher. However, the hurdles required to start a live video interview may well have filtered out members of this group, irrespective of its size, before they could start the interview.

rude, live video respondents could be similarly reluctant to break off. Or it could be some combination of selection and interviewer presence. We cannot determine this based on the data available in the current study, but if the high rate of completion is due at least in part to interviewer presence (and not selection), increasing the rate of scheduling interviews and showing up for scheduled interviews as just described should also increase the proportion of invited sample members who ultimately complete the interview (AAPOR RR1).

Note that as live video interviewing is used more, researchers will need to develop standard definitions of participation metrics, e.g., scheduling rates and show-up rates. This may become particularly tricky in mixed-mode studies since live-video relies heavily on scheduling and thus presents multiple opportunities for nonresponse i.e., for sample members to drop off, which is not usually the case in other modes, complicating how participation is quantified across modes. Moreover, if a sample member shows up at the appointed time, but only to reschedule, or is unable to start due to technical problems, it is hard to know how these scenarios should be classified with current participation metrics.

Completing a prerecorded video interview. Breakoffs were higher in prerecorded video than either live video interviews or the conventional text-based web survey. One likely reason for this concerns the extent to which the features of this mode are consistent with respondents' mode preferences, which seem likely to differ depending on how they communicate every day, their comfort with and trust in technology, how well the particular implementation of the mode comports with their expectations, etc. In the current study, we attribute the higher breakoff rate among the WebPanel than ClinicalPanel members to the former's preference for a text-only experience. Moreover, our implementation required them to play the entire video before being able to register their answer and, on mobile devices, to press a button to play each video. Conversely, Strojil and Cígler, 2025, report lower breakoff rates in prerecorded video than in a textual web survey, presumably because they implemented the former mode so that respondents could answer while the video was playing, and the video auto-played on all devices.

To increase completion in prerecorded video interviews, we propose:

- **Explicitly alert sample members in the invitation that the questionnaire will be administered by a recorded human interviewer, producing speech and visual activity.** This may deter some sample members from starting but could be preferable to a high breakoff rate as the nature of the interview experience becomes clear to respondents; instead, upfront clarity about the features of the mode should help assure that sample members start the interview with appropriate expectations, increasing the chances they will complete it. If the goal is to reach a sample size target rather than maximizing the response rate, researchers can issue as many invitations as are needed to reach the target without incurring monetary costs – at least in a panel such as the WebPanel used in the current study.

- **Allow respondents in a prerecorded video interview to submit a response before the video has finished playing.** In the current study respondents could not answer the question while the video was playing to promote comparability with the administration of the live video interview in which the standardized interviewers were trained to read the questions entirely before accepting a response. That constraint is unlikely to be present in production surveys using prerecorded video interviews.
- **Present question text, including response options, along with the video.** Since some sample members will read faster than the recorded interviewer speaks, textual redundancy with the video recording will allow them to respond as soon as they have come up with the answer. However, this approach may undermine the data quality benefits of prerecorded video interviews (Conrad et al., 2023).
- **Allow the video recording to autoplay on all devices.** This was not possible on mobile devices when the mode was implemented, at least using the platform we used, and seemed to be a deterrent to completion: breakoffs were more than twice as high for respondents on mobile than desktop devices. Clearly, auto-playing video recordings on mobile devices is now possible, as evidenced by the Strojil and Cígler (2025) study.
- **Prepare the respondent for a prerecorded video interview.** The amount of control over the video afforded to respondents should be made clear prior to the start of the interview. This includes whether the video will autoplay, whether respondents can adjust the rate at which the video plays, and whether they can replay, pause, or stop the video.

The current study was based on data derived from an experiment that was designed to investigate the effects of three modes, i.e., a live video interview, a prerecorded video interview, and a textual web survey, on response quality. Thus, it was not designed to examine participation across the three modes. Nonetheless, the effects of the three modes on participation were undeniable, despite our repurposing the data. Further, the two non-probability sample sources were something of a double-edged sword. On the one hand, the very different participation behavior between the two sample sources complicates the mode comparison. On the other hand, this created a natural experiment allowing us to compare the effects of respondent familiarity with a mode, and likely expectations about their subsequent experience.

Online panels, both non-probability and probability, are potentially well-suited for live video interviews, as members of both types of samples can be easily invited, in contrast to members of the public whose email addresses are generally unknown to researchers. Exploring the extent to which the results reported here generalize to other online sample sources is an important next step, especially following the explosive growth in two-way video communication during the pandemic, just after the current data were collected.

Our implementation of prerecorded video interviews was constrained in many ways and probably less optimal than we would have liked. Some version of this approach may become relatively common. For example, Spiegelman and Conrad (2025) found that audio presentation of definitions in a web survey improved response quality relative to text only, and both audio and text led to even higher quality results. Similarly, as LLM-powered chatbots and avatars become more common, it is likely they will be incorporated into survey data collection. Thus, members of different kinds of panels may accept the successors of prerecorded video interviews to a greater extent than was generally the case when the current data were collected.

It seems likely that neither live nor prerecorded video interviews are ideal as stand-alone survey modes. Instead, they seem likely to be most helpful to researchers when they are implemented within a suite of modes from which sample members can choose the mode in which they participate. Video may not be for everyone, but for some it might be just right.

Funding

We gratefully acknowledge financial support from NSF grants SES-1825113 and SES-1825194 (Methodology, Measurement, and Statistics program) and NIA grant P30 AG012846 (Michigan Center on the Demography of Aging), and access to participants through NCATS grant UL1TR002240 (Michigan Institute for Clinical & Health Research).

Conflict of Interest

The authors report no potential conflicts of interest with respect to the research authorship, or publication of this article.

References

Alkilany, R., Tarabichi, Y., & Hong, R. (2021). Telemedicine visits during COVID-19 improved clinic show rates. *JACR Open Rheumatology, 4*(2), 136-141. https://doi.org/10.1002/acr2.11372

Almong, D. M., Devries, J. A., Borrelli, J. A., & Kopycka-Kedzierawski, D. T. (2003). The reduction of broken appointment rates through an automated appointment confirmation system. *Journal of Dental Education, 67*(9), 1016-1022. https://doi.org/10.1002/j.0022-0337.2003.67.9.tb03684.x

Anderson, A. H. (2008). Video-mediated interactions and surveys. In F. G. Conrad & M. F. Schober (Eds.), *Envisioning the survey interview of the future* (pp. 95-118). Wiley. https://doi.org/10.1002/9780470183373.ch5

Bradburn, N. M. (1978). Respondent burden. Proceedings of the American Statistical Association, survey research methods section (pp. 35-40). Alexandria, VA: American Statistical Association. http://www.asasrms.org/Proceedings/papers/1978_007.pdf

Cannell, C. F., Miller, P. V., & Oksenberg, L. (1981). Research on interviewing techniques. *Sociological Methodology, 12*, 389-437. https://doi.org/10.2307/270748

Cibelli Hibben, K., Felderer, B., & Conrad, F. G. (2020). Respondent commitment: applying techniques from face-to-face interviewing to online collection of employment data. *International Journal of Social Research Methodology, 25*(1), 15-27. https://doi.org/10.1080/13645579.2020.1826647

Conrad, F., Broome, J., Benkí, J., Kreuter, F., Groves, R., Vannette, D., & McClain, C. (2013). Interviewer speech and the success of survey invitations. *Journal of the Royal Statistical Society Series A: Statistics in Society, 176*(1), 191-210. https://doi.org/10.1111/j.1467-985X.2012.01064.x

Conrad, F. G., Couper, M. P., Tourangeau, R., & Peytchev, A. (2006). Use and non-use of clarification features in web surveys. *Journal of Official Statistics, 22*(2), 245-269. https://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/use-and-non-use-of-clarification-features-in-web-surveys.pdf

Conrad, F. G., Schober, M. F., Hupp, A. L., West, B. T., Larsen, K. M., Ong, A. R., & Wang, T. (2023). Video in survey interviews: Effects on data quality and respondent experience. *Methods, Data, Analyses, 17*(2), 135-170. https://doi.org/10.12758/mda.2022.13

Conrad, F. G., Schober, M. F., Nielsen, D., & Reichert, H. (2020). Virtual interviewers, social identities, and survey measurement error. In *Interviewer Effects from a Total Survey Error Perspective* (pp. 149-164). Chapman and Hall/CRC. https://www.taylorfrancis.com/chapters/edit/10.1201/9781003020219-15/virtual-interviewers-social-identities-survey-measurement-error-frederick-conrad-michael-schober-daniel-nielsen-heidi-reichert

Couper, M.P., Antoun, C., & Mavletova, A. (2017). Mobile Web Surveys. In Total Survey Error in Practice (eds P.P. Biemer, E. Leeuw, S. Eckman, B. Edwards, F. Kreuter, L.E. Lyberg, N.C. Tucker and B.T. West). https://doi.org/10.1002/9781119041702.ch7

Dantas, L. F., Fleck, J. L., Cyrino Oliveira, F. L., & Hamacher, S. (2018). No-shows in appointment scheduling - a systematic literature review. *Health Policy, 122*(4), 412-421. https://doi.org/10.1016/j.healthpol.2018.02.002

DeBell, M., Amsbary, M., Brader, T., Brock, S., Good, C., Kamens, J., Maisel, N., & Pinto, S. (2022). Methodology Report for the 2020 ANES Time Series Study. Palo Alto, CA and Ann Arbor, MI: Stanford University and the University of Michigan. https://electionstudies.org/wp-content/uploads/2022/08/anes_timeseries_2020_methodology_report.pdf

Deyo, R. A., & Inui, T. S. (1980). Dropouts and broken appointments. *Medical Care, 18*(11), 1146-1157. https://www.jstor.org/stable/3764264

Drerup, B., Espenschied, J., Wiederner, J., & Hamilton, L. (2021). Reduced no-show rates and sustained patient satisfaction of telehealth during the COVID-19 pandemic. *Telemedicine and e-Health, 27*(12), 1409-1415. https://doi.org/10.1089/tmj.2021.0002

Endres, K., Hillygus, D. S., DeBell, M., & Iyengar, S. (2022). A randomized experiment evaluating survey mode effects for video interviewing. *Political Science Research and Methods*, *11*(1), 144-159. https://doi.org/10.1017/psrm.2022.30

Fuchs, M. (2009). Gender-of-interviewer effects in a video-enhanced web survey: Results from a randomized field experiment. *Social Psychology, 40*(1), 37-42. https://doi.org/10.1027/1864-9335.40.1.37

Fuchs, M., & Funke, F. (2007). Video web survey - Results of an experimental comparison with a text-based web survey. In M. Trotman, T. Burrell, L. Gerrard, K. Anderton, G. Basi, M. Couper, K. Morris, K. Birks, A. Johnson, R. B., M. R., S. T., & A. W. (Eds.), *Proceedings of the Fifth International Conference of the Association for Survey Computing: The Challenges of a Changing World* (pp. 63-80). https://1stdirectory.co.uk/_assets/files_comp/41ba0402-5c62-4d6e-9460-07f82f314249.pdf

Gerich, J. (2008). Real or virtual? Response behavior in video-enhanced self-administered computer interviews. *Field Methods, 20*(4), 356-376. https://doi.org/10.1177/1525822X08320057

Gray, W. D., & Fu, W. T. (2004). Soft constraints in interactive behavior: The case of ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. *Cognitive Science*, *28*(3), 359-382. https://doi.org/10.1207/s15516709cog2803_3

Guggenheim, L., Valentino, N. A., Hupp, A. L., DeBell, M., Hillygus, D. S., & Shaw, D. R. (2026). Understanding nonresponse in live video survey interviews: Evidence from two tests conducted as part of the American National Election Studies (ANES). *Methods, Data, Analyses, 2*0(1), 46–86. https://doi.org/10.25521/mda.641

Haan, M., Ongena, Y. P., Vannieuwenhuyze, J. T. A., & De Glopper, K. (2017). Response behavior in a video-web survey: A mode comparison study. *Journal of Survey Statistics and Methodology*, *5*(1), 48-69. https://doi.org/10.1093/jssam/smw023

Höhne, J., Conrad, F., Neuert, C. & Claassen, J. (2024). Exploring effects of life-like virtual interviewers on respondents' answers in a smartphone survey. Paper presented at the 79th Annual Conference of the American Association for Public Opinion Research, Atlanta, GA. https://aapor.confex.com/aapor/2024/meetingapp.cgi/Paper/2812

Krysan, M., & Couper, M. P. (2003). Race in the live and the virtual interview: Racial deference, social desirability, and activation effects in attitude surveys. *Social Psychology Quarterly, 66*(4), 364-383. https://doi.org/10.2307/1519835

Lind, L. H., Schober, M. F., Conrad, F. G., & Reichert, H. (2013). Why do survey respondents disclose more when computers ask the questions? *Public Opinion Quarterly, 77*(4), 888-935. https://doi.org/10.1093/poq/nft038

McGonagle, K. & Sastry, N. (2021). An experimental evaluation of an online interview scheduler: Effects on fieldwork outcomes, *Journal of Survey Statistics and Methodology, 9*(3), 412-428. https://doi.org/10.1093/jssam/smaa031

Mittereder, F., & West, B. T. (2022). A dynamic survival modeling approach to the prediction of web survey breakoff. *Journal of Survey Statistics and Methodology, 10*(4), 945-978. https://doi.org/10.1093/jssam/smab015

Olson, K., Smyth, J. D., & Wood, H. M. (2012) Does giving people their preferred survey mode actually increase survey participation rates? An experimental examination. *Public Opinion Quarterly, 76*(4), 611-635. https://doi.org/10.1093/poq/nfs024

Opon, S. O., Tenambergen, W. M., & Njoroge, K. M. (2020). The effect of patient reminders in reducing missed appointment in medical settings: a systematic review. *PAMJ One Health, 2,* Article 9. https://doi.org/10.11604/pamj-oh.2020.2.9.21839

Peytchev, A. (2009). Survey breakoff. *Public Opinion Quarterly, 73*(1), 74-97. https://doi.org/10.1093/poq/nfp014

Schober, M. F., Okon, S., Conrad, F. G., Hupp, A. L., Ong, A. R., & Larsen, K. M. (2023). Predictors of willingness to participate in survey interviews conducted by live video. *Technology, Mind, and Behavior, 4*(2), 215-229. https://doi.org/10.1037/tmb0000100

Spiegelman, M. & Conrad, F. (2025). Improving understanding of survey questions with multimodal clarification. *Methods, Data, Analyses, 19*(1), 137-165. https://doi.org/10.12758/mda.2025.09

Spitzer, S. (2020). Biases in health expectancies due to educational differences in survey participation of older Europeans: It's worth weighting for. *The European Journal of Health Economics, 21*(4), 573-605. https://doi.org/10.1007/s10198-019-01152-0

Strojil, A., & Cígler, H. (2025). Video-administered questionnaire: Psychometric properties and comparison with a text-based format. *Computers in Human Behavior Reports, 19*, 100714. https://doi.org/10.1016/j.chbr.2025.100714

Sun, H., Conrad, F. G., & Kreuter, F. (2021). The relationship between interviewer-respondent rapport and data quality. *Journal of Survey Statistics and Methodology, 9*(3), 429-448. https://doi.org/10.1093/jssam/smz043

West, B. T., Ong, A. R., Conrad, F. G., Schober, M. F., Larsen, K. M., & Hupp, A. L. (2022). Interviewer effects in live video and prerecorded video interviewing. *Journal of Survey Statistics and Methodology, 10*(2), 317-336. https://doi.org/10.1093/jssam/smab040

Zacharias, C., & Pinedo, M. (2014). Appointment scheduling with no-shows and overbooking. *Production and Operations Management, 23*(5), 788-801. https://doi.org/10.1111/poms.12065

methods, data, analyses | Vol. 20(1), 2026, pp. 46-86 DOI: 10.25521/mda.641

Understanding Nonresponse in Live Video Survey Interviews: Evidence from Two Tests Conducted as Part of the American National Election Studies (ANES)

Lauren Guggenheim[1], Nicholas A. Valentino[1], Andrew L. Hupp[1], Matthew DeBell[2], D. Sunshine Hillygus[3], Daron R. Shaw[4]

[1] *University of Michigan*
[2] *Stanford University*
[3] *Duke University*
[4] *University of Texas at Austin*

© The Author(s) 2026. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Abstract

Live video interviewing is an attractive mode for survey research because it has the potential to supplement or, someday, even replace in-person interviewing at a reduced cost and without sacrificing data quality. The reasons for this optimism are several: live video interviewing maintains some of the desirable elements of in-person interviews, such as visual cues and rapport from the interaction between respondents and interviewers, while potentially conferring additional advantages, such as access to respondents living in remote locations. However, previous evidence and experience suggest respondents are reluctant to be interviewed via video, and that reluctance is non-ignorable. This paper explores these recruitment and non-response challenges, informed by the experience of two tests conducted by the American National Election Studies (ANES). The first study is the ANES 2020 Time Series Study, where respondents were randomly assigned to three sequential mode conditions, one of which included live video interviewing as the initial mode. The second study is the ANES 2023 Video Interviewing Pretest, in which respondents were recruited to an in-person interview with a one-month follow-up conducted by video. Respondents to the 2020 ANES video mode were more educated relative to benchmarks and more likely to be Democrats than the rest of the sample. Both qualitative and quantitative data suggest video non-response is heavily related to discomfort with technology, lack of skills, and distrust or dislike of video, and lack of access to equipment. Finally, offering prepaid incentives resulted in more video completions than postpaid incentives.

Keywords: live video interviewing, survey recruitment, representative samples, incentives, unit nonresponse

Direct correspondence to
Lauren Guggenheim, Center for Political Studies, University of Michigan, Ann Arbor, MI, USA
E-mail: lguggen@umich.edu

While video communication has become increasingly common in domains such as videoconferencing and medical visits, its adoption has lagged in both academic and commercial opinion surveys. This is likely because it is a relatively new mode of communication for most members of the public. Survey researchers have long considered how the capabilities of live video might shape the survey interview, reflecting on its strengths and weaknesses relative to other interviewer-administered modes (e.g., Anderson, 2008; Jeannis et al., 2013). However, the mode was not widely studied before the COVID-19 pandemic made its use imperative. The pandemic both accelerated the adoption of video communication among the public and convinced some survey organizations to deploy live video alternatives for in-person interviews.

Live video interviewing involves the use of real-time, two-way communication between the interviewer and respondent, facilitated by online platforms such as Zoom or Webex. Live video interviews hold particular appeal as a substitute for in-person interviews, given the increasing cost and logistical challenge of in-person approaches. Often considered the gold standard for surveys, in-person interviews are now prohibitively expensive for most research (Schober, 2018). Live video interviewing could preserve many advantages of in-person interviews, including the direct interaction between the interviewer and respondent, while also reducing travel and logistical complexity. Despite this promise, however, we currently know little about the data quality of this mode in large-scale surveys, and we are only beginning to understand its practical challenges (Schober et al., 2020).

This paper focuses on the recruitment of respondents into live video interviews—which has been difficult in both small and large-scale study designs (e.g., Bearfield & Smith, 2021; Conrad et al., 2023)—using data from the American National Election Studies (ANES). The ANES incorporated live video interviewing into its 2020 election study to compensate for the loss of in-person interviews during the COVID-19 pandemic. At the time, sending interviewers door-to-door for traditional in-person interviews was impossible, so ANES eliminated these interviews and attempted to shift some to video. Respondents also completed the survey online and by phone. Although the proportion of completed video interviews was low, the data helped inform a follow-up test conducted in 2023. This second study assessed a procedure where respondents initially completed in-person interviews with follow-up interviews completed by live video after a one-month break.

In the paper, we will: (a) review the current literature and discuss how features of live video may affect recruiting respondents into video interviews, (b) discuss the design and implementation of the studies, (c) examine the representativeness of the video components of the studies and examine potential underlying reasons for unit nonresponse, (d) review the results of efforts designed to increase response rates, and (e) consider how video might best be incorporated into surveys in the future.

Background and Literature Review

Live video interviewing has often been compared with in-person interviews (Anderson, 2008; Endres et al., 2023) and web-based surveys (Conrad et al., 2023; Endres et al., 2023). This is because the technical features of video platforms overlap with—and even rely on—those modes. They therefore share some of their advantages and disadvantages, while also having strengths and weaknesses of their own. Shared advantages of in-person and video stem from the face-to-face nature of the interview experience, especially the inclusion of an interviewer. Video interviewers seem to develop similar levels of rapport with respondents relative to in-person modes (Sun et al., 2021). This may boost data quality compared to non-interviewer modes, although interviewer effects are possible (West et al., 2022). In lab studies comparing in-person and video (Sun et al., 2021), web and video (Conrad et al., 2023), or in-person, web, and video (Endres et al., 2023), data quality via video appears similar to other interviewer-administered modes. For example, relative to web, the advantages of video include less satisficing—that is, more response differentiation and reduced straightlining—(Conrad et al., 2023; Endres et al., 2023), longer, more effortful responses to open-ended questions (Endres et al., 2023), and fewer breakoffs (Hupp et al., 2026). Overall, to date, the quality of data from live video interviews is high and similar to that of in-person interviews.

Live video interviewing also shares advantages with web-based surveys. Because it eliminates interviewer travel, live video could extend survey access to respondents in remote locations who could be missed by in-person designs (Hupp et al., 2026). On the downside, video interviews generally require scheduled appointments or on-demand access to interviewers, restricting some of the respondents' ability to be flexible with timing. However, live video allows participation from various locations and devices. Nevertheless, just as with in person interviews, some benefits of web surveys—quick turnaround times, large samples (Couper, 2000), less rounding of numerical answers (Conrad et al., 2023; Liu & Wang, 2015), or fewer socially desirable responses (Conrad et al., 2023; Endres et al., 2023; Heerwegh, 2007; Kreuter et al., 2008)—may be missing from live video.

One potential drawback of live video is the possibility of exacerbating unit nonresponse compared to online modes. Although most American adults use the internet, some demographic groups, such as older people and those with lower income and less education, are less likely to use or have access to the web (Pew Research Center, 2024). Video interviews obviously require a high-speed internet connection but also the necessary equipment, including a camera, microphone, and speakers (Schober et al., 2020). User abilities and skills also vary meaningfully, especially among older or less educated respondents. Even those with access to video technology may not feel confident setting it up, using it, or troubleshooting problems (Pew Research Center, 2021). The Pew Research Center (2021) found that while video communication was widespread during the pandemic, 19% of adults had not used video, and this gap was especially pronounced among those over 65 and with a high school education or less. Thus, both skills and access issues could explain why some groups rarely participate in video interviews.

Technological features of video may have a distinct impact on nonresponse relative to in-person interviews. In video interactions, participants view each other on screen. Differences in screen dimensions and settings cause face size to vary, and respondents may sit closer to the screen than to an interviewer in an in-person interview. The field of view is confined, and the presence of the camera may remind users that they are being watched (Shockley et al., 2021). Likewise, seeing one's own image via self-view could trigger self-focusing, an effect similar to the presence of a mirror (Bailenson, 2021; Fauville et al., 2023; but see Feuer & Schober, 2026). During the pandemic, both researchers (Bailenson, 2021) and the media (e.g., Fosslien & Duffy, 2020; Sklar, 2020) discussed the concept of "videoconference fatigue," with at least one study documenting its existence (Shockley et al., 2021). Research also found that individuals' attitudes toward live video in general are predictive of their willingness to participate in video interviews (Schober et al., 2023). For some, video could feel uncomfortable or require additional effort compared to an in-person interaction, and past exposure to video could shape their view of the technology, reducing willingness to participate in a video interview.

Distrust may also predict willingness to respond to video interviews. Although trust is important in survey response across modes (Dillman et al., 2014), security and privacy concerns specific to video—such as camera use and ease of recording—may change expectations of the privacy of the interview. Perceptions about online safety and privacy in general could also influence nonresponse; as more people are exposed to identity theft (Federal Trade Commission, 2024), computer viruses, and security breaches, individuals may be growing more skeptical of survey requests (Dillman, 2020). Such negative experiences are linked to distrust (Blank & Lutz, 2016; Dutton & Shepherd, 2006), although the relationship between online trust and internet use is less straightforward (e.g., Barth & de Jong, 2017; Blank & Lutz, 2016). Nevertheless, privacy and security concerns have sometimes been associated with engaging in privacy-protective behaviors (e.g., Baruh et al., 2017), such as limiting online use (Pew Research Center, 2000a, 2000b) or avoiding purchases and financial transactions (National Telecommunications and Information Administration, 2016), which could extend to the use of video communication.

Additional recruitment and unit nonresponse challenges may stem from the lack of an in-person interviewer to overcome respondents' reluctance to participate. Interviewers can play an important role in recruitment because they are able to address respondent concerns (Groves & Couper, 1996; Groves et al., 2004; West & Blom, 2016). More generally, the presence of the interviewer is thought to increase response rates (de Leeuw, 1992). Although recruitment methods for video can vary, respondents must be recruited by a mode other than video at present[1] since there are no video sampling frames. Given these challenges, it is not surprising to find low response rates for live video. For example, Conrad and colleagues (Conrad et al., 2023) found that address-based sampling achieved such low response rates—less than 1%—that they switched to opt-in sources. In general, prior experience suggests that recruitment to video may be easier when respondents are recruited via interviewer-led methods (Hanson et al., 2026) or are recruited as part of a longitudinal design (Wood et al., 2020).

Low response rates are a challenge to high-quality surveys because recruitment into a study can be costly and because low response rates (along with the systematic nonresponse to surveys among some groups) could lead to nonresponse bias. There has long been concern about declining response rates (de Leeuw & de Heer, 2002; Luiten et al., 2020), which necessitate greater fieldwork effort (Williams & Brick, 2018). Increased incentives, contacts with respondents, reminders, and other methods of increasing response rates are labor-intensive and costly. If respondents are reluctant to participate in video interviews, recruitment costs could make video prohibitively expensive or erode savings gained by eliminating travel for the interviewers. Also, although low response rates on their own do not necessarily increase nonresponse bias (e.g., Curtin et al., 2000; Groves, 2006; Groves & Peytcheva, 2008; Keeter et al., 2000), survey nonresponse can lead to bias when it is associated with the outcomes of interest (Groves, 1989). Perhaps more relevantly, differential nonresponse due to mode is a potential source of nonresponse bias (Groves & Peytcheva, 2008). If nonrespondents to video interviews differ from respondents in systematic ways related to the survey questions and topic, bias would surely grow. Nonresponse issues for video might be exacerbated for panel studies, where participants might drop out from subsequent waves because they cannot be contacted, they are unable to participate, or they refuse. Similar to nonresponse at the outset, panel attrition can lower sample size and potentially contribute to bias. Research is slim on attrition in video, but one study found that those who remained in video mode when given the option were younger, more educated, and married compared to those who opted to switch away from video in a subsequent wave of a panel study (Arrue et al., 2023).

---

1 We acknowledge that it may be possible to build a frame of iPhone mobile numbers and use Facetime to complete the video interview.

Research Questions

Based on the state of the literature reviewed above, we were interested in the following five research questions:

RQ1: *How do response rates to video compare to other modes?*

RQ2: *How do respondent demographic characteristics differ in video from population benchmarks, and how do they compare to web and in-person interviews?*

RQ3: *Which respondents are most likely to drop out of panel studies using video?*

RQ4: *What sorts of strategies might help decrease nonresponse to video?*

RQ5: *What factors are associated with nonresponse for video interview requests?*

Design and Implementation of the Studies

The ANES 2020 Time Series Study

The 2020 ANES is a two-wave panel consisting of pre- and post-election interview waves conducted among U.S. citizens ages 18 and over (American National Election Studies, 2022). The main study consisted of a freshly drawn cross-section from an address-based sampling (ABS) frame in all 50 states and D.C.[2] Households received invitation letters by mail and were invited to complete a survey online. The survey began with a screener that randomly selected a household member and assigned them to one of three conditions: web-only, web-mixed, and video-mixed, as shown in Figure 1 below. In the web-only condition, respondents completed the pre- and post-election questionnaires online. The remaining two groups followed a sequential mode ordering; in the web-mixed condition, respondents were asked to complete a questionnaire online, with nonrespondents subsequently pushed to phone. In the video-mixed condition, respondents received a request to complete the survey via live video interview, and nonrespondents were pushed first to the web and finally to the phone if they did not respond. All respondents completed the post-election interview in the same mode as their pre-election interview. Pre-election interviews took place between August 18 and November 3, 2020, and the post-election survey was conducted between November 8 and January 4, 2021. Median administration times were 71 minutes for the pre-election and 78 minutes for the post-election survey.

[2] Two additional sample groups in the study, which were administered at the same time as the cross-sectional sample, consisted of reinterviews with 2016 ANES respondents to create a panel and interviews with General Social Survey (GSS) respondents. These two groups are excluded from the discussion since the focus of this paper is on the fresh cross-sectional sample recruited in 2020.

Incentives were used in both pre- and post-election surveys, ranging from $40 to $200 with non-response escalations.[3] Figure 1 shows the ANES 2020 design, along with the mode information.

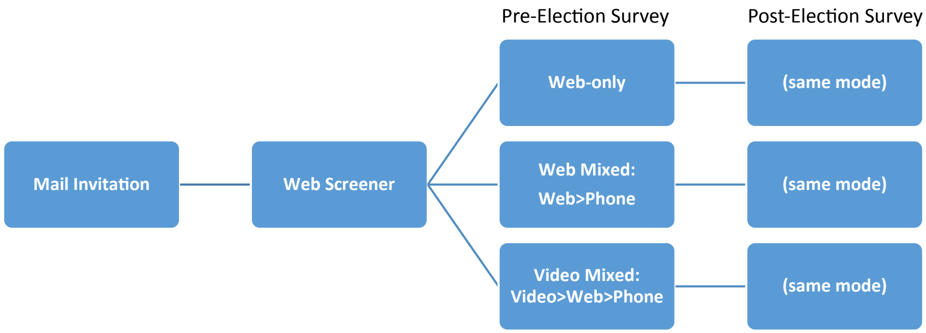


Figure 1 ANES 2020 Sample Design with Mode Information

To administer the video interviews and recruit respondents using this mode, the study relied primarily on inbound contacts, in a method combining mostly on-demand interviews with some appointment setting (DeBell et al., 2022). After completing the screener, the selected respondent was redirected to a virtual waiting room (or received a follow-up email directing them there if they were not currently available) that provided information about the video interview process, the equipment needed, the Zoom video software, and, most importantly, a Zoom link if an interviewer was available. The waiting room also described current interviewer availability in terms of expected wait time (if any), allowed respondents to set appointments, or gave respondents the opportunity to speak with someone on a live help desk by chat, phone, or email. More information on the interview setup is in Appendix A.

Respondents who scheduled an appointment were emailed a confirmation of the appointment date/time and a reminder on the morning of the interview. During the field period, it became clear that more reminders were necessary to reduce the rate of missed appointments, which was 65%. A new protocol was established that added a text reminder 30 minutes before the appointment, five minutes after the scheduled time if the respondent had not connected by then, and a missed appointment email 30-60 minutes after the appointment time if the respondent failed to connect, respond, or reschedule with any of the previous contacts.

[3] More information about the study design and administration can be found in the Methodology Report for the ANES 2020 Time Series Study (DeBell et al., 2022) or the User Guide (ANES, 2022).

The ANES 2023 Video Interviewing Pretest

The 2023 Pretest is a two-wave panel study where respondents completed interviews in person followed by a live video interview four weeks later. Figure 2 shows the design of the pretest, where households first received recruitment letters by mail, followed by an in-person interview visit. Those completing the first interview in person were asked if they wanted to participate in a second interview, one month later, via video. Respondents who agreed provided their email and phone number and scheduled an appointment. For those who scheduled an appointment and were able to be reached, a follow-up interview took place over Zoom. At the end of the interview, interviewers copied a link into the chat and asked respondents to fill out a few final questions about the interview. The reminder protocol included email or text messages around 24 hours before the appointment and on the day of the appointment.

Figure 2 ANES 2023 Pretest Design with Mode Information

The pretest used a three-stage sample design. First, nine geographical areas were strategically identified in different parts of the country near the location of twelve experienced interviewers and covered locations that were rural, suburban, and urban. In five areas, the Republican candidate won a majority or plurality of votes in the 2020 presidential election, while the Democratic candidate won in the other four. In the second stage, a probability sample of addresses was selected from an address-based sampling frame. In the final stage, a respondent within the household was interviewed based on their availability and willingness to participate. Data collection took place between September 7 and November 27, 2023.

Two procedures were employed to encourage respondents to set appointments and complete the video interview: a prepaid-postpaid incentive experiment and an offer to run a connection test with respondents—that is, an offer to show respondents during the in-person interview how to connect to Zoom for the follow-up interview. Studies show prepaid incentives tend to be more effective than postpaid incentives at increasing response rates (Singer et al., 1999), but it is not known whether this extends to video. Respondents to the in-person wave were randomly assigned to receive either a prepaid or postpaid incentive, both for $25. At the end of the in-person wave, respondents were told that an interviewer wanted to complete a video interview with them in a few weeks. They were either told that if they agreed to participate, they would receive the cash right now (i.e., prepaid) or would receive the cash after completing the interview (i.e., postpaid). The connection test was also offered at that point. Then the respondent either agreed or declined the video invitation. Interviewers recorded the incentive, whether (and how successfully) the respondent participated in the connection test, and listed an appointment time if the respondent made one. In 12 cases, there was a discrepancy between the assigned experimental group and the interviewer record of the incentive in the questionnaire. We used actual payment records to resolve the discrepancies for those cases, except for one case that could not be resolved and was dropped. Exact wording for the experiment can be found in Appendix B.

In both the 2020 study and 2023 pretest, non-cooperative cases were asked by interviewers why they did not want to participate via video. In the ANES 2020, non-cooperative cases were asked during an intervention that was designed to increase response rates, while in the pretest, respondents were asked as a follow-up immediately after they declined the request to participate by video.

Methods

RQ1 focuses on response rates, and is answered by calculating the response rates from ANES 2020 using AAPOR RR1 (American Association of Public Opinion Research, 2023) and the reinterview rates, including the reinterview rate for the 2023 ANES Video Pretest. Differences between response rates were calculated using chi-squared tests. For RQ2, ANES responses were recoded for ease of interpretation or to match the Current Population Survey's (CPS) November 2020 Supplement, and then video respondents were compared to respondents in other modes and benchmarks using chi-squared tests and differences in percentage points. RQ3, about panel attrition, is answered using the same method, except comparing the respondent distributions at two time points for both studies. Strategies to help decrease nonresponse (RQ4) are investigated by looking at missed appointment rates, analyzing the results of an incentive experiment using chi-squared tests, assessing the results of a connection test, and examining the frequencies of respondent concerns with video. Finally, to examine differential nonresponse (RQ5) a logistic regression was estimated predicting video completion relative to completion through other modes in the video-mixed group from the 2020 study, and the differences in respondent attitudes towards video and technology between those who responded in person (Wave 1) and by video (Wave 2) were examined using chi-squared tests. All analyses were completed using Stata MP version 18.0.

Results

Response and Reinterview Rates (RQ1)

Table 1 below shows the response and reinterview rates from the 2020 ANES by each of the three sequential mixed-mode groups to which respondents were assigned. Response and reinterview rates from the ANES 2016 Time Series Study are also shown, broken out by mode, to serve as a comparison. For the full 2020 study, the response rate (AAPOR RR1) was 36.7%, while the video-mixed condition had a 27.6% response rate, the lowest of all groups (it also had the lowest re-interview rate).[4] All groups in the 2020 study had a slightly lower response rate than in 2016, which is unsurprising as the general trend is toward declining response rates. However, the response rate for the 2016 in-person mode appears to be quite different from the video-mixed group, with an RR1 of 50.2% compared with an RR1 for the video-mixed of 27.6%. Table 2 shows the contribution of interviews in each mode toward the response rate of the video-mixed condition in Table 1. Table 2 shows that only about half (i.e., 47.8%) of the respondents recruited into the video-mixed group completed the study using video interviewing. The rest were pushed to the web or phone. Moreover, despite the similar reinterview rates across groups in Table 1, Table 2 shows that, in fact, the reinterview rate is also lower for the video mode (76.3%) than the web (93.6%) in the mixed video condition[5] and lower than the web-only (88.1%) and web-mixed groups (88.7%) in Table 1.

Turning to the 2023 Pretest, there were 157 in-person completions from Wave 1 and 48 video completions from Wave 2, a reinterview rate of 30.6%. It is important to note that follow-up efforts were not comparable to standard ANES reinterviews that achieve much higher reinterview rates. This is in contrast to the 2020 study's mode experiment, where video-mixed efforts were comparable to web-only efforts. A total of 35 respondents (72.9% of the video respondents) followed the link the interviewers placed in the chat, completing the online follow-up questionnaire.

4 The response rates between the conditions are significantly different, $\chi^2(2, N = 14{,}827) = 124.83$, $p < .001$, and a series of three pairwise comparisons with Bonferroni adjustments indicate that the video-mixed response rate is significantly lower than the web-only response rate ($\chi^2(1, N = 8{,}771) = 86.09$, $p < .001$) and the web-mixed response rate ($\chi^2(1, N = 8{,}780) = 120.95$, $p < .001$). The web-only and web-mixed groups are not significantly different from each other.

5 Within the video-mixed group, the reinterview rates by mode are significantly different, $\chi^2(2, N = 751) = 39.16$, $p < .001$. This is driven by web, which is significantly higher than the other two modes. A series of three pairwise comparisons with Bonferroni adjustments indicate that the web reinterview rate was significantly higher than video ($\chi^2(1, N = 687) = 39.20$, $p < .001$) and phone ($\chi^2(1, N = 392) = 6.25$, $p = .012$), but the video to phone comparison was not significant.

Table 1 Response and Reinterview Rates in ANES 2020 and 2016, Cross-Sectional Samples

	Pre-Election		Post-Election	
	N	RR1 (%)	*N*	Reinterview (%)
2020 Fresh Cross-Sectional Sample	5,441	36.7	4,783	87.9
Web-only (Web)	2,283	37.8	2,012	88.1
Web-mixed (Web>Phone)	2,407	39.7	2,136	88.7
Video-mixed (Video>Web>Phone)	751	27.6	635	84.6
2016 Web	3,090	44.4	2,590	83.8
2016 Face-to-Face	1,180	50.2	1,058	89.7

Notes: Response rates are based on AAPOR Response Rate 1. Reinterview rates are the percentage of respondents from the pre-election that completed the post-election study. Response rates are republished from the ANES Users Guides (ANES 2019, ANES 2022).

Table 2 Contribution of Each Mode to Video-Mixed Condition and Reinterview Rate for Each Mode in Video-Mixed (Video>Web>Phone) Condition, ANES 2020

	Pre-Election Contribution		Post-Election Reinterview Rate	
	N	%	*N*	%
Video	359	47.8	274	76.3
Web	328	43.7	307	93.6
Phone	64	8.5	54	84.4
Total	751	100.0	635	84.6

Both studies' response and reinterview rates indicate that recruiting respondents into video interviews and maintaining their interest for further completions presents a challenge. Next, we examine the characteristics of the video cases relative to other modes and benchmarks.

Representativeness of Video by Demographic Subgroups in ANES 2020 (RQ2)

We compared demographic characteristics of the 2020 ANES video respondents with the population and to other modes using three sources: (a) respondents to the 2020 ANES who responded by a mode other than video, which combine all web and phone cases across the three study groups,[6] (b) the U.S. adult citizen

6 Phone cases only make up 1.7% of the freshly sampled cross-sectional cases.

population as measured by the Current Population Survey's (CPS) November 2020 Supplement (Flood et al., 2023); and (c) the in-person cases in the 2016 ANES (American National Election Studies, 2019). Unweighted results are provided for the ANES samples, whereas CPS data are weighted to adjust to the population. The unweighted ANES results show the characteristics of the respondents who were recruited into video after contact attempts and incentives (more on this shortly). Note that for most purposes, researchers would use the weighted results, which attempt to minimize the difference between the sample and the population (DeBell et al., 2022). We use unweighted results because weights would adjust respondent characteristics to the population, thereby masking the differences between respondent demographics by mode.

Table 3 shows the demographic characteristics of the samples described above, focusing on age, education, income, sex, race/ethnicity, party identification, home tenure, and urbanicity. The first column shows the CPS benchmarks, which are limited to adult citizens aged 18 and over for a more direct comparison to the ANES. The second column shows the unweighted in-person respondents to the 2016 ANES survey, which serves as a comparison despite the different context and year. The third column shows the unweighted non-video (i.e., web and phone respondents across the groups) modes in the 2020 ANES. The distribution for the unweighted video cases is shown in the fourth column. The differences between the benchmarks and the unweighted distributions for the ANES 2020 video mode are shown in the fifth column. These differences give us an indication of the representativeness of the video mode. Following Chang and Krosnick (2009), column 5 also shows the mean of the absolute values of the differences for each demographic distribution. The largest of these mean deviations from the benchmarks is education ($M = 8.3$ percentage points). Relative to the population benchmark, video respondents were far more educated. This is especially notable for those in the highest education category; among those who responded via video, 25.8% had a graduate degree compared with only 12.4% of Americans (as measured by the CPS benchmark); this is a 13.4 percentage point difference. Though other differences were less striking, the video mode respondents were higher in income, more white, and less Hispanic.

To better understand how the characteristics of respondents in the video mode might differ from other modes, the demographics of the video respondents can also be compared to the rest of the cross-sectional sample in 2020 in Table 3. Columns three and four show the unweighted distributions of the video respondents (column 4) compared to the web and phone modes (column 3). These comparisons show significant differences between the mode groups for education, $\chi^2(4, N = 5{,}343) = 20.84$, $p < .001$; income, $\chi^2(3, N = 5{,}230) = 17.72$, $p = .001$; and party identification ($\chi^2(6, N = 5{,}407) = 51.76$, $p < .001$). Put simply, relative to the rest of the cross-sectional sample in 2020, the video respondents were more educated, had a higher income, and were more likely to be Democrats.

Table 3 Comparisons of Respondent Characteristics Between CPS Benchmarks and ANES Modes (2016 In-person Mode; 2020 Web and Phone Modes; and 2020 Video Mode) (Unweighted)

Characteristic	CPS Nov 2020 benchmarks	ANES 2016 In-Person	ANES 2020 Web and Phone Modes (*N* = 5082)	ANES 2020 Video Mode (*N* = 359)	ANES 2020 Video Difference (CPS-ANES Video)
	Percent	Percent	Percent	Percent	Percent
Age					
18-29	20.3	16.7	14.2	14.6	5.7
30-39	16.5	17.3	17.5	17.7	−1.2
40-49	15.0	15.0	14.8	14.9	0.1
50-59	16.4	18.4	16.5	15.7	0.7
60-69	16.2	17.0	19.2	20.8	−4.6
70 or older	15.6	15.7	17.7	16.3	−0.7
Mean deviation					2.2
χ^2			1.004		
Education					
Less than HS cred.	7.6	9.0	4.8	2.2	5.4
HS credential	29.2	22.0	16.6	13.8	15.4
Some college/ AA degree	28.5	33.1	35.0	28.9	−0.4
Bachelor‘s degree	22.4	22.6	24.7	29.2	−6.8
Graduate degree	12.4	13.3	18.9	25.8	−13.4
Mean deviation					8.3
χ^2			20.843***		
Income					
Under $25,000	15.0	26.0	21.2	16.7	−1.7
$25,000-$49,999	21.6	23.3	18.9	12.7	8.9
$50,000-$99,999	32.1	30.4	28.2	35.9	−3.8
$100,000 or more	31.3	20.3	31.7	34.7	−3.4
Mean deviation					4.5
χ^2			17.716***		
Sex					
Male	48.1	47.5	45.7	44.0	4.1
Female	51.9	52.5	54.3	56.0	−4.1
Mean deviation					4.1
χ^2			0.376		

Table 3 (continued)

Characteristic	CPS Nov 2020 benchmarks	ANES 2016 In-Person	ANES 2020 Web and Phone Modes (N = 5082)	ANES 2020 Video Mode (N = 359)	ANES 2020 Video Difference (CPS-ANES Video)
	Percent	Percent	Percent	Percent	Percent
Race/Ethnicity					
White, non-Hispanic	66.9	67.7	72.6	75.9	−9.0
Black, non-Hispanic	12.4	10.1	9.0	10.2	2.2
Hispanic	13.2	14.4	9.0	7.6	5.6
Other, non-Hispanic	7.5	7.8	9.4	6.2	1.3
Mean deviation					4.5
χ^2			5.269		
Party ID 7-point					
Strong Democrat	–	18.9	22.8	29.7	–
Not strong Democrat	–	11.8	11.0	12.6	–
Lean Democrat	–	15.0	12.1	17.6	–
Independent	–	10.3	12.7	3.1	–
Lean Republican	–	16.6	10.4	13.7	–
Not strong Republican	–	11.3	10.1	8.7	–
Strong Republican	–	16.0	20.8	14.6	–
Mean Deviation					–
χ^2			51.764***		
Home Tenure[a]					
Own	73.8	59.8	71.9	70.0	3.8
Rent/Other	26.2	40.2	28.1	30.0	−3.8
Mean deviation					3.8
χ^2			0.279		
Urbanicity[b]					
Rural	–	–	15.2	11.8	–
Small town	–	–	25.1	21.0	–
Suburb	–	–	30.3	31.0	–
City	–	–	29.4	36.2	–
Mean deviation		–			–
χ^2			7.804#		

Notes: The Pearson's χ^2 and significance levels in the table are for the comparison of the video respondents (column 4) compared to the web and phone modes (column 3). ANES data are unweighted. #$p < .10$; *$p < .05$; **$p < .01$; ***$p < .001$

Table 3 (continued)

[a] To compare to CPS benchmarks for home tenure, the ANES used a split half design where only half the sample was asked the CPS version of the question and the other half was asked an older ANES version of the home tenure item. In this table, the analysis is based on the CPS version of the question.
[b] Urbanicity is based on respondent self-report and not on geographical data.

Attrition Across Waves (RQ3)

Attrition Between the Pre- and Post-Election Survey in ANES 2020

To examine the nature of cases lost between pre- and post-election surveys, we next measure attrition in the video mode. Respondents who were initially interviewed by video were all invited to complete the post-election interview by video after the election. Recall from the reinterview rates in Table 2 that the video-mixed condition in 2020 had the largest attrition (15.4% of respondents were not reinterviewed in this condition), and within that condition, the video mode had the highest attrition (i.e., 23.7% of video-mode respondents were not reinterviewed). Table 4 compares respondents who completed both the pre- and post-election survey in the video mode ($N = 274$) to those who only completed the pre-election study by video (i.e., the attrition cases; $N = 85$).

There are significant differences in age ($\chi^2(5, N = 356) = 19.33$, $p = .002$) and education $\chi^2(4, N = 356) = 16.88$, $p = .002$) between those who remained in the study and those who did not in 2020. Those who completed both the pre and post tended to be older and more highly educated than those who dropped out. For example, respondents age 30-39 were 11.5 points more likely to drop out than to complete the post-election survey (26.5% to 15.0%). That younger respondents tend to attrite from the video mode is in keeping with prior research. Likewise, attrition from the video mode among the less educated is consistent with the literature. This attrition may also be related to issues with technology access and skills among the less educated. By contrast, the distributions of party identification among pre-only and those who completed both pre- and post-election survey are not significantly different. Readers who wish to compare characteristics between those who attrit from the video mode with other modes will find a similar table for attrition from the web in the Appendix C (Table C1).

ANES 2023 Pretest Video Attrition

In Table 5 we examine how respondents who participated in the video component of the 2023 Pretest during Wave 2 compared to those who only engaged in the face-to-face portion in Wave 1. Respondents who participated in the video follow-up may be demographically similar to the 2020 ANES video respondents because video is more appealing or accessible to certain respondents. However, nonresponse to video in Wave 2 is likely not solely due to the request to partici-

Table 4 Attrition among Video Cases in the 2020 ANES (Unweighted)

Characteristic	Completed Pre Only (Attrition Cases) (*N* = 85)	Completed Pre and Post (Reinterview Cases) (*N* = 274)	Difference (Attrition-Reinterview)
	Percent	Percent	Percent
Age			
18-29	22.9	12.1	10.8
30-39	26.5	15.0	11.5
40-49	18.1	13.9	4.2
50-59	10.8	17.2	−6.4
60-69	10.8	23.8	−13.0
70 or older	10.8	17.9	−7.1
Mean deviation			8.8
χ^2	19.333**		
Education			
Less than HS cred.	5.9	1.1	4.8
HS credential	20.0	11.8	8.2
Some college/AA degree	35.3	26.9	8.4
Bachelor‘s degree	22.4	31.4	−9.0
Graduate degree	16.5	28.8	−12.3
Mean deviation			8.5
χ^2	16.877**		
Income			
Under $25,000	24.1	14.4	9.7
$25,000-$49,999	8.4	14.0	−5.6
$50,000-$99,999	33.7	36.5	−2.8
$100,000 or more	33.7	35.1	−1.4
Mean deviation			4.9
χ^2	5.322		
Sex			
Male	42.4	44.5	−2.1
Female	57.6	55.5	2.1
Mean deviation			2.1
χ^2	0.124		
Race/Ethnicity			
White, non-Hispanic	78.6	75.1	3.5
Black, non-Hispanic	9.5	10.4	−0.9
Hispanic	9.5	7.1	2.4

Table 4 (continued)

Characteristic	Completed Pre Only (Attrition Cases) (N = 85)	Completed Pre and Post (Reinterview Cases) (N = 274)	Difference (Attrition-Reinterview)
	Percent	Percent	Percent
Other, non-Hispanic	2.4	7.4	−5.0
Mean deviation			3.0
χ^2	3.281		
Party ID 7-point			
Strong Democrat	25.3	31.0	−5.7
Not strong Democrat	14.5	12.0	2.5
Lean Democrat	20.5	16.8	3.7
Independent	1.2	3.6	−2.4
Lean Republican	15.7	13.1	2.6
Not strong Republican	6.0	9.5	−3.5
Strong Republican	16.9	13.9	3.0
Mean deviation			3.3
χ^2	4.295		
Home Tenure[a]			
Own	55.3	65.4	−10.1
Rent/Other	44.7	34.6	10.1
Mean deviation			10.1
χ^2	2.861[#]		

Notes: The Pearson's χ^2 values and significance levels in the table are for the comparison of respondents who completed both the pre- and post-election survey in the video mode to those who only completed the pre-election survey. [#]$p < .10$; *$p < .05$; **$p < .01$; ***$p < .001$
[a] Home tenure combined two versions of a question that were assigned as split half: the CPS and ANES versions.

pate by video but also to other causes of attrition, such as respondent experience in the earlier wave (Frankel & Hillygus, 2014). It is also important to note that pretest cases are not worked as hard as they would be for a full study. Lastly, because the pretest was not designed to be nationally representative, comparing it to the benchmarks will be less useful; rather, the data presented are meant to be suggestive of who might be persuaded to participate once they were familiar with the survey and the interviewer and were presented with the request to participate by video.

Table 5 shows the differences between those who only completed Wave 1 (in person) and those who completed both Waves 1 and 2 (video). The demographics of Wave 2 respondents, who completed the earlier wave in person and went on to complete the video interview in Wave 2, were younger, more Democratic, and had higher levels of education and income than those who only completed Wave 1. The only difference approaching significance is education $\chi^2(4, N = 149) = 9.05$, $p = .06$. However, the small sample size makes detecting potentially significant differences difficult. The differences appearing in these data, though, suggest that video is attracting the same sorts of respondents as in the ANES 2020 study. What is potentially notable (although not significant) is that the respondents to the video follow-up tended to be younger, i.e., 30-39, which is surprising given the expected attrition among younger respondents. This finding is similar to one by Arrue et al. (2023), who found younger respondents tended to remain in video when offered a mode switch.

Table 5 ANES 2023 Video Pretest Wave 1 (In-person) and Wave 2 (Video) Comparison of Demographic Characteristics (Unweighted)

Characteristic	Total ($N = 157$)	In-Person Only ($N = 109$)	Video ($N = 48$)	Difference (In-Person -Video)
	Percent	Percent	Percent	Percent
Age				
18-29	15.4	14.8	16.7	−1.9
30-39	15.4	13.0	20.8	−7.9
40-49	15.4	13.9	18.8	−4.9
50-59	10.3	11.1	8.3	2.8
60-69	17.9	20.4	12.5	7.9
70 or older	25.6	26.9	22.9	3.9
χ^2		3.520		
Education				
Less than HS cred.	13.4	16.2	6.8	9.4
HS credential	26.2	28.6	20.5	8.1
Some college/AA degree	24.8	24.8	25.0	−0.2
Bachelor‘s degree	22.1	21.9	22.7	−0.8
Graduate degree	13.4	8.6	25.0	−16.4
χ^2		$9.054^{\#}$		
Income				
Under $25,000	7.6	10.1	2.6	7.6
$25,000-$49,999	21.2	22.8	17.9	4.8
$50,000-$99,999	33.1	31.6	35.9	−4.3
$100,000 or more	38.1	35.4	43.6	−8.1
χ^2		2.843		

Table 5 (continued)

Characteristic	Total (N = 157)	In-Person Only (N = 109)	Video (N = 48)	Difference (In-Person -Video)
	Percent	Percent	Percent	Percent
Home tenure				
Own	68.8	71.6	62.5	9.1
Rent/Other	31.2	28.4	37.5	−9.1
χ^2		1.274		
Sex				
Male	45.2	44.0	47.9	−3.9
Female	54.8	56.0	52.1	3.9
χ^2		0.203		
Race/Ethnicity				
White, non-Hispanic	73.4	72.6	75.0	−2.4
Black, non-Hispanic	9.1	10.4	6.3	4.1
Hispanic	13.0	12.3	14.6	−2.3
Other, non-Hispanic	4.5	4.7	4.2	0.6
χ^2		0.803		
Party ID 7-point				
Strong Democrat	19.6	21.8	14.9	6.9
Not strong Democrat	10.1	5.9	19.1	−13.2
Lean Democrat	8.8	7.9	10.6	−2.7
Independent	6.8	7.9	4.3	3.7
Lean Republican	20.9	18.8	25.5	−6.7
Not strong Republican	12.2	11.9	12.8	−0.9
Strong Republican	21.6	25.7	12.8	13.0
χ^2		10.415		

Notes: The Pearson's χ^2 values and significance levels in the table are for the comparison of the in-person and video cases. Data are unweighted. $^{\#}p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Understanding and Attempting to Reduce Unit Nonresponse to Video (RQ4)

Appointment Setting and Reminders

One way to encourage respondents who were offered video to complete the ANES 2020 study by that mode was to set appointments and provide reminder(s). As described above, cases that set appointments for video in the pre-election survey initially received only an appointment confirmation and a single reminder, automated and sent by email. The single reminder was offered until mid-September. The missed appointment rate during that period was 65.4%. A change in protocol after that date provided cases with two additional text reminders and a missed appointment email, sent manually. Before the reminders, 34.6% of appointments were on time, and after the additional reminders, 43.6% were on time (DeBell et al., 2022). In the 2023 Pretest, those who scheduled an appointment for the Wave 2 interview were also reminded of their appointments. Depending on the information provided by the respondent, the interviewer sent an email, a text message, or both the day before the scheduled appointment and a few hours prior to the scheduled appointment. A total of 83 respondents set an appointment, and 48 completed the video interview.

Pretest Incentive Experiment and Connection Test

As described above, the 2023 Pretest included an incentive experiment at the end of the in-person wave. Table 6 shows the outcome of the incentive offered on appointment setting and completions. Among those who were offered the prepaid incentive, 57.7% set an appointment with the interviewer, compared with only 47.4% among those offered the postpaid incentive. Ultimately, 38.5% of respondents who were offered the prepaid incentive completed a video interview, compared to only 23.1% of postpaid respondents, a statistically significant 15.4 percentage point difference, $\chi^2(1, N = 156) = 4.33$, $p < .05$. However, the cost was more than four times as large for the prepaid incentive.[7] Results suggest that the prepaid incentive could encourage appointment setting. More importantly, the prepaid incentive appears to significantly increase video completions.

Table 6 ANES 2023 Pretest: Video Appointment Setting and Completions by Incentive Type

Incentive offered	Set Appointment		Video Completion		Total
	No %	Yes %	No %	Yes %	%
Offered prepaid (N = 78)	42.3	57.7	61.5	38.5	100.0
Offered postpaid (N = 78)	52.6	47.4	76.9	23.1	100.0
χ^2	1.645		4.333*		

Notes: $^{\#}p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

7 This number is calculated based on the incentive amount and number of respondents receiving the incentive. All 78 respondents assigned to the prepaid group were offered the $25 incentive, but only 18 in the postpaid group completed the study and received the $25 incentive.

All respondents who reached the end of the in-person interview were offered the connection test. Connection test results are reported in Table 7. Only 7.8% of respondents successfully completed the test, while an additional 3.2% tried unsuccessfully, and the rest (88.9%) declined. In addition, 30.5% of respondents declined to try because they said they already knew Zoom, and 58.4% of respondents declined without explanation. To understand whether experiences with technology influenced respondents' decisions to participate in the test, we looked at whether prior video call experience or comfort using a smartphone was associated with completing the test. We found that among those who were not at all comfortable with smartphones, 100% were not willing to try the connection test, compared with 89.1% of those who were either slightly or moderately comfortable with smartphones and 86.4% of those who were very comfortable with smartphones. Among those who had never participated in a video call, 90.9% did not try to connect, compared with 88.6% of respondents who had participated in calling. These results suggest that this procedure designed to help overcome respondent misgivings was less successful among those who would benefit most from just such an intervention–those who are less comfortable with technology.

Nonresponse Video Intervention in ANES 2020

Approximately two weeks into the field period of the 2020 ANES, 16% of cases from the video-mixed group had completed the screener but opted not to continue. In the other cross-sectional groups, failure to continue was less than 3%. The underperformance of the video-mixed group relative to the other two groups led to the creation of an intervention designed to increase the response rate for video (ANES, 2021; DeBell et al., 2022). Two experienced interviewers placed outbound calls to cases who had been selected upon screener completion but had not started the interview. Although 422 cases qualified for a contact, interviewers were only able to complete calls with 67 potential respondents or members of their household during this intervention period. Table 8 shows the reasons these participants gave to the interviewers for their nonresponse.[8] Overall, the conversion team recorded more concerns about video privacy or image appearance than other problems. Respondents also gave a variety of reasons for not wanting to participate that had nothing to do with video or technology; it is possible that some respondents had reasons that they were unable or unwilling to articulate, although some likely declined for reasons other than mode.

Additionally, these two interviewers acted as a refusal conversion team. In addition to using standard techniques for gaining cooperation, the interviewers could offer: (a) use of a different video platform, such as Facetime, and (b) an

---

[8] This analysis used closed-ended questions from the call outcome questionnaire completed by the video conversion team interviewers. More information can be found in Appendix D.

increased incentive of $200 for completing the live video interview. The refusal team was allowed to provide one or both offers at their discretion, based on what they thought would gain cooperation. During the intervention, the team chose not to offer any nonrespondents the use of a different video platform but did offer 17 nonrespondents the incentive. Additionally, 29 nonrespondents reached during the intervention agreed to complete a video interview later. However, only three nonrespondents ultimately ended up completing a video interview, and none of those who were offered the increased incentive completed by video.

Table 7 ANES 2023 Pretest: Which respondents were more likely to complete the connection test?

Test Completion	Comfort with Smartphone			Ever participated in a video call		Total
	Not at all %	Slightly/ Moderately %	Very %	No %	Yes %	%
Yes, completed	0.0	3.6	12.3	9.1	7.6	7.8
Tried unsuccessfully	0.0	7.3	1.2	0.0	3.8	3.2
No, did not try	100.0	89.1	86.4	90.9	88.6	88.9
Total	100.0	100.0	100.0	100.0	100.0	100.0
χ^2		9.056$^{\#}$		0.903		

Notes: $^{\#}p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Table 8 ANES 2020: Issues Mentioned by Respondent (Completed Calls) in Nonresponse Intervention

Issue	*N*
Video Privacy or Image Concerns:	
Did not want their image appearing on video	11
Privacy/confidentiality of videoconferencing generally	6
Did not want their surroundings appearing on video	3
Concerned that they are being recorded	2
Subtotal	22
Lacks Equipment:	
Does not have a capable device (smartphone, tablet, computer, laptop) at all	7
Does not have a microphone	4
Does not have a video camera	3
Does not have speakers	3

Table 8 (continued)

Issue	*N*
Subtotal	17
Lacks Skills or is uncomfortable with Video/Technology:	
Uncomfortable with videoconferencing generally	7
Technical difficulties in installing the Zoom app	4
Uncomfortable with technology generally	3
Subtotal	14
Other	21
Total Number of Issues	74
Total Number of Completed Calls	67

Notes: Interviewers could record multiple issues for each respondent. An earlier version of this table appeared in DeBell et al. (2022) but has been revised and extended.

Response to Interviewers in 2023 ANES Pretest

Interviewers in the 2023 Pretest attempted to gain cooperation to participate in the video interviews at the end of Wave 1 in-person interview. Respondents who did not agree to participate in the video interview one month later were asked why they did not want to participate. Respondents could provide as many explanations as they wanted, and interviewers took notes on their answers. Responses were recorded for a total of 75 respondents. These were then coded for the presence or absence of a number of concerns related to non-response and then organized by category type into Table 9. More information about the coding can be found in the Appendix D. Although only 24 responses were about video, there were a number of responses about technology generally. Overall, the main reasons respondents did not want to complete the video interview could be categorized as concerns about privacy, lack of trust in technology, dislike of videoconferencing or technology, or lack of skills. Not many respondents lacked equipment. Many respondents gave general or idiosyncratic reasons for not participating. These were related to the usual concerns about completing surveys (e.g., “too busy”), or respondents may have had reasons for declining the interview that were related to video that they did not share with the interviewer.

Differential Nonresponse (RQ5)

Next, we look within the sequential video-mixed condition to see which respondents were more likely to respond to the video request and which were more likely to wait and respond in other modes. We examine whether respondent attitudes (i.e., social trust and political interest) and access to technology (i.e.,

Table 9 ANES 2023 Pretest: Issues mentioned by respondents

Issue	*N*
General lack of trust in technology	13
Video Privacy or Image Concerns:	
Did not want their image appearing on video	6
Did not want their surroundings appearing on video	1
Concerned that they are being recorded	3
Privacy/confidentiality of video or image	3
Subtotal	13
Dislikes technology or using technology:	
Doesn't like internet	4
Doesn't like videoconferencing	7
Generally doesn't like technology	4
Subtotal	15
Lacks Skills or uncomfortable with video/technology:	
Don't use videoconferencing	4
Not familiar/comfortable with video technology	5
Not familiar/comfortable with technology generally	2
Subtotal	11
Lacks equipment:	
Does not have a capable device (smartphone, computer)	1
Does not have internet	2
Subtotal	3
Other issues:	
Don't like surveys	3
General privacy concerns	4
Too busy	18
Don't want to	10
Other	14
Subtotal	49
Total number of concerns related specifically to video	24
Total concerns over all	104
Total number of respondents	75

Notes: Interviewers could record multiple issues for each respondent. Responses were recorded for respondents who declined a follow-up interview during the first in-person interview.

internet at home and access to the internet via phone) are associated with this outcome. Table 10 shows the results of a logistic regression predicting video completion relative to completion through other modes (i.e., web, phone) among those respondents assigned to the video-mixed group. These results underscore previous findings. Those with higher levels of education were more likely to participate by video, as were Democrats and Republicans relative to independents. Those with higher levels of trust and access to the internet on their phones were also more likely to participate by video.

Switching to the 2023 pretest results, Table 11 shows a comparison of the difference in respondent attitudes among those who responded in person (Wave 1) compared to the video (Wave 2) follow-up. First, the table shows that 67.1% of respondents dislike video surveys, with a lower proportion of respondents disliking paper and web. Similarly, Schober and colleagues (Schober et al., 2023) show that respondents in their study gave lower ratings for their willingness to participate in live video surveys relative to other modes, including in-person, phone, and web. There are several significant differences between those who participated in the in-person wave and those who completed the video follow-up, which are shown in Table 11. However, the largest differences between those who completed the video follow-up and those who completed only the in-person are (a) their comfort levels with a smartphone, (b) their difficulty with video calling, and (c) whether they like video surveys.

Table 10 Logistic Regression Predicting Video Completion to Pre-Election Survey in Video-Mixed (Video>Web>Phone) Group, 2020 ANES (Unweighted)

Independent variables	*B*	*SE*	*z*	*p*
Age	−0.10	0.35	−0.30	.764
Education	1.03	0.34	3.05	.002
Female	0.18	0.16	1.11	.266
Income	0.16	0.32	0.50	.617
Black (r: White)	0.23	0.30	0.78	.435
Hispanic	0.12	0.32	0.36	.719
Other race	−0.05	0.33	−0.14	.886
Democrat (r: Independent)	1.73	0.40	4.36	.000
Republican	1.37	0.41	3.37	.001
Homeowner[a] (r: Rent/Other)	−0.41	0.20	−2.05	.040
Social trust	1.34	0.41	3.27	.001
Political interest	0.11	0.18	0.63	.530
Internet at home	1.01	0.69	1.46	.143
Access to internet on cell phone	0.87	0.36	2.44	.015
Constant	−4.65	0.86	−5.42	.000
N	698			
Pseudo R^2	.094			

Notes: Logistic regression is unweighted. All variables are from the pre-election survey. Non-dichotomous variables are recoded 0-1.

[a] Home tenure (i.e., homeowner) combined two versions of a question that were assigned as split half: the CPS and ANES versions.

Table 11 ANES 2023 Video Pretest Wave 1 (In-person) and Wave 2 (Video) Comparison (Unweighted)

Characteristic	In-person Only (*N* = 109)	Video (*N* = 48)	Total (*N* = 157)
	Percent	Percent	Percent
Comfort with smartphone			
Not at all comfortable	15.0	0.0	10.3
Slightly comfortable	14.0	6.3	11.6
Moderately comfortable	27.1	20.8	25.2
Very comfortable	43.9	72.9	52.9
χ^2	14.682*		

Table 11 (continued)

Characteristic	In-person Only (*N* = 109)	Video (*N* = 48)	Total (*N* = 157)
	Percent	Percent	Percent
Ever participated in a video call			
Yes	81.5	95.8	85.9
No	18.5	4.2	14.1
χ^2	5.650*		
Difficulty video calling today			
Easy	63.2	93.6	72.5
Not easy	36.8	6.4	27.5
χ^2	15.120***		
Use the internet at home			
Yes	87.2	95.8	89.8
No	12.8	4.2	10.2
χ^2	2.742#		
Like web surveys			
Like	43.0	51.1	45.5
Neither like nor dislike	14.0	25.5	17.5
Dislike	43.0	23.4	37.0
χ^2	6.322*		
Like paper surveys			
Like	33.3	48.9	38.1
Neither like nor dislike	31.5	40.4	34.2
Dislike	35.2	10.6	27.7
χ^2	9.974*		
Like video surveys			
Like	11.0	34.8	18.1
Neither like nor dislike	12.8	19.6	14.8
Dislike	76.1	45.7	67.1
χ^2	15.589***		

Notes: The Pearson's χ^2 and significance levels in the table are for the comparison of the in-person and video cases. Data are unweighted. #$p < .10$; *$p < .05$; **$p < .01$; ***$p < .001$

Conclusion

Findings and Limitations

Findings from the ANES 2020 show that response rates were lower and that attrition was higher for video than for other modes (RQ1). Those who were more likely to be successfully recruited into the video mode tended to be more highly educated than benchmarks and other modes and were more likely to identify as Democrats (RQ2). Those who stayed in the 2020 study for the post-election survey by video mode tended to be older and more educated than those who dropped out of the video mode before the post-election survey (RQ3). However, relative to Wave 1 of the 2023 pretest, the data suggest younger respondents were less likely to drop out of the study for Wave 2, which was a switch from in-person to video and is in line with other research examining attrition in a live video panel survey (Arrue et al., 2023). Educated respondents were also less likely to attrite. Both qualitative and quantitative data indicated that respondents who declined to participate in video interviews disliked or were distrustful of technology in general (and video, in particular), had privacy concerns, were uncomfortable with technology (including smartphones), or lacked equipment (RQ5).

Many potential designs and implementations can be used to recruit respondents, and despite using two different designs (i.e., ABS recruitment to video in two waves relative to in-person recruitment to video in the second wave only) with different implementations (i.e., using a combination of on-demand interviews and appointment-setting relative to appointment-setting only), some conclusions can be drawn about recruitment practices (RQ4) and reasons for refusals (RQ5). In both studies, missed appointment rates were high. One useful strategy to increase response was to increase appointment reminders, potentially including reminder modes with more immediacy and visibility, such as text message. Another strategy, offering prepaid incentives (relative to postpaid incentives), produced a modest increase in response, an effect consistent with the literature on the relative impact of incentives (e.g., Church, 1993; Singer & Ye, 2013). One strategy that was tried but did not work as anticipated was the offer from interviewers to show respondents how to connect to video so they would be prepared for Wave 2; most respondents turned down the offer. This was an attempt to alleviate respondent concerns or inexperience with the mode, but it suggests some of these concerns may be difficult for interviewers to overcome. Other failed strategies included increasing the incentive amount (which was relatively high to begin with) and the possibility of using an alternative video platform.

Recruitment to video was one of the main challenges of these studies, but the study designs themselves limited the video response rates and potential comparability to other studies using video. As part of the sequential mixed mode approach in the 2020 ANES, the design quickly shifted respondents from video to other modes in the video-mixed condition to prioritize overall response rate and sample size rather than video response. Because only 48% of the video-mixed group completed the video component, the analyses focused on mode rather than experimental groups. The 2023 Pretest used a convenience sample and had fewer contact attempts than a typical ANES study. In both studies, a longer field period might have allowed more respondents in hard-to-reach groups to be recruited into the studies and for larger sample sizes overall. However, many studies using video may also face similar constraints. Moreover, both studies faced similar hesitancy from respondents despite the different designs.

Discussion and Future Directions

The differences between who can be recruited into video relative to the population may be reduced over time as the technology continues to diffuse and people become more comfortable with it. The advantages of different modes can change over time because of changes in technology, changing respondent preferences for mode, and best practices in survey research. For example, phone surveys were once common, but the change from landline to cell phones along with the way people use their phones has contributed to the mode's declining popularity (e.g., Kennedy & Hartig, 2019; Olson, Smyth et al., 2021). Currently, 90% of American adults have a smartphone (Pew Research Center, 2024) and thus are equipped for video interviews, even if they have not yet participated in video communication. As respondents come to use video communication more frequently in their everyday lives, they may find the request to participate by video to be unremarkable, and recruitment will be less challenging. However, some differences are likely to remain in terms of who can be recruited to the mode, as respondent characteristics tend to differ by mode. This suggests that live video interviewing's best contribution to surveys may be as part of a mixed-mode design.

It is now common for studies to mix modes to address issues such as declining response rates and the potential for nonresponse bias (Atkeson et al., 2014, 2018; de Leeuw, 2005). In a mixed-mode design, individuals less likely to participate in one mode may participate in another, allowing for better representation of the population across modes. Because video appears more appealing to educated respondents and Democrats, a mixed-mode design with video could employ another mode to compensate. This could also extend beyond response rates and groups that are less likely to participate to take advantage of some of the other strengths and compensate for weaknesses across modes.

In assessing how viable live video is as a supplement or replacement for in-person interviews, questions remain. One of the largest is cost. Although expenses could be lower in video than traditional in-person designs, understanding cost will be challenging because it can be hard to disentangle those attributable to the mode itself from other elements of the study (see, e.g., Olson, Wagner et al., 2021, on difficulties in estimating costs in surveys). Additionally, best practices in recruitment have not yet been established, including the best ways to invite sampled respondents, establish a video connection with respondents, pursue a reminder strategy, or overcome respondent concerns about the mode. More research is also needed about data quality, but the evidence so far does seem to indicate that live video interviews produce data of similar quality to in-person interviews (Endres et al., 2023) and can exceed online questionnaires in terms of data quality (Conrad et al., 2023). While further investigation is needed into the methods and best practices for live video interviewing, we are cautiously optimistic about its potential to supplement or replace some in-person interviewing in mixed-mode survey designs.

Funding

This paper is based on work that was funded by the National Science Foundation under grant numbers SES-1835721 and SES-2209438 to the University of Michigan and grant number SES-1835022 to Stanford University.

Conflict of Interest

The authors declare no conflicts of interest.

References

American Association for Public Opinion Research. (2023). *Standard definitions: Final dispositions of case codes and outcome rates for surveys.* https://aapor.org/wp-content/uploads/2023/05/Standards-Definitions-10th-edition.pdf

American National Election Studies. (2019). *ANES 2016 Time Series Study full release* [dataset and documentation]. September 4, 2018 version. https://electionstudies.org/

American National Election Studies. (2021). *ANES 2020 Methodology Dataset release* [dataset and documentation]. July 19, 2021 version. www.electionstudies.org

American National Election Studies. (2022). *ANES 2020 Time Series Study full release* [dataset and documentation]. February 10, 2022 version. https://electionstudies.org/

Anderson, A. H. (2008). Video-mediated interactions and surveys. In F. G.Conrad & M. F. Schober (Eds.), *Envisioning the survey interview of the future* (pp. 95–118). Wiley. https://doi.org/10.1002/9780470183373.ch5

Arrue, J., Kelley, J., Edwards, B., Centeno, L., & Dulaney, R. (2023). Are video interviews for everyone? A comparison of CAVI ‘Stayers’ and ‘Adopters’ in a longitudinal household survey. [Conference session]. The *2023 FCSM Research and Policy Conference.* https://www.fcsm.gov/assets/files/docs/2023-conference-docs/E1.3_Arrue.pdf

Atkeson, L. R., Adams, A. N., & Alvarez, R. M. (2014). Nonresponse and mode effects in self- and interviewer-administered surveys. *Political Analysis, 22*(3), 304–320. https://doi.org/10.1093/pan/mpt049

Atkeson, L. R., Adams, A. N., & Alvarez, R. M. (2018). Mixing survey modes and its implications. In L. R. Atkeson, & R. M. Alvarez (Eds.), *The Oxford handbook of polling and survey methods,* (pp. 53-75). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190213299.001.0001

Bailenson, J. N. (2021). Nonverbal overload: A theoretical argument for the causes of Zoom fatigue. *Technology, Mind, and Behavior, 2*(1). https://doi.org/10.1037/tmb0000030

Barth, S. & de Jong, M. D. T. (2017). The privacy paradox – Investigating discrepancies between expressed privacy concerns and actual online behavior – A systematic literature review. *Telematics and Informatics, 34*(7), 1038-1058. https://doi.org/10.1016/j.tele.2017.04.013

Baruh, L., Secinti, E., & Cemalcilar, Z. (2017). Online privacy concerns and privacy management: A meta-analytical review, *Journal of Communication, 67*(1), 26–53. https://doi.org/10.1111/jcom.12276

Bearfield, A. & Smith, P. (2021). Covid-19 and face-to-face interviewing. *The Social Research Association Blog.* https://the-sra.org.uk/SRA/SRA/Blog/Covid19andfacetofacefieldwork.aspx

Blank, G., & Lutz, C. (2016). Benefits and harms from Internet use: A differentiated analysis of Great Britain. *New Media & Society, 20*(2), 618-640. https://doi.org/10.1177/1461444816667135

Chang, L., & Krosnick, J. A. (2009). National surveys via RDD telephone interviewing versus the Internet: Comparing sample representativeness and response quality. *Public Opinion Quarterly, 73*(4), 641-678. https://doi.org/10.1093/poq/nfp075

Church, A. H. (1993). Estimating the effect of incentives on mail survey response rates: A meta-analysis. *Public Opinion Quarterly, 57*(1), 62–79. http://www.jstor.org/stable/2749438

Conrad, F. G., Schober, M. F., Hupp, A. L., West, B. T., Larsen, K. M., Ong, A. R., & Wang, T. (2023). Video in survey interviews: Effects on data quality and respondent experience. *Methods, Data, Analyses, 17*(2), 135-170. https://doi.org/10.12758/mda.2022.13

Couper, M. P. (2000). Web surveys: A review of issues and approaches. *Public Opinion Quarterly, 64*(4), 464-494. https://doi.org/10.1086/318641

Current Population Survey. (November, 2020). *Voting and registration supplement [machine-readable data file] conducted by the Bureau of the Census for the Bureau of Labor Statistics*. Bureau of the Census [producer and distributor].

Curtin, R., Presser, S., & Singer, E. (2000). The effects of response rate changes on the Index of Consumer Sentiment. *Public Opinion Quarterly, 64*(4), 413–28. https://www.jstor.org/stable/3078736

DeBell, M., Amsbary, M., Brader, T., Brock, S., Good, C., Kamens, J., Maisel, N., & Pinto, S. (2022). *Methodology report for the ANES 2020 Time Series Study*. Stanford University and the University of Michigan. https://electionstudies.org/wp-content/uploads/2022/08/anes_timeseries_2020_methodology_report.pdf

de Leeuw, E. D. (1992). *Data quality in mail, telephone and face to face surveys*. Amsterdam: TT Publikaties.

de Leeuw, E. D. (2005). To mix or not to mix data collection modes in surveys. *Journal of Official Statistics, 21*(2), 233-255.

de Leeuw, E. D. & de Heer, W. (2002). Trends in household survey nonresponse: A longitudinal and international perspective. In R. M. Groves, D. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey Nonresponse*, (pp. 41–54). New York: John Wiley & Sons.

Dillman, D. A. (2020). Towards survey response rate theories that no longer pass each other like strangers in the night. In P. S. Brenner (Ed.), *Understanding survey methodology: Sociological theory and applications*, (pp. 15-44). Springer. https://doi.org/10.1007/978-3-030-47256-6_2

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed mode surveys: The tailored design method* (4th ed.). John Wiley & Sons Inc. https://doi.org/10.1002/9781394260645

Dutton W. H., & Shepherd A. (2006). Trust in the Internet as an experience technology. *Information, Communication & Society, 9*(4), 433–451. https://doi.org/10.1080/13691180600858606

Endres, K., Hillygus, D. S., DeBell, M., & Iyengar, S. (2023). A randomized experiment evaluating survey mode effects for video interviewing. *Political Science Research and Methods, 11*(1), 144–159. https://doi.org/10.1017/psrm.2022.30

Fauville, G., Luo, M., Queiroz, A. C. M., Lee, A., Bailenson, J. N., & Hancock, J. (2023). Videoconferencing usage dynamics and nonverbal mechanisms exacerbate Zoom Fatigue, particularly for women. *Computers in Human Behavior Reports, 10*, 100271. https://doi.org/10.1016/j.chbr.2023.100271

Federal Trade Commission (FTC). (2024). *Consumer Sentinel Network Annual Data Book 2023.* https://www.ftc.gov/system/files/ftc_gov/pdf/CSN-Annual-Data-Book-2023.pdf

Feuer, S., & Schober, M. F. (2026). Effects of the self-view window in live video survey interviews. *Methods, Data, Analyses, 20*(1), 124–178. https://doi.org/10.12758/mda.2025.07

Flood, S., King, M., Rodgers, R., Ruggles, S., Warren, J. R., Backman, D., Chen, A., Cooper, G., Richards, S. Schouweiler, M., & Westberry, M. (2023). *IPUMS CPS: Version 11.0* [Data set]. IPUMS. https://doi.org/10.18128/D030.V11.0

Fosslien, L., & Duffy, M. W. (2020). How to combat Zoom fatigue. *Harvard Business Review.* https://www.harvardbusiness.org/wp-content/uploads/2020/05/HBR-How-to-Combat-Zoom-Fatigue.pdf

Frankel L. L., & Hillygus D. S. (2014). Looking beyond demographics: Panel attrition in the ANES and GSS. *Political Analysis, 22*(3), 336-353. https://doi.org/10.1093/pan/mpt020

Groves, R. M. (1989). *Survey errors and survey costs.* Wiley. https://onlinelibrary.wiley.com/doi/book/10.1002/0471725277

Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly, 70*(5), 646–675. https://doi.org/10.1093/poq/nfl033

Groves, R. M., & Couper, M. P. (1996). Contact-level influences on cooperation in face-to-face surveys. *Journal of Official Statistics*, *12,* 63-84.

Groves, R. M., Fowler, Jr., F., Couper, M. P., Lepkowski, J., Singer, E., & Tourangeau R. (2004). *Survey methodology.* John Wiley and Sons.

Groves, R. M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: A meta-analysis, *Public Opinion Quarterly, 72*(2), 167–189. https://doi.org/10.1093/poq/nfn011

Hanson, T., Briceno-Rosas, R., Ainsaar, M., & Kartau, H. K. (2026). Live online video interviewing as a complementary mode to in-person interviews: Evidence from the European Social Survey. *Methods, Data, Analyses, 20*(1), 179–208. https://doi.org/10.25521/mda.663

Heerwegh, D. (2007). Mode differences between face-to-face and web surveys: An experimental investigation of data quality and social desirability. *International Journal of Public Opinion Research, 21*(1), 111-121. https://doi.org/10.1093/ijpor/edn054

Hupp, A. L., Conrad, F. G., Larsen, K. M., Schober, M. F., West, B. T., & Harrison, M. (2026). Dynamics of participating in video survey interviews. *Methods, Data, Analyses, 20*(1), 21–45. https://doi.org/10.25521/mda.6355

Jeannis, M., Terry, T., Heman-Ackah, R., & Price, M. (2013). Video interviewing: An exploration of the feasibility as a mode of survey application. *Survey Practice,* 6(1). https://doi.org/10.29115/SP-2013-0001

Keeter, S., Miller, K., Kohut, A., Groves, R.M. & Presser, S. (2000). Consequences of reducing nonresponse in a national telephone survey. *Public Opinion Quarterly, 64*(2), 125–148. https://doi.org/10.1086/317759

Kennedy, C., & Hartig, H. (2019). Response rates in telephone surveys have resumed their decline. *Pew Research Center.* https://www.pewresearch.org/short-reads/2019/02/27/response-rates-in-telephone-surveys-have-resumed-their-decline/

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

Liu, M., & Wang, Y. (2015). Data collection mode effect on feeling thermometer questions: A comparison of face-to-face and web surveys. *Computers in Human Behavior, 48*, 212–218. https://doi.org/10.1016/j.chb.2015.01.057

Luiten, A., Hox, J. & de Leeuw, E. D. (2020). Survey nonresponse trends and fieldwork effort in the 21st century: Results of an international study across countries and surveys. *Journal of Official Statistics, 36*(3), 469-487. https://doi.org/10.2478/jos-2020-0025

National Telecommunications and Information Administration (NTIA) (2016). *Lack of trust in Internet privacy and security may deter economic and other online activities*. https://www.ntia.gov/blog/2016/lack-trust-internet-privacy-and-security-may-deter-economic-and-other-online-activities

Olson, K., Smyth, J. D., Horwitz, R., Keeter, S., Lesser, V., Marken, S., Mathiowetz, N.A., McCarthy, J.S., O'Brien, E., Opsomer, J.D., Steiger, D., Sterrett, D., Su, J., Suzer-Gurtekin, Z. T., Turakhia, C., & Wagner, J. (2021). Transitions from telephone surveys to self-administered and mixed-mode surveys: AAPOR Task Force report. *Journal of Survey Statistics and Methodology, 9*(3), 381–411. https://doi.org/10.1093/jssam/smz062

Olson, K., Wagner, J., & Anderson, R. (2021). Survey costs: Where are we and what is the way forward?, *Journal of Survey Statistics and Methodology, 9*(5), 921–942. https://doi.org/10.1093/jssam/smaa014

Pew Research Center (2000a). *Who's not online*. https://www.pewresearch.org/internet/2000/09/21/whos-not-online-2/

Pew Research Center (2000b). *Trust and privacy online: Why Americans want to rewrite the rules.* https://www.pewresearch.org/internet/wp-content/uploads/sites/9/media/Files/Reports/2000/PIP_Trust_Privacy_Report.pdf.pdf

Pew Research Center (2021). *The Internet and the pandemic.* https://www.pewresearch.org/internet/2021/09/01/the-internet-and-the-pandemic/

Pew Research Center (2024). *Internet, Broadband Fact Sheet.* https://www.pewresearch.org/internet/fact-sheet/internet-broadband/

Schober, M.F. (2018). The future of face-to-face interviewing. *Quality Assurance in Education, 26*(2), 290-302. https://doi.org/10.1108/QAE-06-2017-0033

Schober, M. F., Conrad, F. G., Hupp, A. L., Larsen, K. M., Ong, A. R., & West, B. T. (2020). Design considerations for live video survey interviews. *Survey Practice, 13*(1). https://doi.org/10.29115/SP-2020-0014

Schober, M. F., Okon, S., Conrad, F. G., Hupp, A. L., Ong, A. R., & Larsen, K. M. (2023). Predictors of willingness to participate in survey interviews conducted by live video. *Technology, Mind, and Behavior, 4*(2). https://doi.org/10.1037/tmb0000100

Shockley, K. M., Gabriel, A. S., Robertson, D., Rosen, C. C., Chawla, N., Ganster, M. L., & Ezerins, M. E. (2021). The fatiguing effects of camera use in virtual meetings: A within-person field experiment. *Journal of Applied Psychology, 106*(8), 1137–1155. https://doi.org/10.1037/apl0000948

Singer, E., & Ye, C. (2013). The use and effects of incentives in surveys. *The Annals of the American Academy of Political and Social Science, 645,* 112–141. http://www.jstor.org/stable/23479084

Singer, E., Van Hoewyk, J., Gebler, N., & McGonagle, K. (1999). The effect of incentives on response rates in interviewer-mediated surveys. *Journal of official statistics, 15*(2), 217-230.

Sklar, J. (2020). "Zoom Fatigue" is taxing the brain. Here's why that happens. *National Geographic.* https://www.nationalgeographic.com/science/article/coronavirus-zoom-fatigue-is-taxing-the-brain-here-is-why-that-happens

Sun, H., Conrad, F. G., & Kreuter, F. (2021). The relationship between interviewer-respondent rapport and data quality. *Journal of Survey Statistics and Methodology, 9*(3), 429-448. https://doi.org/10.1093/jssam/smz043

West, B. T., & Blom, A. G. (2016). Explaining interviewer effects: A research synthesis. *Journal of Survey Statistics and Methodology, 5*(2). 175–211. https://doi.org/10.1093/jssam/smw024

West, B.T., Ong, A.R., Conrad, F. G., Schober, M.F., Larsen, K.M., & Hupp, A.L. (2022). Interviewer effects in live video and prerecorded video interviewing. *Journal of Survey Statistics and Methodology, 10*(2), 317-336. https://doi.org/10.1093/jssam/smab040

Williams, D., & Brick, J. M. (2018). Trends in U.S. face-to-face household survey nonresponse and level of effort. *Journal of Survey Statistics and Methodology*, 6(2), 186-211. https://doi.org/10.1093/jssam/smx019

Wood, M., Cole, K., Dangerfield, P., Stockman, A., & Spencer, S. (2020). *BCS70 Video Call Pilot: Report of pilot stage.* https://cls.ucl.ac.uk/wp-content/uploads/2017/02/BCS70-Video-Call-Pilot.pdf

Appendix A

Implementation of the Video Interviews in the 2020 ANES and 2023 Pretest

In the 2020 ANES Time Series studies, when an interviewer was ready to take an available live video interview, the interviewer would launch a case management system within a web browser. The case management system would then provide the interviewer with a list of available respondents who had entered the live video interviewing queue. The interviewer could select among the available respondents, after which the interviewer and respondent would be prompted to launch Zoom. Zoom would launch for the interviewer in a second tab within their web browser, at which time the interviewer could see their own image and that of the respondent upon their joining. The interviewer would read a welcome script and then be prompted to launch the Blaise instrument (survey), which appeared in a third tab of their browser. The interviewer would navigate between the Zoom and survey tabs as necessary. Whenever a showcard was indicated for a question, the showcard was displayed to the respondent using screen sharing within Zoom, using a showcard file that was stored locally on the interviewer's computer.

To conduct video interviews in the 2023 Pretest, interviewers used a single laptop display, with the left-hand portion of the screen containing the instrument, and the right-hand portion containing the showcards to screenshare with the respondent. The Zoom window was kept on the right-hand portion of the screen when showcards were not being shared, and between the two windows when sharing the showcard.

Appendix B

Incentive Experiment and Connection Test Question in 2023 Pretest

To conduct the incentive experiment and connection test (described in the Design & Implementation section and the Results section under RQ4), the ANES 2023 Pretest used the following question:

> "We would like to do another interview in a few weeks. [INCENTIVE EXPERIMENT: We will give you $25 in cash right now if you agree to do the interview./ We will give you $25 for completing that interview.] I won't need to come to your home this time. We would like to complete this interview on video using Zoom. I will make it very easy by scheduling a time that is convenient for you, and by emailing you a link to click on to join the interview. There is nothing to download or install. [CONNEC-

TION TEST: I can show you how easy it will be right now if you'd like.] Are you willing to do this second interview over video?"

Appendix C

Attrition Among Fresh Web Cases, 2020 ANES

Table C1 Attrition among Fresh Web Cases in the 2020 ANES (Unweighted)

Characteristic	Completed Pre Only (Attrition Cases) (*N* = 549)	Completed Pre and Post (Reinterview Cases) (*N* = 4,394)	Difference (Attrition-Reinterview)
	Percent	Percent	Percent
Age			
18-29	17.0	14.0	3.0
30-39	17.2	17.8	−0.6
40-49	14.6	15.0	−0.4
50-59	18.0	16.5	1.5
60-69	17.4	19.1	−1.7
70 or older	15.8	17.6	−1.8
Mean deviation			1.5
χ^2	4.996		
Education			
Less than HS cred.	8.2	4.3	3.9
HS credential	22.9	15.5	7.4
Some college/AA degree	34.5	35.1	−0.6
Bachelor's degree	20.0	25.5	−5.5
Graduate degree	14.4	19.6	−5.2
Mean deviation			4.52
χ^2	44.902***		
Income			
Under $25,000	21.4	20.7	0.7
$25,000-$49,999	21.4	18.5	2.9
$50,000-$99,999	24.0	28.9	−4.9
$100,000 or more	33.2	31.9	1.3
Mean deviation			2.5
χ^2	5.818		

Table C1 (continued)

Characteristic	Completed Pre Only (Attrition Cases) (*N* = 549)	Completed Pre and Post (Reinterview Cases) (*N* = 4,394)	Difference (Attrition-Reinterview)
	Percent	Percent	Percent
Sex			
Male	48.9	45.4	3.5
Female	51.1	54.6	−3.5
Mean deviation			3.5
χ^2	2.296		
Race/Ethnicity			
White, non-Hispanic	66.9	73.5	−6.6
Black, non-Hispanic	9.2	8.8	0.4
Hispanic	11.7	8.7	3.0
Other, non-Hispanic	12.2	9.0	3.2
Mean deviation			3.3
χ^2	12.606**		
Party ID 7-point			
Strong Democrat	18.7	23.3	−4.6
Not strong Democrat	13.9	10.6	3.3
Lean Democrat	10.5	12.3	−1.8
Independent	12.9	12.8	0.1
Lean Republican	10.9	10.3	0.6
Not strong Republican	8.7	10.4	−1.7
Strong Republican	24.4	20.3	4.1
Mean deviation			2.3
χ^2	15.709*		
Home Tenure[a]			
Own	62.3	68.6	−6.3
Rent/Other	37.7	31.4	6.3
Mean deviation			6.3
χ^2	8.691**		

Notes: Data include all fresh cross-sectional web cases from the 2020 ANES Time Series Study, which includes web cases from the web-only, web-mixed, and video-mixed group. All web cases were offered web either first or second in the sequential, mixed groups. The Pearson's χ^2 and significance levels in the table are for the comparison of the attrition and reinterview cases in the first two columns. Data are unweighted. $^{\#}p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

[a] Home tenure combined two versions of a question that were assigned as split half: the CPS and ANES versions.

Appendix D

Methods of Analysis for Issue Mentions by the Non-Respondents in ANES 2020 and the 2023 Pretest

The two interviewers who made up the refusal conversion team as part of the ANES 2020 Nonresponse Video Intervention were given an introductory script and a post-contact questionnaire where they recorded the outcome of calls. The script included the question: "Did you encounter any problems or have any concerns about the interview? What are they?" The interviewers recorded whether respondents mentioned any issues using the following question, which was part of the post-contact questionnaire:

Issues mentioned by respondent (check all that apply)

1. Uncomfortable with technology generally
2. Uncomfortable with videoconferencing generally
3. Uncomfortable with Zoom specifically
4. Did not want their image appearing on video
5. Did not want their surroundings appearing on video
6. Concerned that they are being recorded
7. Privacy/confidentiality of videoconferencing generally
8. Privacy/confidentiality of Zoom specifically
9. Security concerns about installing Zoom app
10. Technical difficulties in installing the Zoom app
11. Problems with them seeing us
12. Problems with us seeing them
13. Problems with them hearing us
14. Problems with us hearing them
15. Does not have a video camera
16. Does not have a screen/display
17. Does not have speakers

18. Does not have a microphone

19. Does not have a capable device (smartphone, tablet, computer, laptop) at all

20. Other, specify ___________________

Each issue was included in the ANES 2020 Methodology Dataset Release (2021) as binary response variables. Frequencies were run on each of the 20 issues and included in the table if the issue occurred more than once. The set of issues with a frequency greater than one were then organized by topics into themed categories that make up Table 8.

For the 2023 Pretest, interviewers asked respondents who declined to participate in the video follow-up the question, "Why do you not want to do the video interview? (Write reasons.)" and took notes on their answers. Notes were taken for 75 respondents. A coding scheme was created to code these notes for reasons for non-participation. To create the coding scheme, categories from the issue checklist used in the ANES 2020 Nonresponse Video Intervention were used, additional categories were created, and modifications were made to some of the 2020 checklist items after a review of the responses. Each response was then coded for the presence or absence of the following reasons:

- Too busy
- Don't want to
- Don't have equipment generally
- Does not have a microphone
- Does not have a capable device (phone, computer)
- Does not have a video camera
- Does not have speakers
- No internet
- Video privacy or image concerns generally
- Do not want image appearing on video
- Do not want surroundings appearing
- Concern they are being recorded
- Privacy/confidentiality of video
- General privacy concerns
- General lack of trust in technology
- Dislike technology generally
- Don't like Internet
- Don't like Zoom/video
- Don't use Zoom/video

- Not familiar/comfortable with video
- Not familiar/comfortable with technology generally
- Don't like surveys
- Other

Responses were coded for present or absent for all the categories because some responses contained multiple reasons. For example, the response, "Not comfortable talking to someone I don't know. Don't want my image online. AI Artificial Intelligence is a big issue," includes several reasons. To create Table 9, frequencies were run on each of the coded categories (i.e., reasons for non-participation). Then the reasons were organized by topic into themed categories, and reasons that were not present in the data were cut. The average percent agreement based on double-coding of the variables appearing in the table was 94.1%.

methods, data, analyses | Vol. 20(1), 2026, pp. 87-107 DOI: 10.25521/mda.657

Video Interviewing for Standardized Achievement Assessments with Children

Paula Fomby[1], Narayan Sastry[2], Helen Johnson[2], Sarah Crane[2] & Veronica Connors-Burge[2]

[1] *University of Pennsylvania*
[2] *University of Michigan*

Abstract

We describe the design, execution, and results of a pilot study with a convenience sample ($N = 32$ children aged 5-17 years in 15 families) to evaluate the operational feasibility of video interviewing as a mode to administer standardized assessments of math, reading, and vocabulary skills to children aged 5-17 years in the United States. We review the videoconferencing interface from the perspective of the test examiner and examinee, describe protocols to provide cellular-enabled electronic tablets for examinees' use at home, summarize the fieldwork experience, and describe lessons learned. We conclude that video interviewing is a viable alternative to in person administration from an operational standpoint and can be useful when efforts to reach study participants in-person are cost-prohibitive or ineffective.

Keywords: remote survey administration, standardized testing, children, pilot study, survey operations

© The Author(s) 2026. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Standardized Testing with Children in Survey Research

Children's demonstrated skill in literacy and math is an important indicator of well-being and development for individuals and populations (Organisation for Economic Cooperation and Development, 2020). Consequently, many large-scale surveys of child development include standardized assessments of children's performance in core academic subjects. Raw scores on these assessments are age-standardized against national norms to provide metrics that may be used to describe and explain population variation in learning achievement and associated outcomes.

To date, standardized assessments with children in community-based samples have been conducted in-person during field interviewers' visits to families' homes or other settings where the interviewer can manage the testing environment. In-person administration is attractive because it allows the interviewer to identify the optimal test setting within a child's home, focus children's attention on test stimuli, deflect interference from other family members, and prevent children's efforts to seek outside help to answer questions. However, for at least four reasons, in-person assessment has become increasingly challenging. First, home visits in national samples are expensive compared to telephone and web interviewing (Schober, 2018), and in longitudinal studies, home visits become more expensive over time as participants move and become geographically dispersed. Second, survey respondents have become increasingly resistant to home visits in response to time constraints and preferences regarding privacy. Further, there is some evidence of declining fieldwork effort to recruit resistant respondents to in-person interviews (Bates et al., 2008; Brick & Williams, 2012; Vercruyssen et al., 2011; Williams et al., 2018). Third, scheduling changes for in-person appointments can be costly and inefficient. Especially when interviewers need to travel some distance to a respondent's home, broken appointments and last-minute rescheduling complicate the logistics of completing a home visit. Fourth, the Covid-19 pandemic demonstrated that public health emergencies or other extreme events can make travel or physical proximity unjustifiably risky. These factors have led to the consideration of a range of new approaches to the remote collection of various measures of children's development (see Tsuji et al., 2022).

Video interviewing potentially offers a cost-effective, unintrusive, and safe option to conduct standardized assessments with children at home, but to date there is no published empirical evidence of the operational feasibility of this approach. Video interviewing with respondents in their homes has been adopted successfully as a replacement or complement to face-to-face survey interviewing

Direct correspondence to
Paula Fomby, University of Pennsylvania, Department of Sociology, 3718 Locust Walk, Philadelphia, Pennsylvania 19130, USA
E-mail: pfomby@sas.upenn.edu

with adults (Centeno et al., 2024; Hanson et al., 2026). And prior work has demonstrated that video interviewing is an efficient, valid, and reliable method to conduct assessments when children and adolescents are tested in schools under tightly controlled circumstances (Wright, 2018). We offer the first evaluation of whether such success can translate to standardized assessment with children in home environments. A similar approach has recently been pilot tested for the first time for the remote collection of anthropometric assessments for adolescents and adults (Ghosh-Dastidar et al., 2020) and to assess preschool children's anthropometrics and motor skills (Button et al., 2023). This approach could also be used for other measures that require standardized measurement and supervision by or interaction with a trained interviewer.

Purpose and Aims of the Pilot Study and This Paper

The purpose of the pilot study and this paper is to evaluate the operational feasibility of video interviewing to administer standardized achievement assessments at home with school-aged children in the United States. The primary motivation is to ascertain whether this mode of data collection could be adopted in future waves of the Panel Study of Income Dynamics Child Development Supplement (CDS), a US national study of the well-being of minor children and their caregivers conducted by telephone and in-person interview approximately every five years since 1997 (McGonagle & Sastry, 2015). For several reasons, CDS is well-positioned to adopt video interviewing. First, the CDS sample is geographically diffuse. CDS is embedded in the Panel Study of Income Dynamics (PSID), a household-based panel study started in 1968 that has followed original household members and their descendants to the present day. Sampled households were geographically clustered when PSID began, but families have since become nationally dispersed. Second, CDS includes families living in rural or otherwise geographically remote areas where it is not economically efficient to hire local interviewers. Field interviewers' travel from more densely populated areas to rural or remote areas is expensive and must be carefully coordinated to permit visits to multiple participating households over a few days. Third, families with children are participating in CDS during a busy life stage, when parents may be juggling their own employment and the needs of preschool and school-aged children (Vercruyssen et al., 2011). Beyond the logistical challenge of scheduling a home visit, such a request can risk families' eventual attrition from the study if they perceive continued participation as a significant burden (Sastry et al., 2021). Giving families the choice to participate by video interview or in person could improve participation and retention (see Conrad et al., 2017).

In sum, the use of video interviewing to administer standardized assessments remotely could offer a strategy to control costs and increase efficiency during fieldwork, increase wave-specific response rates, and reduce panel study attrition when offered to families who are unwilling or unable to schedule a home visit. This pilot study is the first step in a larger program to evaluate the merits of this approach.[1] The aims are to assess four aspects of video interviewing to administer standardized assessments: (a) technical development and implementation of a video interface for the standardized tests that was comparable to the in-person format from the test taker's perspective; (b) application development to establish a stable, secure videoconferencing session between the test administrator and test taker; (c) protocols to provide cellular-enabled electronic tablets to test takers for use during the test administration; and (d) data production, as measured by test completeness and perceived test taker burden. The first three aspects of the pilot study are novel to CDS and the survey operations team, and to our knowledge, these protocols have not been tested previously on a sample of children participating at home.

The pilot study was conducted with a small convenience sample (*N* = 32 children ages 5-17 years in 15 families) in one US state. The use of a small convenience sample allowed the study design team to put additional resources into instrument and protocol development and to evaluate these multiple new technologies and procedures with relatively little risk. The tradeoff is that the findings may underestimate the challenges that could arise from using video interviewing for standardized testing in larger and more diverse samples. At a minimum, though, the outcomes provide useful lessons that permit better-informed development and evaluation for large-scale studies as they move toward integrating video interviewing with other more long-standing modes of data collection.

Below we describe the pilot study design and its implementation. We conclude with an assessment of operational feasibility, a discussion of lessons learned, and attention to the study's limitations.

Pilot study design

This section describes the steps to develop and implement a method to conduct standardized testing with children at home, from technical development through test administration. Development, fieldwork, and post-production processing for the pilot study were conducted by Survey Research Operations (SRO) at the Institute for Social Research, University of Michigan, between July 2021 and April 2022. All study materials and protocols were approved by the Health Sciences and Behavioral Sciences Institutional Review Board at the University of Michigan.

1 Building on the pilot study, CDS will evaluate mode equivalence and integrate video interviewing with other modes of data collection in a larger and more representative pretest sample before adopting the method for use in the study's active sample.

Technical Development

Instrument Adaptation for Video Interviewing

The Woodcock-Johnson IV (WJ-IV) is a battery of 20 subtests designed to characterize children's learning performance in various domains relative to their age or grade peers or to assess children's learning progress over time (Mather & Jaffe, 2016). The pilot study included three subtests used in prior waves of CDS: (a) Letter-Word Identification (word and letter recognition or pronunciation); (b) Passage Comprehension (making inferences from written passages); and (c) Applied Problems (problem-solving through quantitative reasoning). For the pilot study, subtests were drawn from WJ-IV Form B, and all testing was completed in English.

During in-person administration, the WJ-IV subtests are presented in a standing easel binder referred to as a stimulus book. The test administrator and test taker sit across from each other, and the stimulus book is positioned between them. The test taker sees one page containing a set of stimuli, and the administrator sees correct responses to those stimuli on a page on the back side of the easel. A child begins each subtest at a point determined by his or her current age or grade in school. This starting point is referred to as a basal. If the child responds incorrectly to each of the first several stimuli, he or she is rerouted to an easier starting point, and the basal is reset. With each correct response, the child advances to a more challenging stimulus. The subtest terminates when the child provides multiple consecutive incorrect responses after the basal has been established or responds to the last stimulus in the subtest. Raw scores are calculated as a function of the child's basal and the number of items answered correctly beyond the basal. These raw scores are then transformed into a set of criterion-based scores to measure individual ability and norm-referenced scores to measure performance relative to age or grade peers.

WJ-IV testing rules had been programmed into Blaise survey software previously to support computer-assisted in-person interviewing in CDS. The program uses preloaded information about a test taker's age and grade in school to establish the preliminary basal for each subtest. After each stimulus is presented, the test administrator records whether the child has provided a correct or incorrect response and follows programmed prompts to move through each subtest based on performance. To adapt the in-person administration design for video interviews, each test stimulus page was digitally scanned, formatted, and uploaded into the existing Blaise application to be screenshared by the test administrator to a test taker.[2]

2 Riverside Insights, the publisher of Woodcock-Johnson IV, granted permission to adapt copyrighted material for electronic use.

Hardware

A laptop computer and external monitor in portrait orientation were provided to each test administrator (Figure 1). The Blaise application for entering test takers' responses to test stimuli was programmed to appear on the laptop monitor and the test stimulus page to appear on the external monitor so that both were visible to the test administrator. A high-resolution webcam and headset with a built-in microphone were also provided.

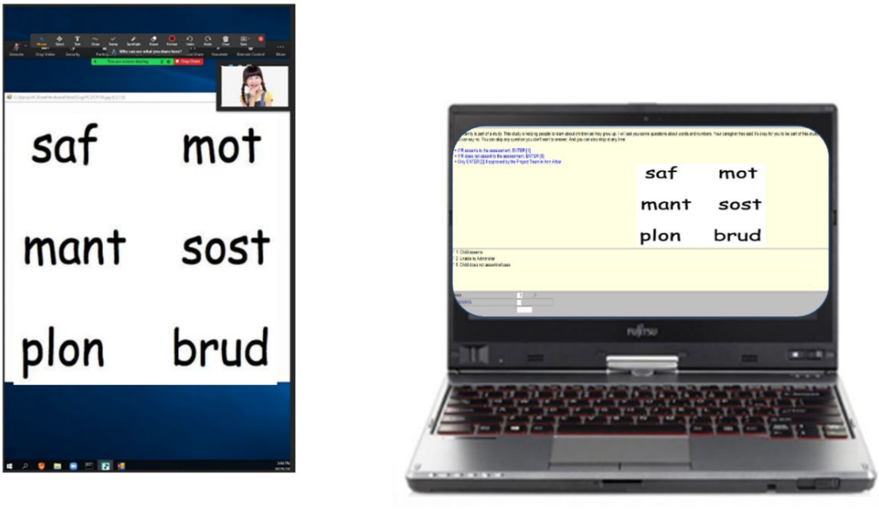


Figure 1 External monitor (left) and laptop computer (right) configured for use by the test administrator during the remote administration of the Woodcock-Johnson IV Letter-Word subtest

Test takers used a Samsung Model 5 electronic tablet configured solely to support remote WJ-IV administration on the Zoom video conferencing platform (Figure 2). Electronic tablets were used in order to standardize the test-taking format across home environments and to allow participation where families lacked a home computer or a high-speed internet connection. Each tablet included a built-in video camera and microphone and a 10.5-inch display, consistent with the WJ-IV publisher's guidelines for electronic administration (Riverside Insights, 2020).[3] A push notification application was installed on the home screen so that

[3] Some recommendations provided by the publisher were not incorporated into the pilot study. These include a peripheral camera directed at the examinee's workspace, a headset with two earphones and a microphone for the examinee's use, and a mouse or trackpad for the examinee's use to point to responses on the screen. The last recommendation was obviated by examinees' use of Zoom's annotation function for the same purpose.

test takers could join the test administrator's active video conferencing session. Mobile device management software was installed on each tablet so that its contents could be erased remotely if it were lost or stolen. No personal information was stored on the tablets. Tablets were enabled with a Verizon cellular data plan to establish the video connection between the test administrator and test taker.[4]

Figure 2 Electronic tablet configuration for use by the respondent during remote administration of the Woodcock-Johnson IV Letter-Word subtest

The video conferencing interface was determined by Zoom's format. During an active session, the test administrator and test taker observed one another through the video feed from their respective web cameras. To minimize distraction, video self-view was disabled for both parties (see Feuer & Schober, 2026, on the effect of self-view on survey response). The test administrator used Zoom's screenshare function to display stimulus pages on the test taker's electronic tab-

[4] As of 2017 (the most recent data available at the time of the pilot study), Verizon mobile wireless coverage reached 97.4% of the US resident population (Federal Communications Commission, 2017).

let screen. During screen sharing, the video tile showed the active speaker in the upper right-hand corner of the screen and was positioned so that it did not block text on the stimulus page.

Most test stimuli required a verbal response, but for young children, some items required the child to point to an object on the screen. To accommodate this, Zoom's annotation function was enabled for both the test administrator and test taker. The annotation feature allows a user to draw a point, line, circle, or other shape on an electronic tablet's touch-enabled screen with their finger. The image appears as a visible overlay on the shared screen, allowing all participants to see the annotation as it is being made. During administration, the test administrator prompted young children to use this feature to select their response on the screenshared stimulus page.

Staffing

Two field interviewers employed by SRO who had experience administering in-person WJ assessments to children were staffed as test administrators. We refer to these staff members as test administrators going forward. In addition to standard training, they received four hours of specialized training in procedures for the remote administration of WJ-IV assessments. This included training in the use of an external monitor and the Zoom interface as well as paired practice with the electronic tablets to experience the test taker's perspective. During fieldwork, test administrators worked from their own homes using computer and communications hardware supplied by SRO.

Sample Recruitment

The planned sample size was 30 children aged 5-17 years in 20 families, which was expected to provide sufficient variation across child age, family socioeconomic status, and geographic location (urban vs. rural) to distinguish contexts in which video interviewing was more or less feasible. Project management and test administrators recruited from their personal social networks a convenience sample of 19 families that included 40 age-eligible children. Sample composition was captured through an electronic screening form programmed in Qualtrics that was completed by the contact person in each recruited family. Four families were unable to participate due to scheduling conflicts during the fieldwork period, and one child was coded out as ineligible due to disability.[5] The final

---

5 Criteria for ineligibility are developed by SRO in consultation with study investigators. In the current case, children with significant vision or hearing impairments or a limiting developmental disability were categorized as ineligible.

sample included 32 children in 15 families, and all eligible families participated in the study. All families resided in the same state in the US Midwest.

Table 1 presents descriptive statistics for the pilot study sample and for families with minor children (5-17 years old) in the same state in 2022 (American Community Survey, Ruggles et al., 2025).[6] Compared to the state population, children were about two years younger in the pilot study sample (median age = 9 years) compared to children in the state (11 years), and pilot study caregivers were more often non-Hispanic white (78% versus 70% among household heads statewide) and highly educated (69% held a Bachelor's degree or higher versus 37%). About one-fifth of recruited families lived in areas classified as rural (vs. 12% in non-metropolitan areas statewide). To the extent that these demographic, socioeconomic, and geographic indicators are correlated with performance, pilot study results may not be generalizable to other groups.

Appointment Scheduling

Test administrators were assigned to families with whom they had no prior personal contact. During a telephone call, the test administrator verified the eligibility of children rostered on the screening form as potential participants. Test administrators then scheduled appointments based on the family's availability. With the caregiver's permission, the administrator sent a Google Calendar invitation by e-mail so that the caregiver could import the appointment to their own electronic calendar. One day prior to the appointment, the administrator called the caregiver with a reminder about the upcoming appointment and to address any questions.

Appointments were scheduled to take place at least one week after recruitment to allow time to ship study materials to a family's home using United Parcel Service (UPS) two-day delivery. Each shipment included a study information sheet, the electronic tablet in a protective case,[7] a printed return label, and instructions directed to the participating child's caregiver (i.e., a parent or guardian) on how to use and return the tablet. Use instructions included written guidance and accompanying annotated screenshots on how to (a) power the tablet on, adjust the volume, and navigate to the home screen; (b) use the push notification application to enter a Zoom videoconferencing session; (c) turn on the video camera and use the annotation function in Zoom; and (d) exit Zoom. The caregiver was also directed to prepare a quiet space in the home free from

6 Child's sex was not recorded.

7 To eliminate the risk of loss, no power cord was included in the shipment. Tablets had an estimated 14.5-15 hours of active battery life before needing to be recharged. When powered off for shipment, the battery can remain charged for several weeks. No instances of data loss or interruption caused by device power issues were reported.

distraction where the child could complete the assessment and, if relevant, to keep other children out of earshot until it was their turn to participate.

Table 1 Descriptive statistics, Woodcock Johnson video interview pilot study sample (*N* = 32 children in 15 families) and same-state families with children age 5-17 years (American Community Survey [ACS], 2022)

	Pilot study proportion (*N*)	State proportion
Number of children[a,d]		
1	0.27 (4)	0.30
2	0.40 (6)	0.40
3+	0.33 (5)	0.30
Child age (years)[e]		
5–8 years	0.28 (9)	0.29
9–12 years	0.38 (12)	0.30
13–17 years	0.34 (11)	0.42
Educational attainment[b]		
Less than high school	0.00 (0)	0.08
High school/GED	0.09 (3)	0.20
Some college	0.06 (2)	0.23
Associate's degree	0.16 (5)	0.11
Bachelor's degree	0.22 (7)	0.22
Master's degree	0.31 (10)	0.12
Professional/doctoral degree	0.16 (5)	0.03
Race and Hispanic ethnicity[b]		
Non-Hispanic white	0.78 (25)	0.70
Non-Hispanic Black	0.09 (3)	0.15
Non-Hispanic Asian	0.06 (2)	0.04
Non-Hispanic other	0.00 (0)	0.05
Hispanic (any race)	0.06 (2)	0.06
Rurality[c]		
City or township	0.82 (26)	0.88
Rural	0.19 (6)	0.12

[a] Refers to the number of minors participating in the pilot study.
[b] Refers to caregiver characteristics in pilot study and to householder characteristics in ACS.
[c] Measured by metropolitan status in the ACS.
[d] Number of children in Pilot: M = 2.13, SD = 0.92; Number of children in State population: M = 2.16, SD = 1.09.
[e] Child age in Pilot: Median = 9, M = 9.5, SD = 3.41; Child age in State population Median = 11, M = 11.2, SD = 3.72.

Test Administration

At the appointed time, the test administrator telephoned the caregiver to confirm that the shipment had arrived,[8] the tablet was powered on, the caregiver was familiar with the tablet's functionality as described in the user guide. and participating children were available and prepared to begin. The administrator then instructed the caregiver to initiate the video conferencing session via the push notification application on the tablet's home screen, after which the telephone connection between the test administrator and caregiver was terminated. In the videoconferencing session, the test administrator obtained informed consent from the caregiver and verbal assent from the child for the child's participation. The test administrator also requested the caregiver's consent to record the session for quality control.[9] The test administrator then launched the first subtest (Letter-Word Identification).[10] The caregiver was asked to remain nearby and to allow the child to complete each subtest independently.

In families with more than one participant, children completed the assessments back-to-back.[11] When all participants had finished, the test administrator reminded the caregiver to package and return the electronic tablet and terminated the video conference session. The test administrator contacted UPS to schedule prepaid pickup of the packaged tablet, usually for the following day. Each child who completed the assessments received a check for $50 as a token of appreciation after the electronic tablet was received at SRO. Blaise data were transferred to SRO using a secure file transfer protocol and removed from the test administrator's laptop daily. SRO technical support staff recharged and sanitized each electronic tablet and checked device settings before shipping the tablet to the next participating family.

Fieldwork occurred over three weeks beginning in the second week of February 2022. At the conclusion of fieldwork, raw data from the assessments were batch uploaded to the proprietary WJ-IV online scoring system with individual records distinguished by a unique numeric identifier. No personally identifying information was included in the upload. Calculated scores were returned from the scoring system in a comma-delimited text file.

---

8 Delivery was also confirmed independently by the service provider.

9 All caregivers provided consent to record. The recording captured the videoconferencing session as it appeared on the test administrator's external monitor. Consent to record was not required to proceed with the test administration.

10 Prior to launching the first subtest, the caregiver was also advised to provide the child with paper and pencil for use in the Applied Problems section, consistent with guidelines for in-person administration.

11 In one family with three children, one child participated the day after the other two.

Outcomes

In this section, we review the outcomes of the pilot study's implementation. Overall, the study demonstrates the operational feasibility of video interviewing for standardized testing with school-aged children while providing some lessons to improve efficiency and ensure test fidelity. We assess outcomes using six sources of information: a debriefing with test administrators at the end of fieldwork; test administrators' responses to closed-ended survey questions about each test taker's experience with each subtest; test administrators' open-ended written comments in Blaise; a log of shipment and return dates for the electronic tablets; timing data from Blaise; and participants' calculated WJ-IV scores.

Technical Systems

Here we review the implementation of study design choices related to Blaise programming, hardware configuration, cellular coverage, and videoconferencing format. Most worked as planned, with some caveats.

The integrated laptop PC and external monitor allowed test administrators to observe test takers, monitor the screenshared presentation of subtest stimulus pages, and record test takers' responses without toggling between screens. In the post-fieldwork debriefing, both test administrators noted that it was helpful to see the active test stimulus page on the external monitor while the test subject was also looking at it, in contrast to in-person administration with an easel, where the test stimulus page is visible only to the test taker. However, both also stated that additional training (beyond the four hours provided) on the hardware configuration and Zoom's functionality before beginning fieldwork would have increased their familiarity and confidence.

We use the test administrators' debriefing and their open-ended comments to describe participating families' experience with the videoconferencing application. All caregivers successfully used the push notification on the electronic tablet's pared-down home screen to join the videoconferencing session. No information is available about whether caregivers found the printed user guide sent in advance or the real-time guidance from the test administrator to be more helpful for getting started. It is important to note that caregivers and children in this relatively highly educated sample may have been familiar with push notifications and the Zoom interface.

Cellular data coverage on the tablets was mostly stable and reliable and provided clear communication. There were three instances of unstable or lost connections (one each in a city, township, and rural area), but these were resolved and did not prevent completion of the assessments. Had connectivity issues persisted, test administrators were trained to ask families to travel to another location, such as a library or the home of a friend or family member, to complete the assessment. However, it was never necessary for a family to travel to receive cellular service for the study. Other studies wishing to implement a similar design should review service coverage in the areas where sample members reside and develop protocols for cases where service is unstable or unavailable.

All caregivers were able to turn on the annotation feature at the beginning of the videoconferencing session. In open-ended comments provided after children completed their subtests, the test administrator noted that in one family the caregiver helped younger children (aged 6 and 9 years) to use the annotation function. No other observations of caregiver assistance related to children's use of the annotation function were reported. However, the test administrator's instruction to young children on how to use the feature added administration time. Children in the youngest age group (aged 5-8 years) were the most likely to respond to stimuli in the Letter-Word Identification subtest that required them to use the annotation function (e.g., to identify a letter of the alphabet), and children in that age group needed about twice as much time to complete the subtest (M = 6.66 minutes) compared to test takers in middle childhood (aged 9-12 years, M = 3.33 minutes). The Passage Comprehension test also required early readers to point to their responses for some stimuli (e.g., to identify the shape represented by a printed word), resulting in longer administration times compared to the middle childhood group (M = 11.4 minutes in early childhood vs. 10 minutes in middle childhood; see Table 2).

Test administrators offered two further recommendations based on their fieldwork observations. First, materials provided to participating families should include a stand to hold the tablet and instructions for the child to sit at a table while completing the subtests. In the pilot study, children held the tablet by hand. During the Applied Problems subtest, children frequently put the tablet down in order to solve problems. Because the test administrator could not see the child, it was not possible to know whether they were solving the problems by hand on their own. Second, Zoom features should be locked down. Test administrators observed that some children changed the name field or attempted to alter the annotation functionality.

Hardware Durability and Shipment

Test administrators and project staff reported that the electronic tablet hardware worked reliably in the field. All tablets were fully charged and operational and the screens intact when the devices were delivered to test takers' households, and there was no reported damage when tablets were returned to SRO throughout fieldwork. No tablets were reported missing or stolen.

All tablets shipped from SRO using UPS two-day delivery service reached recipients as scheduled one to two days before the scheduled appointment. Test administrators usually scheduled UPS to pick up the tablet for return the day after the test administration was completed, and SRO logged receipt two to three days later. In one case, a test administrator drove to a participating family's home to collect the tablet when it had not been picked up by UPS after several days. Upon receipt at SRO, tablets were processed for reuse and available for shipment to the next household one to two business days later. The interval between shipments for a given tablet ranged from seven to 15 calendar days, with a median of 10 days. The log indicates that longer intervals occurred because of lags in scheduled appointments rather than delays in processing or shipping.

The smooth circulation of tablets between SRO and participating families likely benefited from high compliance in this small convenience sample. Participants' willingness to keep their scheduled appointments was another benefit. A request to reschedule after a shipment was already *en route* could have significantly hampered the capacity to send tablets to other families in advance of their scheduled appointments. Yet because tablets were often tied up in shipment or at a participating family's home until the time of a scheduled appointment, the study would have benefitted from having more than five tablets available even in this small sample. Overall, the pilot study achieved the basic objective of demonstrating the feasibility of the shipment and use of electronic tablets in a household-based study with child respondents. But significantly more testing is required to establish a system suitable for larger, more dispersed, and more population-representative samples.

Data Production

All testing was completed with children in their homes. Thirty-one children completed all subtests. One child (age 6 years) did not complete the Passage Comprehension subtest.

Test administrators described all participants as cooperative, but children varied in their level of engagement with the subtests. Two children asked to skip some questions, and one child required frequent redirection to focus on the test stimuli. In one case, a caregiver tried to intervene to help the child answer a question, and in another the child turned to their caregiver for assistance. In both cases, the test administrator advised the caregiver to allow the child to answer each question on their own and reported no further interference. Test administrators described the effort to focus children's attention and redirect caregivers during the video interview as no more challenging than what they had previously encountered during in-person administration of the same subtests.

It was not always possible to keep study participants separated from other family members during testing. Test administrators reported three instances of temporary distractions when family members passed through the room where children were completing the assessment and one instance where a parent was speaking on the telephone in another room nearby. One related concern is that disruption and distraction could have negatively affected the results for the child being tested. Test administrators' ability to monitor the child's environment through the video connection allowed them to assess whether the conditions made the test infeasible to administer, which never occurred.

Another potential consequence is that children who were in line for the assessments may have seen or overheard portions of a subtest's content and the test taker's responses. In theory, there should be little associated risk. Children of different ages would be expected to complete different segments of each subtest, and test takers (and others around them) do not know whether they have provided a correct response. Still, based on this experience, we advise providing headphones or earbuds to test takers to prevent the test administrator's instruction from being overheard and to block out background noise. A tradeoff is that such a strategy would require the child to remove the audio device should the test administrator and caregiver need to communicate during the test administration.

Test Performance

This pilot study was designed to assess whether video interviewing to administer the WJ-IV to children was operationally feasible rather than to establish mode equivalence compared to in-person administration. We present results from the WJ-IV subtests as descriptive. The key takeaways are that nearly all children completed all subtests (with the exception of one incomplete subtest noted above) without greater difficulty than during in-person administration and the distribution of test scores does not suggest systematic upward or downward bias. We describe three outcomes indicative of test performance: time to subtest completion in minutes, proportion of respondents who had some or a lot of difficulty completing the subtest, and age-standardized scores (Table 2). All statistics are reported overall and by age group (5-8, 9-12, and 13-17 years). For comparison, we report weighted standardized scores and frequency of any testing difficulty among same-age participants in the 2019 PSID Child Development Supplement (CDS). Timings for individual subtests in CDS were not available.

Table 2 Standardized scores, time to completion in minutes, and proportion experiencing some or a lot of difficulty, WJ-IV Letter-Word Identification, Passage Comprehension, and Applied Problems, overall and by age group with comparison to the 2019 Panel Study of Income Dynamics Child Development Supplement (CDS)

Subtest	Metric	Data	Overall ($N = 32$)	5-8 years ($N = 9$)	9-12 years ($N = 12$)	13-17 years ($N = 11$)
			Mean (*SD*) or proportion	Mean (*SD*) or proportion	Mean (*SD*) or proportion	Mean (*SD*) or proportion
Letter-Word Identification	Standardized score	Pilot	101.7 (13.7)	101.3 (13.8)	104.1 (12.2)	99.4 (15.9)
		CDS	100.5 (15.9)	97.3 (18.2)	104.0 (15.4)	100.8 (14.8)
	Time	Pilot	3.9 (2.6)	6.7 (2.9)	3.3 (1.4)	2.2 (0.9)
	Any difficulty	Pilot	0.09	0.11	0.17	0.00
		CDS	0.49	0.58	0.43	0.38
Passage Comprehension	Standardized score	Pilot	97.2 (13.4)	100.5 (10.3)	99.1 (12.0)[a]	92.8 (16.4)
		CDS	94.0 (13.9)	96.1 (14.5)	94.9 (14.0)	91.6 (12.8)
	Time[b]	Pilot	9.8 (3.4)	11.4 (3.9)	10.0 (3.11)	8.2 (2.8)
	Any difficulty	Pilot	0.47	0.67	0.42	0.36
		CDS	0.63	0.75	0.64	0.54
Applied Problems	Standardized score	Pilot	98.6 (18.0)	99.6 (20.2)	96.5 (18.4)	100.0 (17.4)
		CDS	99.0 (17.0)	96.4 (15.4)	100.5 (16.8)	99.3 (17.8)
	Time	Pilot	15.9 (4.5)	13.8 (5.5)	16.0 (3.6)	17.5 (4.1)
	Any difficulty	Pilot	0.34	0.56	0.25	0.27
		CDS	0.66	0.72	0.63	0.56

[a] $N = 11$
[b] Excludes one outlier whose test administration was suspended overnight.
CDS estimates are survey-weighted to represent the population of US children.
Time to module completion is not available for CDS.

The median time to complete all of the assessments was 29.5 minutes, which is similar to in-person administration times in the CDS (Panel Study of Income Dynamics, 2022). The Letter-Word Identification was the shortest (3.9 minutes overall), followed by Passage Comprehension (9.8 minutes)[12] and Applied Problems (15.9 minutes). As noted above, the youngest children required more time than older children to complete the Letter-Word and Passage Comprehension assessments. Older children required the most time to complete the Applied Problems assessment.

Test administrators reported in post-test observations whether a participant had no difficulty, some difficulty, or a lot of difficulty with each subtest. These reports are intended to assess the child's evident cognitive burden when evaluating the test stimuli, but here the items might also have captured difficulty specific to the modality. Fewer than 10% of children had any difficulty with the Letter-Word Identification subtest, but nearly half had at least some difficulty with the Passage Comprehension subtest, and one-third had at least some difficulty with the Applied Problems subtest. The youngest children (aged 5-8 years) had difficulty more often than their older peers on the latter two subtests, while children aged 9-12 years more often had trouble with the Letter-Word Identification subtest (17%). The same age patterns were reported for these subtests when they were administered in person during the 2019 wave of CDS (Panel Study of Income Dynamics, 2022), but the frequency of reported difficulty overall was lower in the pilot study sample, perhaps reflecting its relatively educationally advantaged composition.

For each subtest, the standardized score has an expected mean of 100 and a standard deviation of 15. Overall, the average score on the Letter-Word Identification subtest (101.7) is slightly higher than the expected mean, and the average scores on the Applied Problems (98.6) and Passage Comprehension (97.2) are slightly lower. The magnitude of difference compared to the expected values is consistent with those observed in CDS, and pilot study sample means were not significantly different from the expected mean in a series of one-sample hypothesis tests ($p < .05$). The same pattern holds for each age group except for the Passage Comprehension subtest for adolescents age 13 years and older, where the sample mean (92.8) is about half of a standard deviation below the expected value. This relatively large difference may be attributable to chance, to unobserved characteristics of the sample ($N = 11$), or to some interaction between the subtest content and the video format that disproportionately influenced scores on one subtest for one age group. It is worth noting that the sample mean of the standardized score on this subtest for same-age adolescents in CDS was 91.6 (Table 2), suggesting that the lower-than-expected average performance on this subtest may be attributable to factors other than mode of administration.

12 Statistics exclude one outlier case (age 9 years) for whom administration of the Passage Comprehension subtest was suspended overnight.

Conclusion

The purpose of the pilot study and the objective of this paper were to assess the operational feasibility of video interviewing to administer standardized achievement tests at home with school-aged children. This study represents the first step in a larger program of development and evaluation to inform the adoption of video interviews in large-scale survey research with US children and their families.

We assessed four aspects of video administration. The first two were technical development and implementation of a secure screen-based interface for administering the standardized tests. This effort was largely successful. The integration of a laptop personal computer and external monitor allowed test administrators to observe the test taker's environment, the test stimulus page, and the data collection instrument simultaneously. The annotation function was effective for children with early reading and vocabulary skills, but training in how to use that feature appeared to contribute substantially to test administration time. As an alternative, Riverside Insights, the publisher of the WJ-IV battery of tests, recommends using a mouse or trackpad for test takers to point to their responses. Future work could compare these approaches in terms of their intuitiveness and time requirements.

The use of the screen-based interface relies upon two other aspects of video interviewing that the pilot study assessed: development of a push notification to establish a stable, secure connection between the test administrator and the test taker and protocols to provide cellular-enabled electronic tablets to children in all participating families. The push notification worked as expected. The cellular-enabled internet connection was mostly reliable, but some sessions were dropped due to lost connectivity. In each case, the test administration continued after the connection was re-established, but in a larger sample drawn by other means, participants may be unable or less willing to re-establish a lost connection. The availability of more reliable cellular coverage would improve continuity.

The third aspect was the protocols developed to manage the distribution of electronic tablets with a home-based sample. This poses a potential sticking point for scaling up. In many ways, the pilot study was a best-case scenario because the tablets had relatively little "down time." They were either in shipment, in processing for reuse, or in receipt at the test taker's household for a short time before the scheduled appointment. Further, no tablets were lost, stolen, or damaged, so they all remained in circulation during the short study period. And yet the shipment log and SRO personnel reports suggest that it would have been helpful to have more tablets on hand. Managing similar logistics in a large-scale data collection effort may be untenable or prohibitively costly if the same protocol were established for all participants. However, the approach may be more feasible where video interviews are selectively offered as an option to a subset of participants, such as those who are geographically isolated or strongly resistant to an in-person visit.

Finally, we considered the data production experience. Here we asked whether children as young as 5 years old would be able to focus on the test-taking task at home when they were engaged with a remote test administrator in a videoconferencing session. Our conclusion is that they can. Although test administrators reported occasional background noise or disruption and described some children as less engaged, their overall assessment was that children's attention to the test and apparent ease or difficulty was comparable to what testers had observed during previous in-person administrations. Test administrators offered several suggestions to improve the data production experience: provide a stand for the electronic tablet so that the testing environment remains visible and children may use the tablet hands-free; provide earphones or earbuds to block out distractions and prevent other test takers from overhearing test content; lock down Zoom features to prevent manipulation; and increase training time and practice for test administrators who are new to the video interviewing format.

An important caveat to these conclusions is that the pilot study was conducted with a small convenience sample of families recruited from the personal networks of study staff. These personal connections likely resulted in greater compliance than would be observed in samples drawn by other methods. Further, caregivers in the convenience sample were highly educated compared to the general population. This may have contributed to caregivers' and children's ease with the technology used in this study and may not be generalizable to more population-representative samples. Finally, the small sample size limits the reliability of our conclusions.

Despite these limitations, the results of the current study suggest that video administration is a feasible and attractive alternative when in-person administration is challenging due to high travel costs or in cases where respondents are hard to reach in-person but amenable to remote contact. Future work to establish mode equivalence compared to in-person administration and to identify the circumstances under which video administration is most cost-effective will further inform decision-making about the appropriate use of this innovative approach for standardized testing with children in survey research.

Funding

The research described herein was supported by the *Eunice Kennedy Shriver* National Institute for Child Health and Human Development (R01HD104671-01S1, PI: Sastry).

References

Bates, N., Dahlhamer, J., & Singer, E. (2008). Privacy concerns, too busy, or just not interested: Using doorstep concerns to predict survey nonresponse. *Journal of Official Statistics, 24*(4), 591.

Brick, J. M., & Williams, D. (2012). Explaining Rising Nonresponse Rates in Cross-Sectional Surveys. *The ANNALS of the American Academy of Political and Social Science,* 645(1), 36–59. https://doi.org/10.1177/0002716212456834

Button, A. M., Webster, E. K., Kracht, C. L., Hendrick, C., Okely, A., Chong, K. H., Cross, P., & Staiano, A. E. (2023). Validation of remote assessment of preschool children's anthropometrics and motor skills. *Frontiers in Digital Health, 5,* 1168618. https://doi.org/10.3389/fdgth.2023.1168618

Centeno, L., Arrue, J., Good, C., Edwards, B., & Steiger, D. (2024). Practical considerations for developing and implementing a video mode for survey data collection. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique, 164*(1), 122–146. https://doi.org/10.1177/07591063241277065

Conrad, F. G., Schober, M. F., Antoun, C., Yan, H. Y., Hupp, A. L., Johnston, M., Ehlen, P., Vickers, L., & Zhang, C. (2017). Respondent mode choice in a smartphone survey. *Public Opinion Quarterly, 81*(S1), 307–337. https://doi.org/10.1093/poq/nfw097

Federal Communications Commission. (2017). *Implementation of Section 6002(b) of the Omnibus Budget Reconciliation Act of 1993, Annual Report and Analysis of Competitive Market Conditions with Respect to Mobile Wireless, Including Commercial Mobile Services* (FCC-17-126). https://www.fcc.gov/document/fcc-releases-20th-wireless-competition-report-0

Feuer, S., & Schober, M. F. (2026). Effects of the self-view window in live video survey interviews. *Methods, Data, Analyses, 20*(1), 124–178. https://doi.org/10.12758/mda.2025.07

Ghosh-Dastidar, M., Nicosia, N., & Datar, A. (2020). A novel approach to anthropometric assessment for geographically dispersed samples: A pilot study. *Preventive Medicine Reports, 19,* 101125. https://doi.org/10.1016/j.pmedr.2020.101125

Hanson, T., Briceno-Rosas, R., Ainsaar, M., & Kartau, H. K. (2026). Live online video interviewing as a complementary mode to in-person interviews: Evidence from the European Social Survey. *Methods, Data, Analyses, 20*(1), 179–208. https://doi.org/10.25521/mda.663

Mather, N., & Jaffe, L. E. (2016). *Woodcock-Johnson IV: Reports, recommendations, and strategies.* (pp. ix, 590). John Wiley & Sons, Inc.

McGonagle, K. A., & Sastry, N. (2015). Cohort Profile: The Panel Study of Income Dynamics' Child Development Supplement and Transition into Adulthood Study. *International Journal of Epidemiology, 44*(2), 415–422. https://doi.org/10.1093/ije/dyu076

Organisation for Economic Cooperation and Development. (2020). *Early Learning and Child Well-being.* https://www.oecd.org/content/dam/oecd/en/publications/reports/2020/03/early-learning-and-child-well-being_74042660/3990407f-en.pdf

Panel Study of Income Dynamics. (2022). *Child Development Supplement to the Panel Study of Income Dynamics, public use dataset* (Survey & I. for S. R. Research Center University of Michigan, Eds.). Survey Research Center, Institute for Social Research, University of Michigan.

Riverside Insights. (2020). *Tips for Remote Assessment Using Riverside Insights Clinical Products*. https://info.riversideinsights.com/hubfs/Remote_Assessment%20ProtocolTSdocx-MAY7.pdf

Ruggles, S., Flood, S., Sobek, M., Backman, D., Cooper, G., Rivera Drew, J. A., Richards, S., Rodgers, R., Schroeder, J., & Williams, K. C. W. (2025). *IPUMS USA: Version 16.0* [Dataset]. IPUMS. https://doi.org/10.18128/D010.V16.0

Sastry, N., Fomby, P., & McGonagle, K. A. (2021). Effects on Panel Attrition and Fieldwork Outcomes from Selection for a Supplemental Study: Evidence from the Panel Study of Income Dynamics. In P. Lynn (Ed.), *Advances in Longitudinal Survey Methodology* (pp. 74–99). Wiley. https://doi.org/10.1002/9781119376965.ch4

Schober, M. F. (2018). The future of face-to-face interviewing. *Quality Assurance in Education, 26*(2), 290–302. https://doi.org/10.1108/QAE-06-2017-0033

Tsuji, S., Amso, D., Cusack, R., Kirkham, N., & Oakes, L. M. (2022). Empirical research at a distance: New methods for developmental science. *Frontiers in Psychology, 13*, 938995. https://doi.org/10.3389/fpsyg.2022.938995

Vercruyssen, A., Van de Putte, B., & Stoop, I. A. (2011). Are they really too busy for survey participation? The evolution of busyness and busyness claims in Flanders. *Journal of Official Statistics, 27*(4), 619–632.

Williams, D., Brick, J. M., Williams, D., & Brick, J. M. (2018). Trends in U.S. Face-To-Face Household Survey Nonresponse and Level of Effort. *Journal of Survey Statistics and Methodology, 6*(2), 186–211. https://doi.org/10.1093/jssam/smx019

Wright, A. J. (2018). Equivalence of Remote, Online Administration and Traditional, Face-to-Face Administration of Woodcock-Johnson IV Cognitive and Achievement Tests. *Archives of Assessment Psychology, 8*(1), 23–35.

methods, data, analyses | Vol. 20(1), 2026, pp. 108-123 DOI: 10.12758/mda.2025.06

Video Interviewing and Observed Differences in Mental Health Outcomes

Christine Carr, Heidi Guyer, Katelan McDaniel, Wendy Reed, Carlos Macuada, Paul Geiger, Lisa Carpenter, Andy Peytchev, Jill Dever, R. Suresh & Mark Edlund

RTI International

Abstract

The integration of video interviewing in survey research is relatively new and may offer similar benefits as telehealth visits in mental health research. Methodological evaluations of video interviewing are needed for large-scale surveys. Over 3,000 clinical interviews were conducted by video and over 1,500 by phone for a national study of U.S. adults, the Mental and Substance Use Disorders Prevalence Study (MDPS). Sociodemographic differences were observed among those who completed a clinical interview by video compared to phone respondents. Higher prevalence rates of all disorders, with the exception of schizophrenia spectrum disorders, were found for video respondents. Higher prevalence rates of generalized anxiety disorder (video 16.2% vs. 14.1%, $p < .05$), bipolar 1 (2.1% vs. 0.7%, $p < .05$) and obsessive compulsive disorder (OCD; 3.1% vs. 1.5%, $p < .05$) were observed among those completing an interview by video compared to those interviewed by phone. Individual logistic regression models were calculated for each disorder adjusting for sociodemographic characteristics to assess the difference in prevalence rate by mode of interview. Respondents interviewed by video had higher odds of having bipolar 1 (OR = 2.96, 95% CI [1.42, 6.17]), OCD (OR = 2.16, 95% CI [1.20, 3.90])and having two or more mental health disorders (OR = 1.64, 95% CI [1.23, 2.19]) than those interviewed by phone after adjusting for sociodemographic characteristics. While further investigation using experimental approaches is required, video interviewing may improve the ability to detect mental health conditions in large-scale survey research.

Keywords: video interviewing, mental health, prevalence, data collection, national survey, clinical interview

© The Author(s) 2025. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Because the use of video interviewing is new to large-scale survey data collection, it lacks the methodical study and evaluation that has been applied to other modes of data collection. The COVID-19 pandemic changed video interviewing from an option to a necessity for many surveys. Survey researchers and methodologists have used video interviewing as an alternative or supplement to in-person interviewing (Anderson, 2008; Jeannis et al., 2013), but its successful use during the COVID-19 pandemic proved it to be a viable mode for collecting data while face-to-face interviewing was not an option (Joshi et al., 2020; McClain et al., 2021; Thunberg & Arnell, 2021).

The Mental and Substance Use Disorders Prevalence Study (MDPS) is a national epidemiological survey of U.S. adults conducted between 2020 and 2022 designed to estimate the prevalence of serious mental and substance use disorders among adults in the United States. This study is the largest implementation of video interviewing in a national probability-based survey, as the pandemic started just a few months before data collection was scheduled to begin. Video interviewing was planned as an option, but it became the main mode for clinical interviews.

In many ways, video interviewing mimics telehealth visits, which also increased during the pandemic and have continued to be widely used for medical visits (Shaver, 2022). Telehealth visits and assessments can be especially beneficial for health conditions in which a visual assessment is critical to the overall diagnosis, such as mental health assessments. Telehealth and video interviewing allow for the assessment of behaviors, physical appearance, and potential physical manifestations commonly exhibited with some mental health conditions. Examples include motor activity (psychomotor agitation), grooming (dishevelment), and affect (a limited affect is sometimes seen in individuals with schizophrenia). There is substantial evidence that such assessments, including the *Structured Clinical Interview for the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (SCID-5®; First et al., 2015), a semi-structured clinical inter-

Acknowledgements

The Mental Health and Substance Use Disorders Prevalence Study (MDPS) was a cooperative agreement between RTI International and the Substance Abuse and Mental Health Services Administration (SAMHSA). SAMHSA provided funding for the project (H79FG000030).

Notes

The views and opinions contained in this presentation do not necessarily reflect those of SAMHSA or the U.S. Department of Health and Human Services and should not be construed as such.

Direct correspondence to

Christine Carr, RTI International, Research Triangle Park, North Carolina, USA

E-mail: ccarr@rti.org

view for psychiatric diagnosis, have equivalent reliability and diagnostic accuracy in video-mediated and in-person assessments (Chakrabarti, 2015). In addition, it has been found that assessment via videoconferencing is well accepted among patients with schizophrenia (Kasckow et al., 2014) and other psychoses (Sharp et al., 2011) as well as by members of hard-to-reach populations (Durrant & Hanson, 2023; Gray et al., 2020; Miller & Gibson, 2004; Schober et al., 2023). However, not all individuals are able to participate in a video interview. Some may not have a smartphone, tablet, or computer; others may not have internet access; and some may not understand how to use the technology required for a video-based interview.

In this paper, we compare outcomes from video and phone interviews conducted in the MDPS, including differences in demographic characteristics and disorder-specific prevalence estimates for those completing a clinical interview by video. Due to the COVID-19 pandemic, close to seventy percent of clinical interviews were conducted by video and one-third by phone as in-person data collection was prohibited during much of the study period. Respondents were encouraged to participate in video-based interviews. However, for respondents who could not or would not participate in this mode, interviews were conducted by phone. Although a formal experiment was not conducted, the MDPS is the largest study to date to implement video interviewing in a national survey offering much needed information on the large-scale feasibility of this mode as well as a comparison of the demographic characteristics and the outcomes of interest attributed to video interviewing.

Methods

Study Design

The MDPS recruited participants from national probability-based samples of households and prisons and from convenience samples of state psychiatric hospitals and homeless shelters (Guyer et al., 2024; Ringeisen et al., 2023), the only probability-based study on serious mental health disorders that includes clinical interviews nationally in the United States. Interviews were conducted with 5,679 participants (4,764 household, 321 prison, 425 homeless shelter, 171 state psychiatric hospital); this paper concerns only the household data as other factors influenced the mode of interview within facilities. Data collection was conducted from October 2020 to July 2022. The MDPS was funded by a grant from the Substance Abuse and Mental Health Services Administration, an agency of the U.S. Department of Health and Human Services. The study protocol and informed consent were reviewed and approved by an external Institutional Review Board external to the research organization and funding agency. The study methodology and preliminary findings have been published previously (Bareis, et al., 2024; Chwastiak, et al., 2025; Gibbons, et al., 2024; Guyer et al., 2024; Ringeisen et al., 2023), as well as the technical and qualitative outcomes associated with video interviewing (Jayaram et al., 2023b).

The MDPS employed a three-stage design. First, each household in a national probability sample of households was invited to complete a household roster. Next, up to two randomly chosen age-eligible residents in each household were invited to complete a mental health screener. Following the completion of the screener, respondents were categorized based on their likelihood of having a serious mental health disorder (e.g., schizophrenia spectrum disorder) and sampled differentially to complete a clinical interview. Individuals at greater risk of disorders were oversampled as follows: 100% of individuals endorsing psychotic symptoms, 80% of individuals endorsing symptoms of another MDPS mental or substance use disorder, and 20% of all others. The goal of this approach was to increase the number of participants with the rarest disorders to improve the precision and accuracy of prevalence estimates for these disorders. A total of 234,270 households were invited to complete the initial household roster and 4,764 respondents completed a clinical interview. The household roster and the screener could be completed in person or by web, mail, or telephone. Approximately 61% of respondents completed the roster via web, and this substantial reliance on web completion likely aided the video interviewing mode for the third stage, the clinical interview. Mail did not provide a reasonable number of completed roster and screener surveys, but the other two modes did.

Interview Mode

Clinical interviews were conducted by video or phone. Clinical interviewers contacted respondents selected for the clinical interview by email or phone (using an Interactive Intelligence Client Application on their laptop which masked their phone) and requested that they conduct the interview via video. Respondents who indicated they could not participate via video or would be more comfortable with a non-video interview were interviewed via phone (after the interviewer attempted to address technical or other issues raised by the respondent). Video interviews were conducted using Zoom. Clinical interviewers used a laptop and tablet computers to log into the scheduled Zoom meeting and conduct the video-based personal interviews. Zoom was an ideal platform because respondents could join the interview session through a browser without needing to download an app.

Before beginning the interview, the clinical interviewer identified a private location from which they could conduct the interview. This location was separate from shared living or working areas so others could not overhear. Clinical interviewers were required to have a neutral background (i.e., free of political or religious imagery and other personal items) to avoid biasing the respondent or making them uncomfortable. Clinical interviewers also asked the respondent to find a private location to complete the interview so they would feel comfortable answering the questions honestly and to protect confidentiality. Clinical interviewers used their project tablet to launch and record the Zoom video and used the laptop to administer the clinical interview and record clinical notes. Respondents could join using any device that had a camera, including a smartphone, tablet, laptop, or desktop computer. At the beginning of the interview, clinical interviewers asked respondents for permission to record the interview. If a respondent did not give consent to record, the clinical interviewer stopped all recording but continued the interview. During the interview, the clinical interviewer would occasionally screen-share showcards to display a list of response options. If the interview was completed via phone, the clinical interviewer would direct the respondent to the showcards emailed to them or to the project website where the showcards were available.

Clinical Interview

The MDPS was designed to estimate the past-year prevalence of common and rare mental health disorders, including generalized anxiety disorder (GAD), major depressive disorder (MDD), obsessive compulsive disorder (OCD), post-traumatic stress disorder (PTSD), anorexia nervosa, bipolar 1, and schizophrenia spectrum disorders (SSDs). The clinical interview consisted of basic demographic questions; the NetSCID, a computerized version of the MDPS SCID-5 clinical interview (Brodey et al., 2016), questions on mental health treatment, an assessment of the severity of the disorder, and questions on the impact of the COVID-19 pandemic on mental health status and treatment.

The SCID-5 (First et al., 2015) is designed to be delivered by trained clinicians (e.g., social workers, psychologists, psychiatrists) who are experienced with diagnostic interviewing and the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5). Minimum qualifications for MDPS clinical interviewers included a graduate degree and training in a relevant field (e.g., psychology, psychiatry, social work) and clinical experience administering the SCID-5. All clinical interviewers and clinical supervisors completed extensive project training and certification. Clinical interviewers were trained to read the question on the laptop screen and to then view the respondent as they answered. Various quality control measures were conducted throughout data collection, including standardized reviews of recorded interviews, inter-rater reliability exercises with the clinical interviewers, and data quality checks. Quality control measures have been described elsewhere (Guyer et al., 2024; Ringeisen et al., 2023).

Statistical Analysis

Analyses were conducted to explore differences in demographic characteristics and disorder-specific prevalence estimates by mode of interview. First, the demographic characteristics of those completing the clinical interview by video vs. phone were compared. Demographic characteristics were assessed and included age group, sex, race/ethnicity, education, region, and urbanicity. Next, the weighted prevalence estimate of each disorder was compared for those completing the interview by video vs. by phone. The methods for calculating the MDPS prevalence estimates have been described in detail elsewhere (Guyer et al., 2024; Ringeisen et al., 2023). Finally, separate logistic regression models were calculated for each mental health disorder, with the disorder of interest as the dependent variable (yes/no) and mode as the independent variable (video compared with phone) along with the demographic covariates. This allowed for comparisons of key disorder outcomes by mode, after controlling for any demographic differences. Descriptive statistics and chi-squared tests were conducted to explore the differences by mode in respondent demographic characteristics. Unweighted counts and weighted percentages are presented; analyses were weighted for the complex sample design. Odds ratios and 95% confidence intervals (CIs) were calculated using logistic regression. All analyses were completed using SAS v9.4.

Results

Interview Mode and Mental Health Outcomes

The demographic characteristics for those who completed the clinical interview by phone and by video are shown in Table 1. Compared to those interviewed by phone, respondents who agreed to complete the interview by video were more likely to be younger (44 years or younger vs. 45 years or older), identify as non-Hispanic White or Asian rather than Hispanic or non-Hispanic Black, have a college education, have a higher income ($75,000 or higher), and live in an urban area.

Table 2 shows the disorder-specific prevalence estimates overall and by mode of interview. Higher prevalence estimates were found for those interviewed by video for all disorders except past-year and lifetime SSDs. Significantly higher prevalence rates of GAD (16.2% vs. 14.1%, $p < .05$), bipolar 1 (2.1% vs. 0.7%, $p < .05$), and OCD (3.1% vs. 1.5%, $p < .05$) were found among those completing a video interview compared to those interviewed by phone. Additionally, a 60% increase in the prevalence of having two or more MDPS mental health disorders was observed among those completing a video interview (9.3% vs. 5.8%, $p < .001$).

Table 1 Sociodemographic characteristics, by mode of clinical interview

Respondent characteristics	Video (n = 3,178)		Phone (n = 1,586)	
	N	%	*N*	%
Age*				
18–25	433	19.9	177	12.2
26–44	1,575	45.6	641	36.6
45–65	1,170	34.5	768	51.2
Sex				
Male	1,200	48.6	604	49.4
Female	1,978	51.4	982	50.6
Race/ethnicity*				
Hispanic or Latino	392	13.9	320	25.1
White, not Hispanic	2,189	65.2	897	51.6
Black or African American, not Hispanic	241	10.3	216	15.8
Asian, not Hispanic	197	5.8	65	2.3
American Indian or Alaska Native, not Hispanic	23	0.4	14	0.5
Native Hawaiian or other Pacific Islander, not Hispanic	5	0.2	5	0.1
Multi-racial, not Hispanic	129	4.2	69	4.6
Education*				
Less than a high school diploma	67	6.5	121	16.1
High school degree or equivalent	314	20.7	332	35.2
Some college or associate’s degree	845	31.1	534	32.1
Bachelor’s degree	1,057	27.3	334	10.2
Post-college/professional degree	895	14.4	265	6.4
Urbanicity				
Urban	2,768	84.1	1,324	79.0
Rural	410	15.9	262	21.0
Income*				
Less than $20,000	350	9.7	369	24.6
$20,000–$49,999	620	23.6	385	24.4
$50,000–$74,999	510	13.0	219	14.7
$75,000–$99,999	433	12.0	158	7.4
$100,000–$149,999	508	15.5	154	7.5
$150,000 or more	600	19.0	156	8.6
Missing	157	7.2	145	12.8

Notes: *$p < .001$; all percentages are weighted for the complex sample design.

Table 2 Mental health disorder prevalence estimates, by mode

MDPS disorders (past-year unless otherwise indicated)	All (n = 4,764)		Video (n = 3,178)		Phone (n = 1,586)		
	N	%	N	%	N	%	p-value
Mental health disorders							
Schizophrenia or schizoaffective disorder	73	1.2	32	0.9	41	1.6	.124
Schizophrenia or schizoaffective disorder (lifetime)	114	1.8	50	1.5	64	2.2	.265
Major depressive disorder	1,213	15.4	846	16.2	367	14.1	.317
Generalized anxiety disorder	855	10.0	623	11.3	232	8.0	.042
Bipolar 1	127	1.5	91	2.1	36	0.7	.024
Posttraumatic stress disorder	320	4.0	218	4.4	102	3.4	.326
Obsessive compulsive disorder	209	2.5	154	3.1	55	1.5	.026
Anorexia nervosa	15	0.1	11	0.2	4	0.1	.168
Any MDPS mental health disorder	1,884	24.9	1,293	26.7	591	22.3	.096
Two or more MDPS mental health disorders	731	7.9	525	9.3	206	5.8	<.001

Notes: All percentages are weighted for the complex sample design.

Disorder-specific odds ratios were calculated to further explore the difference in prevalence estimates by mode of interview (Table 3). After adjusting for demographic characteristics, those who completed an interview by video (vs. phone) had significantly higher odds of meeting criteria for bipolar 1 (Odds ratio (OR) = 2.96, 95% CI [1.42, 6.17]) and of meeting criteria for OCD (OR = 2.16; 95% CI [1.20, 3.90]). Additionally, those completing an interview by video were significantly more likely to meet criteria for two or more MDPS mental health disorders than those interviewed by phone, after adjusting for demographic characteristics (OR = 1.64; 95% CI [1.23, 2.19]). Appendix Table 1 provides the full set of outcomes for each disorder-specific logistic regression model including the odds ratios for the sociodemographic characteristics included in each model.

Table 3 Association of interview mode with each MDPS mental health disorder

Mental health disorder model (past-year unless otherwise indicated)	Adjusted odds ratios*	Lower CI limit	Upper CI limit	*p*-value
Schizophrenia spectrum disorders (SSDs)	0.80	0.31	2.06	.642
SSDs (lifetime)	1.07	0.48	2.39	.868
Major depressive disorder (MDD)	1.03	0.75	1.40	.876
Generalized anxiety disorder (GAD)	1.17	0.84	1.65	.350
Bipolar 1	2.96	1.42	6.17	.004
Posttraumatic stress disorder (PTSD)	1.35	0.82	2.23	.232
Obsessive compulsive disorder (OCD)	2.16	1.20	3.90	.011
Any MDPS mental health disorder	1.12	0.85	1.47	.418
Two or more MDPS mental health disorders	1.64	1.23	2.19	.001

Notes: CI = confidence interval.
* Odds ratios for each health disorder-specific model were adjusted for age group, sex, race/ethnicity, education, income, region, and urbanicity. Odds ratios are shown for video interviewing mode with phone mode as the reference category. Higher odds ratios indicate increased likelihood of participating in video interview.

Respondent and Interviewer Experience in Video Interviews

Additionally, clinical interviewers provided positive feedback about video interviews, with approximately 81% reporting that the functionality of the video interviewing process was excellent, 91% of respondents did not have technical difficulties with Zoom, and 93% did not get disconnected during the interview. Clinical interviewers also indicated that 91% of the interviews had very good/good visual quality and visual observations were used in making diagnoses in about 71% of cases where the respondent had at least one mental health disorder. Respondent feedback on video interviewing was positive as well; 90% of video respondents indicated they were comfortable completing the interview by video (which was comparable with other modes), and 94% indicated they were very comfortable or comfortable using Zoom. Similar positive feedback was found in studies conducted by the Australian Bureau of Statistics (Phillips et al., 2023).

Discussion

The MDPS experience provides the opportunity to better understand the potential differences the relatively new mode of video interviewing may introduce.

MDPS respondents interviewed by video tended to be younger, identify as a non-minority race/ethnicity, and have higher levels of education and income than those interviewed by phone. Significantly higher prevalence estimates of GAD, bipolar 1, OCD, and two or more MDPS mental health disorders were found in those interviewed by video compared to those interviewed by phone. To account for differences in the non-random assignment to mode, estimates were examined after adjusting for the sociodemographic characteristics: higher rates of these disorders remained significant among those interviewed by video with the exception of GAD. Respondents interviewed by video were almost 3 times more likely to meet criteria for bipolar 1, more than twice as likely to meet criteria for OCD, and almost 2 times more likely to meet criteria for two or more MDPS mental health disorders than those interviewed by phone. Consequently, differences by mode found in our study are not simply an artifact of the differences in sample composition between modes. The video mode likely helps with identification of serious mental illnesses as interviewers are able to observe the respondent's behavior, compared to telephone. Finding higher rates of diagnoses in the video mode after controlling for demographic characteristics helps to support this finding. However, this was not an experimental design and it is possible that those who are interviewed by telephone are also less likely to have serious mental illnesses.

Video interviewing enabled clinical interviewers to assess the behaviors, physical appearance, and potential physical manifestations commonly exhibited with some of the disorders of interest. It was initially planned to detect typical symptoms of SSDs that can be observed visually by trained clinicians and to accommodate potential respondents with paranoia, which is another common symptom among those with SSDs. However, the odds of meeting the criteria for SSD were no higher in video than phone interviews. Nonetheless, the odds of meeting the diagnostic criteria were higher for those who completed an interview by video for disorders whose detection was not thought, a priori, to rely on visual observation. Although this finding may be somewhat surprising from a clinical observation perspective, it is not surprising when comparing interviews in which the interviewer can see the respondent's face (i.e., video) to phone interviews. Highly skilled and extensively trained interviewers observed (and took into account) respondent behaviors during the average 77-minute video-based clinical interview that were helpful in making these diagnoses (e.g., rearranging objects to ensure a specific order or symmetry, restlessness). It could also be that video interviewing enabled development of a stronger rapport between respondent and interviewer, which could have resulted in the respondent feeling more comfortable disclosing information. Video interviewing appears to be most beneficial in detecting the less impairing mood and mania disorders compared to SSDs.

The inclusion of the telephone mode helped to bring in more respondents from underrepresented groups than the video mode: Hispanic, non-Hispanic Black respondents, those with a high school education or less, and those with lower income. However, if the lower prevalence rates in the telephone mode are the result of lower levels of rapport between interviewers and respondents and because the telephone interviewers cannot observe respondents visually, then including the phone interview data in the overall prevalence estimates may introduce measurement error despite its reducing nonresponse bias among certain subgroups. This is a tradeoff researchers will have to navigate.

A warranted concern with video interviewing is whether some respondents are less likely to participate, increasing coverage error for some subgroups: this mode requires internet access, a video-capable device, willingness to participate in a video interview, and a private location. Such mode differences in coverage could introduce bias into the estimates. We found that those who completed the clinical interview by video were more likely to be younger, to be non-Hispanic White or Asian, and to have a higher socioeconomic status (education and income). These differences were statistically significant, whereas significant differences were not observed by sex or urbanicity. Dulaney et al. (2023) reported similar findings among video respondents for the Medical Expenditure Panel Survey: video respondents tended to be younger, college educated, female and White. Higher response to video-based interviewing compared to in-person interviewing among those with higher socioeconomic status has also been observed on the European Social Survey (Hanson et al., 2023; Martin & Fradier, 2023; Thórólfsson et al., 2023). These analyses highlight the importance of considering under-representation when video is the only mode offered. Additionally, these analyses highlight the benefits of offering another mode, such as phone, to improve representation and coverage.

Data quality is an important consideration when introducing a new mode of data collection. Previous studies have reported that video interviewing has positively affected—or at least has not harmed—data quality compared to in-person interviews, possibly due to the opportunity for the respondent to develop a similar level of rapport with the interviewer in both interview modes (Sanchez et al., 2023; Sun et al., 2021). In comparison of the quality of video interviews and in-person interviews, Kelley et al. (2023) found no strong evidence of measurement effects introduced by video. But comparisons of data quality between video interviews and data collection in other modes are relatively rare. One relevant study was conducted by Conrad et al. (2023) who compared Live Video Interviewing to Web Surveys and Prerecorded Video interviews. They reported that Live Video interviewing produced higher quality in some areas (e.g., less non-differentiation) but lower data quality in other areas (e.g., more rounded numerical responding, more socially desirable responding, i.e., fewer sensitive responses).

The MDPS was designed as a pilot, or feasibility study, to determine national prevalence estimates of serious mental health and substance use disorders among U.S. adults. Many novel aspects of the study are true strengths, especially its use of video interviewing: a national sample of households were included in the study, providing a broad range of generalizable findings; video interviewing provided increased scheduling flexibility for both interviewers and respondents; and video interviewing allowed the study to be conducted during the COVID-19 pandemic rather than halting all data collection efforts. However, there are several limitations of this study as a vehicle for studying and evaluating video interviewing. First, the MDPS was not designed to evaluate mode differences at any stage, specifically not for a comparison of clinical interviews conducted by video with those conducted by phone. Respondents were not randomly assigned to the two modes which led to differently composed samples in the two modes. Second, the MDPS was conducted during the COVID-19 pandemic, which likely influenced study outcomes. For example, had in-person data collection been available, fewer sample members might have elected to participate in video interviews. Finally, video interviewing as an interview mode was available only to households with internet access, those with access to a phone or computer, and those who were comfortable using the video software. Sample members without these characteristics could well have exhibited mental health and substance use disorders at different rates than those who were able to participate by video.

Video interviewing offers many of the benefits of in-person interviewing, such as respondents developing rapport with the interviewer and enabling interviewers to incorporate a visual appraisal of the respondent and their surroundings into their clinical assessment, at a lower cost than with in-person observation. The substantial increase in video interviewing during the pandemic led to rapid integration with little time for thorough investigation of the mode differences prior to beginning data collection. However, many lessons were rapidly learned. Although video interviewing was conducted successfully on the MDPS, our results highlight the need for controlled experiments to estimate any mode effects due to video interviewing. Further investigations of the impact of video interviewing on participation, data quality, and estimates of substantive outcomes will help inform researchers' choice of data collection mode when the options include video interviewing.

References

Anderson, A. H. (2008). Video-mediated interactions and surveys. In F. G. Conrad & M. F. Schober (Eds.), *Envisioning the survey interview of the future* (pp. 95–118). Wiley. https://doi.org/https://doi.org/10.1002/9780470183373.ch5

Bareis, N., Olfson, M., Dixon, L. B., Chwastiak, L., Monroe-Devita, M., Kessler, R. C., Gibbons, R. D., Edlund, M., Guyer, H., Kreski, N. T., Graupensperger, S., Winans, K. S., & Stroup, T. S. (2024). Clinical characteristics and functioning of adults with bipolar I disorder: Evidence from the mental and substance use disorders prevalence study. *Journal of Affective Disorders, 366*, 317–325. https://doi.org/10.1016/j.jad.2024.08.133

Brodey, B. B., First, M., Linthicum, J., Haman, K., Sasiela, J. W., & Ayer, D. (2016). Validation of the NetSCID: An automated web-based adaptive version of the SCID. *Comprehensive Psychiatry, 66*, 67–70. https://doi.org/10.1016/j.comppsych.2015.10.005

Chakrabarti, S. (2015). Usefulness of telepsychiatry: A critical evaluation of videoconferencing-based approaches. *World Journal of Psychiatry, 5*(3), 286–304. https://doi.org/10.5498/wjp.v5.i3.286

Chwastiak, L., Graupensperger, S., Ringeisen, H., Edlund, M., Guyer, H., Bareis, N., Monroe-Devita, M., Dixon, L. B., Stroup, T. S., Swanson, J., Swartz, M., Hancq, E., Gibbons, R., Kessler, R. C., & Olfson, M. (2025). Impact of insurance type on outpatient mental health treatment of US adults. *PLOS Mental Health, 2*(5), 1–15. https://doi.org/10.1371/journal.pmen.0000299

Conrad, F. G., Schober, M. F., Hupp, A. L., West, B. T., Larsen, K. M., Ong, A. R., & Wang, T. (2023). Video in survey interviews: Effects on data quality and respondent experience. *Methods, Data, Analysis, 17*(2), 135–170. https://doi.org/10.12758/mda.2022.13

Dulaney, R., Kelley, J., Arrue, J., & Edwards, B. (2023, July). *Who is choosing video interviewing? A look into respondent characteristics of video adopters* [Conference presentation]. European Survey Research Association Conference, Milan, Italy.

Durrant, G., & Hanson, T. (2023, July). *Video-assisted personal interviewing: Opportunities and challenges* [Conference presentation]. European Survey Research Association Conference, Milan, Italy.

First, M. B., Williams, J. B. W., Karg, R. S., & Spitzer, R. L. (2015). *Structured clinical interview for DSM-5—Research version (SCID-5 for DSM-5, research version; SCID-5-RV) 1–94*. American Psychiatric Association.

Gibbons, R. D., Olfson, M., Saulsberry, L., Edlund, M. J., Zangeneh, S., Bareis, N., Chwastiak, L., Gibbons, J. B., & Kessler, R. C. (2024). Social vulnerability and prevalence and treatment for mental health and substance use disorders. *JAMA Psychiatry, 81*(10), 976–984. https://doi.org/10.1001/jamapsychiatry.2024.1870

Gray, L., Wong-Wylie, G., Rempel, G., & Cook, K. (2020). Expanding qualitative research interviewing strategies: Zoom video communications. *The Qualitative Report, 25*(5), 1292–1301. https://doi.org/10.46743/2160-3715/2020.4212

Guyer, H., Ringeisen, H., Dever, J., Liao, D., Peytchev, A., Carr, C., Geiger, P., Stambaugh, L., Smith, T., Dixon, L., Olfson, M., First, M., Stroup, S., Chwastiak, L., Monroe-Devita, M., Swanson, J., Swartz, M., Kessler, R. C., Gibbons, R., … the MDPS Consortium. (2024). Mental and Substance Use Disorders Prevalence Study: Background and methods. *International Journal of Methods in Psychiatric Research,* 33(1), Article e2000. https://doi.org/10.1002/mpr.2000

Hanson, R., Briceno-Rosas, R., & Mitchell, J. (2023, July). *Video interviewing on the European Social Survey* [Conference presentation]. European Survey Research Association Conference, Milan, Italy.

Jayaram, P., Filippenko, L. V., Spain, C. J., Check, M. L., Reed, W., Carr, C. M., Guyer, H., & Suresh, R. (2023, October). *Video interviewing: An optimal solution for a national behavioral health survey* [Paper presentation]. International Blaise Users Conference, Durham, NC. https://blaiseusers.org/2023/papers/5_2.pdf

Jeannis, M., Terry, T., Heman-Ackah, R., & Price, M. (2013). Video interviewing: An exploration of the feasibility as a mode of survey application. *Survey Practice, 6*(1), 1–5. https://doi.org/10.29115/sp-2013-0001

Joshi, A., Bloom, D. A., Spencer, A., Gaetke-Udager, K., & Cohan, R. H. (2020). Video interviewing: A review and recommendations for implementation in the era of COVID-19 and beyond. *Academic Radiology, 27*(9), 1316–1322. https://doi.org/10.1016/j.acra.2020.05.020

Kasckow, J., Felmet, K., Appelt, C., Thompson, R., Rotondi, A., & Haas, G. (2014). Telepsychiatry in the assessment and treatment of schizophrenia. *Clinical Schizophrenia & Related Psychoses, 8*(1), 21–27A. https://doi.org/10.3371/csrp.kafe.021513

Kelley, J., Edwards, B., Dulaney, R., & Arrue, J. (2023, July). *Evaluating potential mode effects in video interviewing* [Conference presentation]. European Survey Research Association Conference, Milan, Italy.

Martin, L., & Fradier, Y. (2023, July). *CAVIsio ESS response* [Conference presentation]. European Survey Research Association Conference, Milan, Italy.

McClain, C., Vogels, E. A., Perrin, A., Sechopoulos, S., & Rainie, L. (2021). *The internet and the pandemic*. Pew Research Center. https://www.pewresearch.org/internet/2021/09/01/the-internet-and-the-pandemic/

Miller, R., & Gibson, A. M. (2004). Supervision by videoconference with rural probationary psychologists. *International Journal of Innovation in Science and Mathematics Education, 11*(1), 22–28. https://www.proquest.com/docview/2248418855?pq-origsite=gscholar&fromopenview=true&sourcetype=Scholarly%20Journals

Phillips, B., Neiger, D., Slamowicz, S., Lester, G., Luddon, S., Farrell, E., Gerlach, K., & Carmo, P. (2023, July). *Video-assisted live interviewing in comparison to other survey methods in Australia* [Conference presentation]. European Survey Research Association Conference, Milan, Italy.

Ringeisen, H., Edlund, M. J., Guyer, H., Geiger, P., Stambaugh, L., Dever, J., Liao, D., Carr, C., Peytchev, A., Reed, W., McDaniel, K., & Smith, T. (2023). *Mental and Substance Use Disorders Prevalence Study (MDPS): Findings report*. RTI International. https://www.rti.org/publication/mental-substance-use-disorders-prevalence-study-findings-report/fulltext.pdf

Sanchez, C., Brown, M., Taylor, K., Spencer, S., Englefield, L., & Pun, B. (2023, July). *Experience of video interviewing in two UK national cohort studies* [Conference presentation]. European Survey Research Association Conference, Milan, Italy.

Schober, M. F., Conrad, F. G., Hupp, A. L., Larsen, K. M., Ong, A. R., & West, B. T. (2020). Design considerations for live video survey interviews. *Survey Practice, 13*(1). https://doi.org/10.29115/sp-2020-0014

Schober, M. F., Okon, S., Conrad, F. G., Hupp, A. L., Ong, A. R., & Larsen, K. M. (2023). Predictors of willingness to participate in survey interviews conducted by live video. *Technology, Mind, and Behavior, 4*(2). https://doi.org/10.1037/tmb0000100

Sharp, I. R., Kobak, K. A., & Osman, D. A. (2011). The use of videoconferencing with patients with psychosis: A review of the literature. *Annals of General Psychiatry, 10*(1), Article 14. https://doi.org/10.1186/1744-859X-10-14

Shaver, J. (2022). The state of telehealth before and after the COVID-19 pandemic. *Primary Care, 49*(4), 517–530. https://doi.org/10.1016/j.pop.2022.04.002

Sun, H., Conrad, F. G., & Kreuter, F. (2021). The relationship between interviewer-respondent rapport and data quality. *Journal of Survey Statistics and Methodology, 9*(3), 429–448. https://doi.org/10.1093/jssam/smz043

Thórólfsson, A., Jónsdóttir, G., Hjaltason, A., & Guðmundsson, H. (2023, July). *Mode effects of video interviewing as a proxy of face-to-face interviewing in R10 of ESS in Iceland* [Conference presentation]. European Survey Research Association Conference, Milan, Italy.

Thunberg, S., & Arnell, L. (2021). Pioneering the use of technologies in qualitative research—A research review of the use of digital interviews. *International Journal of Social Research Methodology, 25*(6), 757–768. https://doi.org/10.1080/13645579.2021.1935565

Appendix

Table 1 Video interviewing and mental health disorders with adjustment by sociodemographic characteristics: Results of independent multivariate logistic regression models with the disorder of interest as the outcome variable

Variable and levels	SSD	SSD-lifetime	MDD	GAD	Bipolar 1	PTSD	OCD	Any	Two or more
Mode: Video	0.80 [0.31, 2.06]	1.07 [0.48, 2.39]	1.03 [0.75, 1.40]	1.17 [0.84, 1.65]	**2.96** **[1.42, 6.17]**	1.35 [0.82, 2.23]	**2.16** **[1.20, 3.90]**	1.12 [0.85, 1.47]	**1.64** **[1.23, 2.19]**
Age: 18–25	0.79 [0.19, 3.35]	0.95 [0.31, 2.87]	**2.34** **[1.47, 3.73]**	**3.67** **[1.99, 6.79]**	**3.25** **[1.56, 6.75]**	2.61 [1.05, 6.50]	**4.76** **[2.46, 9.22]**	**3.22** **[2.01, 5.13]**	**3.54** **[2.37, 5.28]**
Age: 26–44	1.72 [0.72, 4.10]	1.54 [0.75, 3.19]	**1.87** **[1.36, 2.57]**	**1.70** **[1.25, 2.32]**	1.26 [0.63, 2.54]	**3.52** **[2.24, 5.53]**	2.55 [1.37, 4.76]	**2.01** **[1.53, 2.65]**	**2.02** **[1.52, 2.69]**
Sex: Male	1.78 [0.84, 3.76]	1.33 [0.72, 2.46]	**0.65** **[0.50, 0.84]**	**0.52** **[0.38, 0.73]**	0.78 [0.39, 1.58]	0.59 [0.32, 1.09]	0.66 [0.40, 1.09]	**0.65** **[0.51, 0.83]**	**0.56** **[0.40, 0.80]**
R/E: Hispanic	0.30 [0.09, 1.05]	0.46 [0.20, 1.07]	0.59 [0.37, 0.94]	**0.53** **[0.39, 0.72]**	0.43 [0.14, 1.28]	0.52 [0.27, 0.97]	0.97 [0.49, 1.92]	**0.49** **[0.34, 0.70]**	0.72 [0.44, 1.19]
R/E: Not Hispanic (NH) Black	1.13 [0.38, 3.37]	2.43 [0.71, 8.27]	0.73 [0.44, 1.21]	0.88 [0.50, 1.55]	1.46 [0.34, 6.20]	0.68 [0.30, 1.55]	1.20 [0.47, 3.07]	0.96 [0.61, 1.51]	1.08 [0.55, 2.10]
R/E: NH Multiple	3.20 [0.92, 11.10]	4.04 [1.30, 12.60]	1.20 [0.72, 2.00]	1.99 [0.93, 4.22]	**0.08** **[0.02, 0.39]**	0.79 [0.33, 1.87]	2.44 [1.07, 5.59]	1.79 [1.10, 2.89]	1.90 [0.91, 3.98]
R/E: NH Asian, American Indian/ Alaska Native or Native Hawaiian/ Other Pacific Islander	3.96 [0.67, 23.50]	2.98 [0.46, 19.41]	0.82 [0.44, 1.53]	0.62 [0.30, 1.28]	0.30 [0.08, 1.08]	**0.22** **[0.10, 0.49]**	0.51 [0.18, 1.50]	0.85 [0.48, 1.51]	**0.37** **[0.18, 0.75]**
Education: Less than High School (HS)	**12.65** **[2.94, 54.39]**	**5.27** **[1.64, 16.95]**	0.71 [0.36, 1.40]	0.55 [0.26, 1.16]	0.14 [0.02, 0.83]	1.67 [0.47, 5.96]	2.76 [1.07, 7.13]	0.70 [0.41, 1.20]	1.16 [0.58, 2.33]
Education: HS or equivalent	14.39 [4.54, 45.60]	10.10 [3.75, 27.22]	1.13 [0.72, 1.77]	0.45 [0.23, 0.90]	0.53 [0.22, 1.29]	1.71 [0.77, 3.79]	0.88 [0.36, 2.16]	0.77 [0.48, 1.23]	1.06 [0.63, 1.76]
Education: Some college or Associate	**8.24** **[3.60, 18.86]**	**5.78** **[2.76, 12.09]**	1.12 [0.77, 1.63]	0.77 [0.51, 1.15]	2.88 [0.95, 8.78]	**2.27** **[1.21, 4.26]**	1.36 [0.82, 2.26]	0.98 [0.70, 1.38]	1.57 [0.97, 2.54]
INCOME_C4:1 = Less than $50,000	8.08 [2.12, 30.83]	9.05 [3.20, 25.59]	1.58 [1.05, 2.39]	**2.03** **[1.25, 3.29]**	**5.24** **[2.36, 11.62]**	2.41 [1.13, 5.12]	1.18 [0.54, 2.58]	**2.39** **[1.62, 3.52]**	**1.94** **[1.25, 3.01]**
INCOME_C4:2 = $50,000-$99,999	4.34 [0.78, 24.10]	4.62 [1.23, 17.32]	1.57 [1.00, 2.45]	1.47 [0.99, 2.18]	1.47 [0.63, 3.41]	1.47 [0.78, 2.76]	0.89 [0.43, 1.83]	**1.68** **[1.15, 2.45]**	1.42 [0.95, 2.12]
INCOME_C4:4 = Missing	12.99 [2.45, 68.80]	13.57 [3.71, 49.65]	1.43 [0.78, 2.64]	1.44 [0.75, 2.77]	2.82 [0.95, 8.36]	**0.71** **[0.27, 1.83]**	2.66 [1.11, 6.37]	**1.97** **[1.17, 3.32]**	1.72 [0.87, 3.40]
REGION:1 = Midwest	1.07 [0.39, 2.92]	0.53 [0.18, 1.55]	1.11 [0.76, 1.63]	1.27 [0.80, 2.02]	1.42 [0.54, 3.73]	0.97 [0.40, 2.35]	0.89 [0.47, 1.68]	0.91 [0.62, 1.34]	1.31 [0.81, 2.14]
REGION:2 = Northeast	0.78 [0.18, 3.45]	0.30 [0.05, 1.70]	1.16 [0.74, 1.81]	1.06 [0.69, 1.61]	0.94 [0.32, 2.77]	0.74 [0.35, 1.60]	0.72 [0.34, 1.54]	0.93 [0.63, 1.37]	1.05 [0.67, 1.64]
REGION:3 = South	1.36 [0.46, 3.95]	0.57 [0.17, 1.90]	1.09 [0.72, 1.66]	1.23 [0.86, 1.77]	2.22 [0.78, 6.32]	1.20 [0.48, 2.96]	1.34 [0.76, 2.38]	1.04 [0.73, 1.49]	1.37 [0.90, 2.09]
Urban: Urban	0.86 [0.22, 3.27]	0.96 [0.31, 3.03]	1.46 [1.00, 2.14]	0.69 [0.37, 1.26]	0.70 [0.27, 1.82]	0.71 [0.31, 1.65]	0.66 [0.32, 1.35]	0.90 [0.57, 1.42]	0.83 [0.53, 1.28]

Notes: SSD = Schizophrenia spectrum disorder; MDD = major depressive disorder; GAD = generalized anxiety disorder; PTSD = posttraumatic stress disorder; OCD = obsessive-compulsive disorder; R/E: race/ethnicity; NH: Non Hispanic.
Reference groups: mode = phone, age = 45–65, sex = female, race/ethnicity = NH White, education = Bachelor's degree or higher, income > $100,000, region = west, urbanicity = rural; **Bold:** $p < .01$

Effects of the Self-View Window in Live Video Survey Interviews

Shelley Feuer[1] & Michael F. Schober[2]

[1] *National Center for Science and Engineering Statistics, U.S. National Science Foundation*

[2] *New School for Social Research*

Abstract

The studies reported here explore how the "self-view" window (a live video feed of oneself) affects live video survey respondents' likelihood of disclosing sensitive information and their feelings about the interview. In Study 1 (2012), 124 laboratory respondents answered sensitive and nonsensitive questions taken from US government and social scientific surveys over Skype, either with or without a self-view window. Respondents randomly assigned to having a self-view disclosed no less sensitive information than those without a self-view, and on a few questions, they disclosed more (more frequent alcohol use and more sex partners). Self-view respondents also perceived the interview as less sensitive, and they reported less copresence with the interviewer, reduced self-consciousness, and greater comfort answering many of the sensitive questions. Study 2 (2017) replicates these findings in a second sample of 133 respondents by (a) tracking where video survey respondents look on the screen—at the interviewer, at the self-view, or elsewhere—while answering the same survey questions and (b) examining how gaze location and duration differ for sensitive vs. nonsensitive questions and for more and less socially desirable answers. Findings include that self-view respondents looked less at the self-view while answering sensitive (vs. nonsensitive) questions, and that respondents who looked more at the self-view window reported feeling less self-conscious and less worried about how they presented to the interviewer. Results demonstrate that the self-view can change respondents' experience and where they look during a video interview. They also document, for the first time in video surveys, surprising individual variability in looking at the self-view, with some respondents never once looking and others looking at their self-view as much as 50% of the time. Attending to how self-view and respondents' choices (e.g., turning it on or off) affect respondent experience and data quality will be important as live video surveys are increasingly deployed.

Keywords: live video interviews, self-view, survey methodology, gaze, interviewing, disclosure

© The Author(s) 2025. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

As live video is tested in the field as a plausible mode for conducting surveys on a range of topics in different populations (Centeno et al., 2023; Hanson et al., 2023; Neiger et al., 2023; Sanchez et al., 2023; Thórólfsson et al., 2023; Zavala-Rojas et al., 2023), more needs to be understood about how the particular features of video will affect survey data quality and respondent experience relative to other survey modes. Thus far, the emerging evidence from studies that randomly assign participants to video surveys or to other modes (Conrad et al., 2023; Endres & Hillygus, 2023) suggests that across several measures of data quality video surveys can elicit responses comparable in quality (for good and ill) to those elicited in in-person interviews, and different from those elicited in self-administered interviews. For example, respondents randomly assigned to live video interviewing differentiate their responses in a series of questions that use the same response scale—presumably reflecting more thoughtful responding and better data quality—more than respondents in web surveys (Conrad et al., 2023; Endres et al., 2023), and at similar levels to respondents in in-person interviews (Endres et al., 2023). But live video respondents disclose less sensitive information—presumably reflecting social desirability bias and poorer data quality—than respondents in a web survey (Conrad et al., 2023; Endres et al., 2023), and at similar levels to respondents in in-person interviews (Endres et al., 2023).[1]

The evidence thus far is based on particular implementations of live video surveys, in studies using particular video platforms that bring their own features and options and that may connect more or less well with different potential respondent populations depending on the devices they are optimized for (Schober et al., 2020). But even within particular platforms (e.g., Zoom, Teams, FaceTime, WhatsApp, Skype) there are options for how a video interview is

[1] Self-administered "prerecorded video" interviews, in which respondents play video of a recorded interviewer asking the questions before entering responses, seem to provide a different mix of the benefits and drawbacks of live video–more disclosure and less rounding of numerical information, but also more nondifferentiation (Conrad et al., 2023).

Acknowledgements
This work was supported by a National Science Foundation Doctoral Dissertation Research Improvement grant (Methodology, Measurement, and Statistics program) [SES-1632015]; the National Science Foundation (Methodology, Measurement, and Statistics program) [SES-1025645]; the Charles Cannell Fund in Survey Methodology; and New School for Social Research faculty research support. We thank Heidi Reichert for her consultation on statistical procedures, and Courtney Snavely for support preparing video files and linking gaze data for Study 2 analyses.

Notes
The views represented in this paper are not necessarily those of the US Federal government, the National Center for Science and Engineering Statistics, or the National Science Foundation.

Direct correspondence to
Shelley Feuer, National Science Foundation, Alexandria, VA, USA
E-mail: sfeuer@nsf.gov

deployed that have the potential to affect data quality and respondent experience, for example whether and when the interviewer's or the respondent's camera is on or off, where the interviewer's face appears on the respondent's device's screen, whether screen sharing or an attention-competing chat feature is used, and whether the respondent can see their self-view image during the interview or not. The current study focuses on effects of one of these features: the presence or absence of the respondent's self-view during a video survey interview, and in particular an interview that includes sensitive questions.

How might seeing oneself affect respondents' experience and data quality in a standardized video survey interview? To our knowledge, there is no direct evidence to date. The evidence on how video self-view can affect people's behavior and their feelings in other domains of interaction suggest that the effects may differ in different task contexts and for different populations, just as positive and negative effects of video interaction more generally may differ across contexts and tasks (see Seitz et al., 2024). For example, in one study (Shockley et al., 2021) employees of a US corporation reported more fatigue and less engagement during group video meetings when their camera was required to be on (likely entailing a self-view). In contrast, in another study (Abramova et al., 2021) Western European workers who self-reported looking at themselves more while speaking during video work meetings reported greater satisfaction with the meeting process and enjoying video meetings more than those who looked at themselves less, although those who reported looking at themselves more while *listening* reported less satisfaction with the meeting process. In an educational context, university undergraduates in one study who had a self-view, playing the roles of boss and employee in a disciplinary workplace conversation, evaluated their own performance in the conversation and the outcome more negatively than those who did not have a self-view (Shin et al., 2022); yet US undergraduates learning from a Zoom psychology lecture in another study who saw themselves during the lecture performed as well on a quiz as those who did not (Austin et al., 2022).

Potential mechanisms underlying self-view effects on behavior and feelings in other contexts include that a self-view increases self-focus or self-awareness, much as a mirror in in-person situations does (Carver & Scheier, 1978). Increased private self-awareness has been demonstrated in video conversations where strangers asked each other scripted questions about non-sensitive matters (e.g., "What do you like to eat on your pizza?"): participants with a self-view scored higher on a self-awareness scale than those with no self-view (Miller et al., 2017). The self-view can also provide an ongoing reminder that others can see you right now—what has been called public self-awareness—which also can increase self-focus (Carver & Scheier, 1978). On the negative end, this may lead to a greater sensitivity to potential or actual negative evaluation (Fenigstein, 1979), as well as to self-consciousness about one's appearance or "appearance anxiety" (Tien et al., 2023). The self-view, as additional visual input not ordinarily experienced in in-person interaction, may also contribute to the "nonverbal overload" that has been argued to lead to video fatigue more generally (Bailenson, 2021). On the positive end, people with the increased private self-awareness that a self-view brings may have better access to their own physical and emotional states (Joinson, 2001), potentially leading to their feeling less defensive. They may also be more comfortable knowing that the video connection is working, and that they look okay, than non-self-view respondents.

To complicate matters, the desirability or aversiveness of self-view may also vary across individuals, plausibly depending on their ongoing or trait levels of public self-consciousness (Kuhn, 2022) and the extent to which they generally "self-objectify" (Pfund et al., 2020). In one study, open-ended descriptions by college-age business students of their reactions to self-view in the video classroom (Kuhn, 2022) ranged from people disliking seeing their face during a class and finding it distracting to liking seeing themselves. In another study (Balogova & Brumby, 2022) many online participants reported not minding a self-view at all, but some reported that it made them feel somewhat or very uncomfortable and that they find the self-view somewhat or very distracting.

How might these self-view effects and proposed mechanisms apply in standardized video survey interviews? While it is plausible that video interlocutors' goals and behaviors in video survey interviews may overlap in some ways with what they feel and do in other video contexts, the video survey interviewing situation is distinct in several ways: respondents do not typically initiate the interaction, which is intended to elicit information from the respondent (different than situations in which a participant is seeking information, see Schober et al., 2003); an interviewer follows a standardized script that can include sensitive or threatening questions; the interaction is typically dyadic as opposed to in a group, etc. This suggests that the mechanisms that have been proposed to account for negative and positive effects of self-view in other contexts could well be relevant to the specific situation of participating in a standardized survey interview that includes sensitive questions—but not necessarily.

In the studies presented here, we experimentally measure the effects of a self-view on survey respondents' disclosure of sensitive information (their responses to questions about survey questions that are likely to be sensitive or threatening), their reported post-interview comfort answering the questions, and their feelings of connection with the interviewer. We compare survey responses and comfort for respondents randomly assigned to participate in an interview with a self-view to those assigned to participate in an interview with no self-view. Based on evidence and theorizing in other contexts of video use that self-view and its concomitant increased self-awareness may be harmful or helpful, we test competing hypotheses:

Hypothesis 1.1: Self-view respondents will disclose less sensitive information and report feeling less comfortable answering sensitive questions than respondents with no self-view. This effect should be observed if a greater proportion of self-view respondents feel a need to present themselves in a positive light than non-self-view respondents, whether because of greater private self-awareness or greater self-consciousness about their appearance or both.

Hypothesis 1.2: Self-view respondents will disclose more sensitive information and report feeling more comfortable answering sensitive questions than non-self-view respondents. This effect should be observed if respondents with greater self-awareness from having a self-view have better access to their internal states than non-self-view respondents, as well as if the self-view leads more respondents to feel comfortable knowing that the connection is working and that they look okay.

Study 1

Respondents

Respondents were recruited from a Craigslist New York ad and offered a $20 cash incentive to participate in a laboratory study. Of the 124 participants, 60 were randomly assigned to the *self-view* condition (respondents could see themselves and the interviewer) and 64 to the *no-self-view* (respondents could only see the interviewer) condition. (See Figure 1 for interface examples.) As one would expect with random assignment to conditions, respondents in the two groups did not differ in age (averaging 37.4 years (SD = 13.7) in the self-view group and 33.5 (SD = 10.4) in the no self-view group, $F(1, 123) = 3.13$, n.s.), sex (χ^2 (1, N = 124) = .99, n.s.), level of education (χ^2 (3, N = 124) = 5.35, n.s.) or racial / ethnic identities (χ^2 (5, N = 124) = 6.44, n.s.). See Table 1 for details.

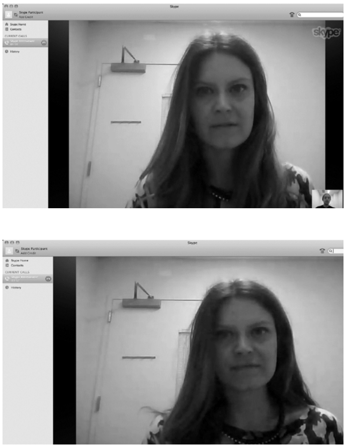


Figure 1 Respondent's view of interviewer with self-view window (top) and without self-view window (bottom)

Respondents were also asked to provide details about their computer use: to estimate how often (if ever) they used text-based and video-based chat programs (e.g., gChat, Skype), on a scale from never to multiple times a day, and to rate how comfortable they were using video chat programs, from 'not at all' to 'extremely.' Analysis of variance showed no reliable differences between the conditions for any of these measures, *F*s = n.s.

Table 1 Demographic characteristics of respondents in Study 1 and Study 2, as self-reported at the end of the questionnaire

	Study 1 Self-view (N = 64)		Study 1 No self-view (N = 60)		Study 2 Self-view (N = 66)		Study 2 No self-view (N = 67)	
Characteristic	*n*	%	*n*	%	*n*	%	*n*	%
Sex								
Male	23	35.9	22	36.7	38	55.6	30	44.8
Female	41	64.1	28	63.3	27	40.9	35	52.2
Level of education								
Some high school (no degree)	0	0	0	0	2	3.0	0	0
High school diploma/GED	2	3.1	7	11.7	9	13.6	12	17.9
Some college or Associate's degree	21	32.8	23	38.3	22	33.4	25	37.3
Bachelor's degree	32	50.0	23	28.3	27	40.9	19	28.4
Graduate	9	14.1	7	11.7	6	9.1	11	16.4
Race/Ethnicity								
White	22	34.4	19	31.7	20	30.8	20	30.3
Black/African-American	26	40.6	19	31.7	25	38.5	32	48.5
Asian	5	7.8	11	18.3	8	12.3	8	12.2
Hispanic/Latino	5	7.8	7	11.7	8	12.3	4	6.1
Multiracial	4	6.3	2	3.3	3	4.6	2	3.0
Other	1	1.6	1	1.7	1	1.5	0	0

Notes: For Study 1, M = 35.5 (SD = 12.3) for age. For Study 2, M = 40.6 (SD = 13.6) for age. For Study 2, 3 respondents did not report their sex.

Interviewers

The five interviewers, all of whom identified as female, were doctoral candidates in clinical psychology with notable interviewing experience in clinical settings. The interviewers were trained in standardized survey interviewing protocols, based on Fowler and Mangione's (1990) guidelines. Interviewers, who were blind to the study hypotheses and to the fact that some respondents assigned to them had a self-view and others did not, each conducted between 18 and 28 interviews (interviewing approximately 12 respondents in each condition). Each interview took approximately 15 minutes, ranging from 12–19 minutes. Interviewers conducted the interview in a separate room from the respondents, who did not know that the interviewer was in the same building; the interviewers used a laptop with a computer-assisted interviewing (CAI) interface (in Qualtrics) and a Skype window to enter the respondents' spoken responses on their laptop.

Survey Questions

The 43 survey interview questions, also used in Lind et al. (2013), ask about health and fitness, sexual health and behaviors, alcohol use, personal finances and political activism. The questions were selected from US government and social scientific surveys: the General Social Survey (GSS), Survey of Consumer Finances (SCF) and Behavioral Risk Factor Surveillance System (BRFSS). These questions had been selected for use in Lind et al. (2013) either because 1) they had been shown specifically to demonstrate mode effects on disclosure of sensitive information in previous research—less disclosure of socially undesirable information in more personal (interviewer-administered) than self-administered modes—or 2) there was a chance that respondents would engage in socially desirable responding. Response options among the questions are categorical (yes/no or multiple response options), numerical or ordinal (rating scales like "every day", "a few times per week", "once a week", etc.). The temporal frame (reference period) for the questions varied from the present, to the past 30 days, to the past 12 months, to time since the respondent's 18th birthday or their entire life. A CAI interface was implemented in Qualtrics and controlled by interviewers to initiate the next question and record responses. See Appendix A for a full list of questionnaire items.

Table 2 identifies domains for which there was specific evidence from previous studies using large samples of respondents of mode effects (first column), as well as domains where it was plausible that mode effects might be observed. Given that the lab context, New York City sample and Craigslist recruitment method for the current study are the same as for the Lind et al. study, this previous evidence seemed likely to be particularly relevant.[2]

[2] The domains identified with "yes" for Lind et al. were categorized as sensitive for the primary gaze analyses in Study 2.

Table 2 Mode effects previously demonstrated for specific survey question topics

Question topic	Mode effects previously demonstrated	Mode effects demonstrated in Lind et al. (2013)
Weight	Yes	No
Exercise	Yes	No
Smoking	Yes	Yes
Number of sex partners – in past 12 months	Yes	No
Number of sex partners – since age 18	Yes	Yes
Frequency of sex	No	No
HIV – tested ever	No	No
Alcohol use – days in past 30	Yes	No
Alcohol use – binge drinking in past 30 days	Yes	No
Alcohol use – regularity of use in past 12 months	Yes	Yes
Alcohol use – drinks typically consumed on average day	Yes	No
Newspaper reading	Yes	Yes
Television hours	Yes	Yes
TV – Primetime drama and sitcom viewing	Yes	Yes
TV – News program viewing	No	No
Religious services	Yes	No
Having credit cards	Yes	Yes
New charges on credit card	No	No
Saving habits	Yes	Yes
Spending	No	No
Voting – in 2012 presidential election	No	No
Voting – in local elections	No	No
Donate blood	No	No
Giving food or money to homeless	Yes	Yes
Doing volunteer work	Yes	No
Giving money to charity	No	No
Offering seat to stranger	Yes	Yes

Procedure

Data were collected between April and August of 2012. The study was conducted in a laboratory to fully control administration, with respondents and interviewers Skyping from separate rooms. Respondents were told that they would be answering a series of questions about their lives over the computer, in real time, by a trained interviewer. They were then brought to a different room and seated in front of a computer monitor where they were informed that the interviewer would be calling over Skype in a few minutes. They were told that to start the connection they were to press the green key that had a picture of a video camera on it. Once the connection was established, the experimenter left the room, and the interviewer proceeded to ask the survey interview questions.

Respondents were never informed that the interviewer was in the same building (and none ever guessed it, as evident in their responses to the post-interview questionnaires). After the interview, respondents filled out the post-interview questionnaires on a separate computer.

Post-Interview Questionnaires

Respondents completed a series of post-experiment questionnaires rating their experience with the interviewer and the interview itself. They also answered questions about their demographic characteristics and their experience with and comfort using computers and video technology in general. Self-view respondents also responded to questions asking for self-reported estimations of frequency of glances at the self-view window and general awareness of the image, as well as whether they ever attempted to or wanted to close the image. All respondents were then presented with textual versions of the survey interview questions they had just answered and asked to rate how comfortable they had felt answering each one, as well as how upset they would be if someone else learned their answers to each question. They also filled out a self-consciousness scale (Scheier & Carver, 1985) asking them to rate the degree to which each statement is characteristic of the way they see themselves.[3] See Supplementary material A for all post-interview questionnaires.

[3] Note that for the self-view respondents the comfort and upset ratings and self-consciousness scale were presented after they had reported about their experience of the self-view window (which made sense for that point in the questionnaire), and so there is the possibility that these ratings by self-view respondents could have been affected by their reflection on the self-view experience.

Statistical Methods

We fitted a series of multilevel regression models to evaluate the fixed effect of condition on response, with interviewer included as a random effect.

For the twelve questions requiring open numerical (continuous) responses, models were estimated using restricted maximum likelihood methods in IBM SPSS Statistics version 27 (Linear Mixed-Effects Models using the 'mixed' command). For the seven questions that required binary yes/no responses, we fitted multilevel mixed-effects logistic regressions in Stata/SE version 16.1 (using the 'melogit' command). For the five questions requiring a choice from an unordered response scale, we fitted multilevel multinomial logistic regressions in Stata/SE version 16.1 (using the 'mlogit' and 'gsem' commands). For the nineteen questions requiring selection for an ordered response scale (e.g., *Every day, Several times a week, Several times a month, Rarely,* or *Never*), multilevel ordinal regression models were fitted (using Stata's 'meologit' command). Significance of the model coefficients was evaluated using the z statistic, and R^2 and McFadden's pseudo R^2 (McFadden, 1974) were used as measures of model fit for the continuous and non-continuous question types, respectively. All tests were two-sided using alpha = 0.05.

For the post-interview questions asking for comfort ratings, self-consciousness ratings, and experience with the interview and interviewer, ANOVAs were conducted to test group differences (as the majority of those questions required continuous responses). For the post-interview ratings of each survey question asking *How comfortable were you answering this question?* and *How upset would you be if someone else found out your answer?* a binomial sign test was used to determine if either group reported different mean levels of *comfort* or different mean levels of *upset*.

Results

Survey Questions

For the twelve questions requiring numeric responses, the interviewer random effect was not significant in any model, meaning that the multilevel model was no better a fit than the OLS model for those questions. So, while interviewer was included in the models, Table 3 only reports the coefficients of condition for each question.

As Table 3 shows, for 9 of the questions there was no effect of self-view on response (disclosure). In two cases, self-view respondents disclosed more sensitive information–specifically, they reported having had more sex partners (combining responses about both male and female partners in two separate questions) and drinking more often (R^2 = 3.3% and 3.2%, respectively). Specifically, self-view respondents reported having had more sex partners (M = 16.8, *SD* = 23.3) than those in the no-self view condition (*M* = 9.2, *SD* = 12.4), and more frequent alcohol use (*M* = 1.8 days per week, *SD* = 1.8) than those in the no-self view condition (*M* = 1.2, *SD* = 1.3).

Table 3 Multilevel mixed-effects regression modeling effects of self-view (condition) on response (numeric)

Question	Coefficient	(*SE*)
Servings fruit/vegetables yesterday	-0.02	(0.31)
Sex partners in last year	-0.54	(0.54)
Total sex partners since 18th birthday	-7.54*	(3.38)
Alcohol use (days/week)	-0.64*	(0.29)
Alcohol use (days/month)	-1.01	(1.24)
Binge drink (times/month)	0.12	(0.35)
Avg number of drinks per day	-0.05	(0.26)
TV (hours/day)	-0.40	(0.51)
Number of credit cards	-0.22	(0.16)
New charges on credit cards	-3.78	(2.59)
Credit card balance	-0.15	(0.37)

Notes: Total sex partners since 18th birthday combines male and female partners, which were asked separately.
$^*p < .05$.

For questions requiring binary yes/no responses, Table 4 only reports the coefficients of condition for each question. The random effect for interviewer was not statistically significant for all but 1 of the questions, meaning that the multilevel model was no better a fit than the OLS model for those questions. Looking at the significance and direction of condition for the model where interviewer was significant, there was still no effect of condition. Altogether, looking at questions with binary response options, the multilevel mixed-effects logistic regression models show no effect of self-view on response distributions (McFadden's pseudo R^2 all less than 1%).

Table 4 Multilevel mixed-effects logistic regression modeling effects of self-view (condition) on response (binary)

Question	Coefficient	(*SE*)
Smoked > 100 cigarettes	-0.07	(0.40)
Tested for HIV	-0.43	(0.40)
Read novels	-0.03	(0.44)
Have credit cards	-0.12	(0.37)
Have credit cards paid off over time	-0.13	(0.41)
Attended political meetings/rallies	0.27	(0.41)
Contributed money to politics	-0.14	(0.42)

For the nineteen questions requiring selection for an ordered response scale, Table 5 again only reports the coefficients of condition for each question. The random effect for interviewer was not statistically significant for all the questions, meaning that the multilevel model was no better a fit than the OLS model for those questions. Altogether, looking at questions with ordinal response options, the multilevel mixed-effects ordered logistic regression models show no effect of self-view on response distributions (McFadden's pseudo R^2 all less than 2%).

Table 5 Multilevel mixed-effects ordered logistic regression modeling effects of self-view (condition) on response

Question	Coefficient	(*SE*)
Health (poor – excellent)	-0.23	(0.33)
Physical exam (past 6 months – 5+ years ago)	-0.23	(0.34)
Weight (very underweight – very overweight)	-0.04	(0.33)
Exercise (0–7 days)	-0.40	(0.33)
Sex past year (not at all – 4+ times/week)	0.22	(0.32)
Alcohol past year (not at all – 5+ times/week)	0.45	(0.32)
Read newspaper (never – every day)	-0.14	(0.33)
Watch TV shows (never – every day)	-0.10	(0.32)
Watch news (never – every day)	-0.27	(0.32)
Attend religious services (never – 1+ times/week)	0.51	(0.33)
Pay off credit card balance (never – always)	0.07	(0.41)
Spending (less than income – exceeded income)	0.39	(0.34)
Follow politics (hardly – most of time)	-0.28	(0.33)
Voting local elections (never – always)	-0.40	(0.33)
Donated blood (not at all – 1+ times/week)	-0.09	(0.51)
Donate to homeless (not at all – 1+ times/week)	-0.83	(0.33)

Table 5 (continued)

Question	Coefficient	(*SE*)
Volunteer work (not at all – 1+ times/week)	0.16	(0.33)
Money to charity (not at all – 1+ times/week)	-0.21	(0.34)
Offered seat to stranger (not at all – 1+ times/week)	-0.23	(0.33)

For categorical responses, Table 6 only reports the coefficients of condition for each question. The random effect for interviewer was not statistically significant for most questions, meaning that the multilevel model was no better a fit than the OLS model for those questions. While the interviewer random effect was significant in two of the questions, there was no significant effect of self-view. Altogether, looking at questions with categorical response options, the multinomial models show no effect of self-view on response distributions (McFadden's pseudo R^2 all less than 3%).

Table 6 Multilevel multinomial logistic regression modeling effects of self-view (condition) on response (categorical)

Question	Reference	Category	Coefficient	(*SE*)
Sex of sex partners	Male	Female	0.11	(0.46)
		Both	-1.00	(0.88)
		No partners	0.04	(0.45)
Sexual orientation	Straight	Gay	1.16	(0.84)
		Bisexual	0.24	(0.64)
Employment	Employed	Self	0.49	(0.56)
		No work 1 year	-0.81	(0.75)
		No work > 1 year	1.24	(0.72)
		Homemaker	15.59	(1687.07)
		Student	-0.27	(0.49)
		Retired	15.59	(2385.90)
		Unable	-0.66	(0.91)
Savings	Spend > income	Spend = income	-0.30	(0.55)
		No plan	-14.40	(869.67)
		Save occasionally	0.22	(1.00)
		Save regularly	-0.05	(0.49)
Vote 2012	Yes	No	-0.78	(0.47)
		Not eligible	-0.44	(0.60)

Post-Interview Questionnaires

Self-view respondents reported less sense of copresence with the interviewer (*M* = 3.1, *SD* = 1.4) than no-self-view respondents (*M* = 3.6, *SD* = 1.3), $F(1, 121) = 4.91$, $p = .033$, $\eta_p^2 = .04$. Self-view respondents also reported feeling that the survey, as a whole, was less sensitive (*M* = 3.1, *SD* = 1.2) than no-self-view respondents (*M* = 3.6, *SD* = 1.1), $F(1, 121) = 3.88$, $p = .051$, $\eta_p^2 = .03$.

When presented with textual versions of the survey questions and asked *How comfortable were you answering this question,* there were generally high levels of reported comfort—but self-view respondents reported higher mean levels of comfort than no-self-view respondents for 32 of 42 questions, binomial/sign test $p < .001$. When asked *How upset would you be if someone else found out your answer,* self-view respondents reported lower mean levels of how upset they would be than no-self-view respondents for 39 of 42 questions, binomial/sign test $p < .001$.

With regards to self-awareness, self-view respondents reported "*thinking about themselves*" reliably more than no-self-view respondents (*M* = 3.0, *SD* = 1.0) and (*M* = 2.6, *SD* = 1.0), respectively, $F(1, 122) = 5.67$, $p = .019$, $\eta_p^2 = .04$. Self-view respondents also reported being marginally less "self-conscious about the way they look" than no-self-view respondents, (*M* = 2.4, *SD* = 1.0) and (*M* = 2.7, *SD* = 0.9), $F(1, 122) = 3.16$ $p = .077$, $\eta_p^2 = .03$.

With regards to the presence of the self-view window, 70% of self-view respondents reported being very aware of it and looking at it quite often.

Discussion

Contrary to Hypothesis 1.1, the evidence reported here is that self-view does not reduce disclosure of sensitive information and does not affect disclosure overall; multilevel models showed no overall effect of self-view on responses. If anything, the disclosure findings show modest support for Hypothesis 1.2—increased disclosure among respondents with a self-view in responses to two sensitive questions: self-view respondents reported more frequent alcohol use and more sex partners.[4] Also consistent with Hypothesis 1.2, self-view respondents reported having felt more comfortable answering the questions and being marginally less "self-conscious about the way they look" than no-self-view respondents. The fact that self-view respondents reported feeling less connection with the interviewer than no-self-view-respondents is also inconsistent with Hypothesis 1.1, and more consistent with prior evidence that people disclose more embarrassing information in situations that involve less social presence with an interviewer (e.g., Kreuter et al., 2008; Lind et al., 2013; Schober et al.,

[4] Note that finding mode differences in social desirability for only one or a subset of sensitive questions is the norm across studies (e.g., Corkrey & Parkinson, 2002; Mott, 1985; Tourangeau & Smith, 1996).

2015; Tourangeau & Smith, 1996). Consistent with prior evidence that self-view increases self-awareness, self-view respondents reported “thinking about themselves” reliably more than no self-view respondents. Most self-view respondents reported being very aware of the self-view window and looking at it quite often.

Study 2 tests the replicability of the Study 1 findings and adds additional measures to further explore the potential mechanisms underlying self-view effects. Rather than relying on self-view respondents’ reports of how often they looked at the self-view window, we measure respondents’ gaze during the interview with eye-tracking equipment. Of course, adding eye-tracking measures changes the interviewing context in potentially important ways, as the equipment must be calibrated for each participant and the session is video-recorded; whether the effects that emerged in Study 1 replicate with these changes is an open question. To the extent that they do, these measures allow us to test more specific hypotheses about how self-view affects eye movements and disclosure, exploring when and how often self-view respondents look at the image of themselves, as well as how their gaze location and duration differ for sensitive vs. neutral questions and for less socially desirable answers.

Study 2

Study 2 replicates Study 1’s procedure and adds eye-tracking measures. Analyses vary three main factors: condition (self-view vs. no self-view), question type (sensitive vs. nonsensitive), and response type (whether the actual response given is non-socially desirable vs. socially desirable). The main dependent measure is gaze duration at the relevant points of interest (POI): the self-view window, the interviewer/screen (these are the same because the interviewer’s image filled the entire screen except for the self-view window region), or elsewhere (away from the screen). Additional dependent measures include respondents’ post-interview reports about the interview experience, their perceptions of the interviewer, and their feelings of comfort and self-consciousness. The analyses are based on particular moments in the interviews, specifically each question-answer (Q-A) sequence (from presentation of one question to presentation of the next). Study 2 tests the following five hypotheses:

Hypothesis 2.1: Respondents will look at the self-view region for a greater proportion of time in self-view than no-self-view interviews. This allows for verification of self-view respondents’ reports that they look at the self-view a lot, while also providing a check that the eye-tracking coordinates measurement was sensible.

Hypothesis 2.2: Respondents will avert their gaze (look away from the interviewer) more during question-answer (Q-A) sequences for sensitive than nonsensitive questions. People have been shown to avert their gaze more when answering cognitively difficult questions (e.g., Doherty-Sneddon et al., 1997; Glenberg et al., 1998) and during survey responses that prove to be less reliable (Schober et al., 2012). Gaze aversion is also connected to social anxiety and poorer relational engagement (Clark & Wells, 1995; Horley et al., 2003). If sensitive questions are more cognitively or interpersonally demanding than nonsensitive questions, and if sensitive questions prompt the fear of being evaluated by others and discomfort, then more gaze aversion should be observed for sensitive questions whether or not respondents have a self-view.

This hypothesis also implies that self-view respondents should look at the self-view window less during sensitive than nonsensitive Q-A sequences, because gaze aversion consists of eye movements to empty or uninformative parts of space (Morency et al., 2006) and the self-view is another source of potentially relevant information. If the pattern is different when respondents have a self-view, this would suggest that the self-view window creates a context that differs from no self-view in terms of cognitive demand and/or dynamic with the interviewer.

Hypothesis 2.3: Respondents will avert their gaze more when giving more embarrassing answers (independent of whether the question is sensitive). This hypothesis tests a different locus of disclosure effects: sensitive *responses.* The same question may not be sensitive for all respondents: some respondents may not find a topic embarrassing or have nothing embarrassing to report, and others may be unembarrassed by reporting what most others would find embarrassing to report. Gaze behavior should then be predictable based on respondents' own feelings of discomfort when answering each question, which we can determine from respondents' ratings immediately after the interview. To further test this hypothesis, similar question-level analyses are carried out based on outside raters' judgments of the sensitivity of answers.

Hypothesis 2.4: Self-view respondents who report greater comfort will have looked at the self-view region more. This hypothesis tests whether the Study 1 respondents' reports of greater reported comfort in the self-view condition occur because these respondents actually look at the self-view window or because they feel comfortable simply because of the mere presence of the self-view window, independent of looking.

Hypothesis 2.5: Self-view respondents who looked at the self-view region more during the interview will report having thought about themselves more during the interview. This hypothesis tests whether the Study 1 finding of greater self-awareness in the self-view condition is related to actually looking at the self-view window or to the mere presence of the self-view window independent of looking. Analysis of gaze at the self-view window also allows for a clearer understanding of the distinction between self-awareness and self-consciousness (heightened awareness of being viewed by another); more frequent looking may be associated with *increased* self-awareness while *decreasing* feelings of self-consciousness (as measured by post-interview reports of the interview experience).

Respondents

One hundred and thirty-three new respondents were recruited using an advertisement on Craigslist New York and offered a $20 cash incentive. As in Study 1, they were randomly assigned to either the self-view (n = 66) or no self-view condition (n = 67) and interviewed over Skype by interviewers unaware of this manipulation. We kept a record of respondents' self-reported demographic characteristics, including those that could be associated with eye movement and gazing, such as age (Romano Bergstrom et al., 2013).

As in Study 1, respondents in the two groups did not differ reliably in age, averaging 40.3 years (SD = 14.2) in the self-view group and 40.8 (SD = 13.1) in the no-self-view group, F (1, 130) = 0.05, n.s. As shown in Table 1, they also did not differ in sex, χ^2 (1, N = 133) = 4.97, n.s.; level of education, χ^2 (4, N = 133) = 5.48, n.s.; racial/ethnic identities, χ^2 (5, N = 133) = 3.39, n.s.; or self-reported computer-use frequency and comfort, Fs = n.s. Focusing only on respondents who had reliable gaze data (N = 119), the two groups also did not differ reliably in age, averaging 40.4 years (SD = 14.6) in the self-view group and 40.1 (SD = 13.2) in the no-self-view group, F (1, 116) = 0.01, n.s. They also did not differ in sex, χ^2 (1, N = 116) = 4.29, n.s.; level of education, χ^2 (4, N = 119) = 4.09, n.s.; racial/ethnic identities, χ^2 (5, N = 118) = 3.32, n.s.; or self-reported computer-use frequency and comfort, Fs = n.s.

Interviewers

As in Study 1, the five new interviewers (4 female, 1 male) were doctoral candidates in clinical psychology with notable interviewing experience in clinical settings. They followed the same procedures as Study 1.

Survey Questions

The survey questions were the same as those used in Study 1. This time the interviewer's online instrument also generated timestamps for the start of each question page and duration on each page, for subsequent linkage with the gaze data.

Eye-Tracking Equipment

Study 2 used The Eye Tribe, a small eye-tracking device that sits in front of the computer monitor and allows for unobtrusive measurement of gaze with free head movement. We used Eye Tribe data acquisition and control software, which displays graphical representations of respondents' binocular gaze data (x/y screen coordinate), 3D eye position and pupil diameter in millimeters; analog and serial real-time outputs of eye position and pupil size were acquired through text files and imported to SPSS for subsequent analyses. The accuracy of recordings was sufficient for study purposes: the margin of error for timing of eye movements (latency) is ±1.6 milliseconds and for position of eye fixations ± 0.5–1 degrees, where a degree of visual angle amounts to approximately 1 cm on the screen at a 50-cm distance (Ooms et al., 2015). The tracking area is 40 cm x 30 cm at a 65 cm distance, with an operating range of 45 cm – 75 cm. The sampling/frame rate was approximately 30 Hz, which means that 30 data points were produced each second (or a data point is produced every 33 milliseconds). The 30 Hz frame rate was selected, rather than the 60 frames/second option that the Eye Tribe (Figure 2) also allows, in order to allow maximally free head movement for respondents (the higher rate would require respondents to be more stationary).

Figure 2 The Eye Tribe

Video-Recording Software

Each interview was video recorded using Camtasia Studio video editing software. The video captured the interviewer's screen, and thus where the participant was looking, as well as the CAI interface. Each video was also saved as an MP4 as an alternate format for subsequent coding.

Procedure

Data were collected between February and May 2017. The procedure was nearly identical to that in Study 1. Respondents were brought to the laboratory and asked to sit comfortably in front of a computer monitor with the Eye Tribe tracker directly in front of it (see Figure 3). The Eye Tribe was moved to ensure respondent-specific centering and an appropriate distance for gaze capture. After respondents completed a short Eye Tribe calibration exercise (in which they follow dots displayed at different parts of the screen with their eyes), the interviewer called them over Skype. The entire interview was video recorded using Camtasia. After the interview, the experimenter walked the respondent to a different room where they completed the same questionnaires as in Study 1.

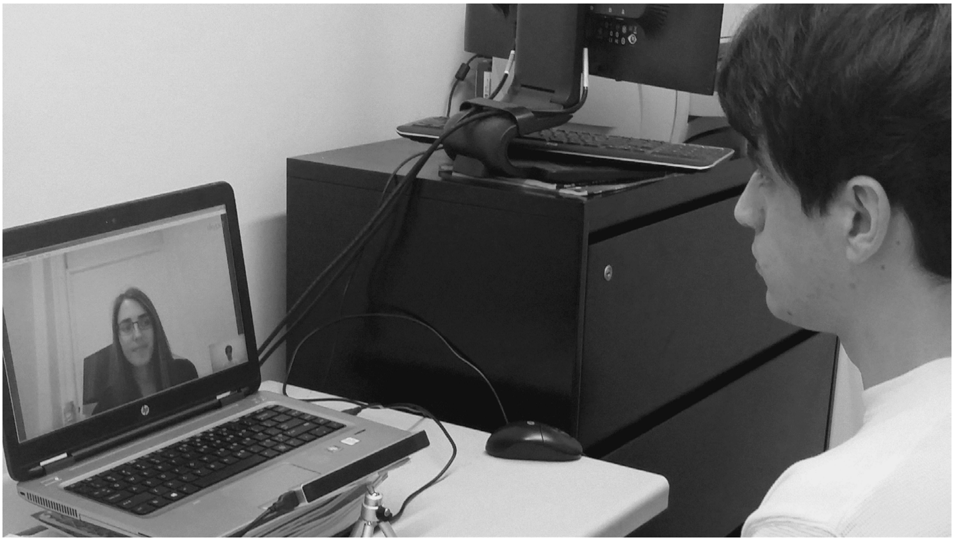


Figure 3 Participant and interviewer engaging in a survey interview

Post-Interview Questionnaires

The post-interview questionnaires were the same as those used in Study 1.

Data Preparation

To test the study's hypotheses, the survey and gaze data needed to undergo several preparatory procedures.

Annotation and Coding

Timestamps were generated for the start of each question (and therefore, the question-answer sequence) using Adobe Premiere software. These were then linked with the stream of gaze behavior from the eye tracker, which generated x and y coordinates of gaze at a rate of 30 frames per second.

The three points of interest (looking at self-view region, looking at interviewer/screen, looking elsewhere) were identified using the pixel-based coordinates from the eye tracker output. The computer screen itself was 1365 x 767 pixels, and so gaze measurements that fit within the following regions were counted accordingly (see Figure 4):

- Looking at self-view window (x > 1130 AND x < 1365) AND (y > 633 AND y < 767)
- Looking away from screen: (x < 0 OR x > 1365) OR (y < 0 OR y > 767)
- Looking elsewhere on screen (x < 1130 OR x > 1365) AND (y < 633 OR y > 767) and NOT looking at self

For a more detailed description of the annotation and coding, see Appendix B.

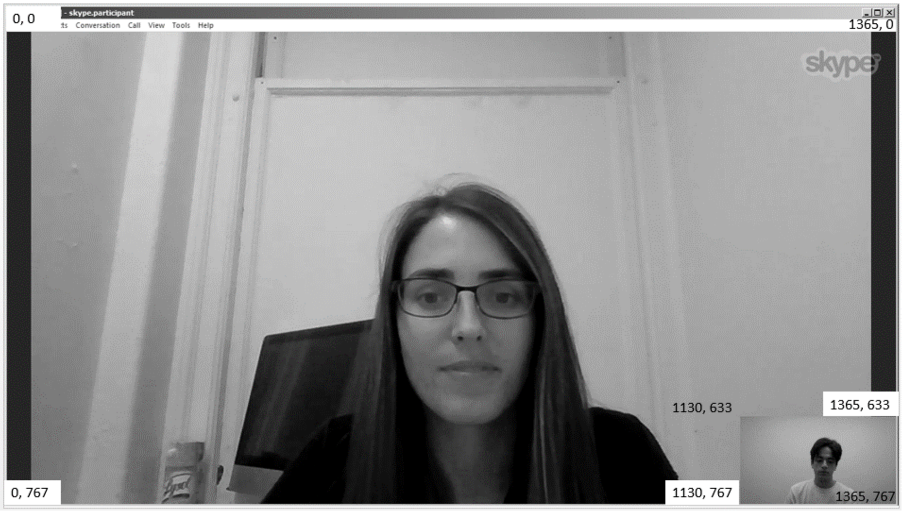


Figure 4 Pixel-based coordinates

Online Rating Data

To test the hypothesis that gaze aversion will be greater for responses that are more embarrassing, question-level analyses were carried out based on the sensitivity judgments of 100 online respondents, recruited to match US national demographic characteristics on sex, race/ethnicity, and education (see Conrad et al., 2023; Fail et al., 2021 for alternate versions of this procedure that differed in minor ways). For each question, raters were asked to judge how embarrassed most people would be simply being asked each question and then how embarrassed they thought most people would be to give each possible response option. There were 3 rating options: *Not at all embarrassed, somewhat embarrassed, very embarrassed*. For more details about the online rating survey see Supplementary material B.

Gaze Data

One-hundred and nineteen respondents (of the 133 total participants) had complete usable gaze data, with no recording errors from the eye tracker or video recorder. For 7 of the 14 respondents whose data could not be used, the eye tracker could not calibrate. The other cases were excluded because either no video data were recorded due to video software malfunction or there were recording problems with the audio tracks that made coding impossible.

Results

Comparability to Study 1

First, to determine whether the self-view and no-self-view findings replicated with the eye tracker[5], we used the same statistical approach as in Study 1. The pattern of responding and post-interview ratings largely replicates. The various multilevel regressions modeling the effects of self-view on response (for each type: numeric, ordinal, binary, unordered categorical), with interviewer as a random effect (see Tables 7–10), generally showed no effect of self-view (all R^2 and pseudo $R^2 < 2.0\%$) with a few exceptions: self-view respondents again reported a higher number of total sex partners (M = 180.1, SD = 55) than no-self-view respondents (M = 21.2, SD = 55.5) with R^2 of 10.2%, and self-view respondents again reported drinking more (M = 5.9 days per month, SD = 8.4) than no-self-view respondents (M = 3.7 days per month, SD = 4.6) with R^2 of 2.0%. (For Study 2, the greater report of alcohol use emerged in responses to the question about days of drinking per month rather than days per week.)

[5] Again, the addition of the eye tracker, though unobtrusive, could change respondents' experience just enough that the original self-view manipulation might differ in important ways. Additionally, Study 2 respondents were video-recorded, which they were not in the previous study. This could also increase self-awareness at different levels (i.e., private and public) and thereby socially desirable responding.

Table 7 Multilevel mixed-effects regression modeling effects of self-view (condition) on response (numeric)

Question	Coefficient	(*SE*)
Servings fruit/vegetables yesterday	-0.13	(0.28)
Sex partners in last year	0.21	(0.57)
Total sex partners since 18th birthday	-158.90*	(78.20)
Alcohol use (days/week)	-0.40	(0.34)
Alcohol use (days/month)	-2.26*	(1.17)
Binge drink (times/month)	-1.16	(0.72)
Avg # drinks/day	-0.23	(0.31)
TV (hours/day)	-0.89	(0.59)
# of credit cards	-0.22	(0.16)
New charges on credit cards	-53.19	(174.20)
Credit card balance	0.42	(0.25)

Notes: *$p < .05$.

Table 8 Multilevel mixed-effects ordered logistic regression modeling effects of self-view (condition) on response

Question	Coefficient	(*SE*)
Health (poor – excellent)	0.52	(0.33)
Physical exam (past 6 months – 5+ years ago)	0.46	(0.33)
Weight (very underweight – very overweight)	-0.61	(0.34)
Exercise (0 – 7 days)	0.46	(0.31)
Sex past year (not at all – 4+ times/week)	0.11	(0.31)
Alcohol past year (not at all – 5+ times/week)	-0.09	(0.31)
Read newspaper (never – every day)	0.43	(0.33)
Watch TV shows (never – every day)	-0.14	(0.31)
Watch news (never – every day)	-0.27	(0.32)
Attend religious services (never – 1+ times/week)	0.26	(0.33)
Pay off credit card balance (never – always)	-0.13	(0.43)
Spending (less than income – exceeded income)	-0.19	(0.33)
Follow politics (hardly – most of time)	0.25	(0.33)
Voting local elections (never – always)	-0.09	(0.32)
Donated blood (not at all – 1+ times/week)	0.67	(0.45)
Donate to homeless (not at all – 1+ times/week)	-0.01	(0.30)
Volunteer work (not at all – 1+ times/week)	0.05	(0.33)
Money to charity (not at all – 1+ times/week)	0.37	(0.34)
Offered seat to stranger (not at all – 1+ times/week)	0.24	(0.31)

Table 9 Multilevel mixed-effects logistic regression modeling effects of self-view on response (binary)

Question	Coefficient	(*SE*)
Smoked > 100 cigarettes	-2.80	(0.38)
Tested for HIV	-0.38	(0.43)
Read novels	-0.22	(0.38)
Have credit cards	-0.04	(0.36)
Have credit cards paid off over time	-2.25	(1.02)
Attended political meetings/rallies	-0.09	(0.38)
Contributed money to politics	-0.41	(0.46)

Table 10 Multilevel multinomial logistic regression modeling effects of self-view (condition) on response (categorical)

Question	Reference	Category	Coefficient	(*SE*)
Sex of sex partners	Male	Female	0.50	(0.42)
		Both	0.67	(0.78)
		No partners	-0.31	(0.46)
Sexual orientation	Straight	Gay	-0.31	(0.52)
		Bisexual	1.13	(0.62)
Employment	Employed	Self	-0.37	(0.47)
		No work 1 year	-1.40	(0.86)
		No work > 1 year	1.80*	(0.83)
		Homemaker	-0.30	(1.03)
		Student	-0.30	(0.64)
		Retired	0.62	(0.88)
		Unable	-0.30	(0.69)
Savings	Spend > income	Spend = income	-0.27	(0.53)
		No plan	-0.07	(1.47)
		Save occasionally	-0.92	(0.79)
		Save regularly	0.08	(0.46)
Vote 2012	Yes	No	0.52	(0.40)
		Not eligible	0.02	(0.57)

Notes: $^{*}p < .05$.

With regards to the Self-Consciousness Scale-Revised (Scheier & Carver, 1985), as in Study 1, self-view respondents again reported being marginally less self-conscious about the way they look than the no-self-view respondents (M = 1.4, SD = 1.0) and (M = 1.7, SD = 1.1), respectively, $F(1, 131) = 3.04$, $p = .081$, $\eta_p^2 = .02$. Self-view respondents also reported thinking about what they said during the interview reliably more than no-self-view respondents (M = 4.5, SD = 0.7) and (M = 4.1, SD = 0.9), respectively, $F(1, 128) = 4.39$, $p = .038$, $\eta_p^2 = .03$. Self-view respondents (M = 3.3, SD = 1.1) also reported that they felt that the interviewer was more empathetic than no-self-view respondents (M = 2.8, SD = 1.2), $F(1, 131) = 4.76$, $p = .031$, $\eta_p^2 = .04$.[6]

Gaze Data

The gaze data consisted of proportions of time spent looking at each of the three POIs (self-view region, the screen/interviewer, elsewhere) during every question-answer sequence for each respondent. Overall, across conditions, respondents spent about 83.8% of the time looking at the screen and 15.7% looking away. As Figure 5 shows, respondents on average did not look at the self-view window very much, and looked elsewhere much more.

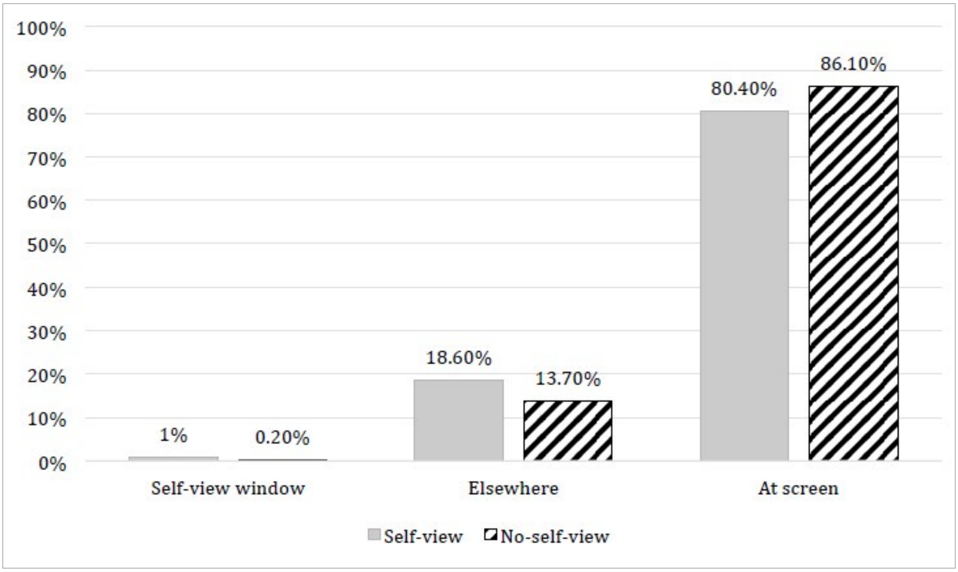


Figure 5 Average gaze at POI, by condition

[6] The one finding from Study 1 that did not replicate was that self-view respondents no longer reported feeling significantly less copresent with the interviewer.

Hypothesis Tests

For Hypotheses 2.1 and 2.2, we fitted a series of multilevel mixed-effects regression models to evaluate the fixed effect of condition on gaze (at the self-view, at the screen, or elsewhere). Because sensitivity is a feature of the question, it was included in the models as a fixed effect to reduce some residual variance, along with potential interactions between condition and question sensitivity. Questions were categorized as sensitive or not based on the Lind et al. (2013) findings. Question duration (length of time from start of one question to the start of the next) was included in the models to explain some of the residual variance. Random effects for interviewer and respondent were included. For Hypothesis 2.3, we fitted a series of multilevel mixed-effects regression models to evaluate the fixed effects of condition and response sensitivity on gaze (at the self-view, at the screen, or elsewhere), along with potential interactions. Response sensitivity was assessed by classifying responses as sensitive if at least 40% of the online judges rated the response as somewhat or very embarrassing (see Appendix D). Random effects for interviewer and respondent were included. For Hypothesis 2.4 question level analyses—using gaze for each question and that question's comfort rating—were conducted. For Hypothesis 2.5, we conducted correlations between responses to questions related to self-awareness and also self-consciousness and self-view gaze.

As Hypothesis 2.1 proposed, Table 11 shows that respondents who had a self-view window looked at this region for a significantly greater proportion of the time than non-self-view respondents. The variables added into the models as plausible contributors to gaze at self-view did not have significant effects on gaze. (Self-view respondents varied enormously in how much time they spent looking at the self-view window; a few looked at it a great deal, and some never looked at it all.) The random effects of interviewer and respondent were not significant.

Table 11 Multilevel regression modeling effects of self-view (condition) on gaze at self

Predictor	Reference category	Coefficient	(*SE*)
Condition	Self-view	0.01***	(0.00)
Q duration		0.00	(0.00)
Q sensitivity	Sensitive	-0.00	(0.00)
Condition & Q sensitivity	Self-view	-0.00	(0.00)

Notes: Q = question. ***p < .001.

For Hypothesis 2.2, that respondents would avert gaze more when answering sensitive than nonsensitive questions, Table 12 shows an effect of condition and question duration, but no main effect of question sensitivity. The corresponding effect of looking at the screen is reported in Table 13. The random effects for interviewer and respondent were not significant. This pattern is not consistent with the first part of Hypothesis 2.2.[7] However, with regards to the hypothesis that respondents with a self-view will look less at the self-view window during sensitive questions compared to nonsensitive, a repeated ANOVA revealed that indeed self-view respondents looked at the self-view window less during sensitive questions (M = 0.0096, SD = 0.016) than nonsensitive questions (M = 0.0074, SD = 0.016), $F(1, 60) = 6.42$, $p = .014$, $\eta_p^2 = .10$.

Table 12 Multilevel regression modeling effects of self-view (condition) on gaze elsewhere

Predictor	Reference category	Coefficient	(*SE*)
Condition	Self-view	0.07*	(0.03)
Q duration		0.00**	(0.00)
Q sensitivity	Sensitive	0.96	(0.34)
Condition & Q sensitivity	Self-view	-0.01	(0.01)

Notes: Q = question. *$p < .05$, **$p < .01$.

Table 13 Multilevel regression modeling effects of self-view (condition) on gaze at screen

Predictor	Reference category	Coefficient	(*SE*)
Condition	Self-view	-0.07**	(0.03)
Q duration		-0.00**	(0.00)
Q sensitivity	Sensitive	-0.00	(0.01)
Condition & Q sensitivity	Self-view	0.01	(0.01)

Notes: Q = question. **$p < .01$.

Other question categorization strategies, including using various cutoffs of the panel data, were also employed. For the classification procedures and a list of this classification of sensitive questions, see Appendix C.

[7] While duration was significant in the models for gaze at screen and gaze elsewhere, its inclusion didn't change the effects of condition, question sensitivity or the interaction between them.

So, Hypothesis 2.2 is only partially supported: self-view respondents looked less at the self-view window during sensitive question-answer sequences, but they did not avert their gaze more while answering sensitive questions than non-self-view respondents.

Hypothesis 2.3, that duration of gaze aversion would be greater for responses that are rated as more embarrassing to provide, was assessed by classifying responses as sensitive if at least 40% of the online judges rated the response as somewhat or very embarrassing (see Appendix D). These values were linked to each respondent's answer and comfort rating for each survey question. Table 14 shows that there was no effect of response sensitivity on gaze aversion.

Table 14 Multilevel regression modeling effects of response sensitivity on gaze elsewhere

Predictor	Reference category	Coefficient	(*SE*)
Condition	Self-view	0.06*	(0.03)
R sensitivity	Sensitive	0.01	(0.01)
Condition & R sensitivity	Self-view	-0.01	(0.01)

Notes: R = response. *p < .05.

An alternative analysis was also carried out using each respondent's post-interview ratings about how their comfort with the interviewer had changed over the course of the interview, as well as their own self-reported comfort ratings in answering each question. (Note that the self-reported comfort ratings—while in one sense providing the most specific respondent-level evidence that should be relevant to these analyses—have the limitation that they didn't decouple question and response sensitivity in the way that online raters' judgments did, and they also ask about comfort rather than embarrassment.) With regards to comfort change, an ANOVA revealed that respondents who reported that their comfort increased also averted their gaze less (M = 0.12, SE = 0.03) than those who reported that their comfort just stayed the same (M = 0.19, SE = 0.02), $F(2, 112) = 3.11$, $p = .048$, $\eta_p^2 = .05$.

Question level analyses—using gaze for each particular question and that question's comfort rating (based on classifying questions rated 5 or higher on the 7-point scale as comfortable)—were conducted. We used a repeated ANOVA with comfort as the within subjects factor and gaze elsewhere as the dependent variable. Condition was also added as a between-subjects factor. Respondents did not avert their gaze differently for questions they reported having been comfortable answering and questions they reported having been uncomfortable answering. So, Hypothesis 2.3 is not supported using response sensitivity ratings from the online raters nor respondents' own post-interview comfort ratings.

To test Hypothesis 2.4, that people in the self-view condition who report greater comfort will have looked at the self-view region more, question level analyses—using gaze for each question and that question's comfort rating—were conducted. We used a repeated ANOVA with comfort as the within subjects factor and gaze at the self-view window as the dependent variable. As predicted, self-view respondents looked at the self-view window for a substantially greater proportion of the time during Q-A sequences for questions they reported having been comfortable answering ($M = 0.10$, $SD = 0.2$) than during Q-A sequences for questions they reported being uncomfortable answering ($M = 0.07$, $SD = 0.01$), $F(1,44) = 4.98$, $p = .031$, $\eta_p^2 = .10$.

To test Hypothesis 2.5, that more frequent looking would be associated with increased self-awareness (while decreasing feelings of self-consciousness), several correlations were conducted. First, respondents who reported looking at themselves more did, in fact, look more at the self-view window, $r(59) = .28$, $p = .032$. Respondents who looked more at the self-view window reported feeling less self-conscious, $r(59) = -.28$, $p = .027$. They were also marginally less concerned about how they presented themselves, and they reported examining their motives less, ps $< .07$. However, for self-view respondents there was no relationship between *Thinking about what they said during the interview* and gaze at the self-view window ($p > .05$).

Additional potential relationships between gaze aversion and experience with the interview, interviewer, and dispositional self-consciousness were also explored. With or without the self-view window, respondents who felt greater copresence during the interview, and perceived the interviewer to be more personal and empathetic, averted their gaze less.

One additional observation: the range of time spent gazing at the self-view window was very large, with some respondents never once looking at that area of the screen, and others looking at the self-view as much as 50% of the interview time. The fact that there is a wide range is consistent with self-reports of attraction to or aversion to the self-view in other arenas of video interaction (e.g., Kuhn, 2022; Pfund et al., 2020), but it also suggests that there is more to be understood about how group effects of self-view are constituted from potentially varying effects on individuals.

Discussion

Study 2 replicates Study 1's demonstration that having a self-view window in a live video survey interview changes respondents' experience and has the potential to change their answers. As in Study 1 (which did not measure respondents' gaze), self-view respondents did not disclose any less than non-self-view respondents; if anything, they disclosed more—again reporting more alcohol drinking and sex partners. They also reported being less self-conscious during the interview and being more self-reflective. Additionally, in Study 2, self-view respondents rated the interviewer as more empathetic. The fact that this pattern of findings replicated even when respondents' gaze was tracked and their interviews video-recorded provides further support for the generality of the self-view effects as measured here.

Study 2's findings also demonstrate that gaze patterns in live video interviews can be informative about respondents' experience and response processes. First, as one would expect, respondents in the self-view condition did look more at the region of the screen containing the self-view window than respondents in the no-self-view condition. Regarding the major hypotheses about gaze direction and duration, the hypothesis that duration of gaze aversion would be greater for sensitive than nonsensitive questions and duration of gaze at self would be lesser for sensitive than nonsensitive questions was partially supported. Categorizing questions as sensitive or not based on the Lind et al. (2013) findings, self-view respondents looked less at the self-view window during sensitive than nonsensitive questions, as predicted. Relatedly, these respondents also looked less at the self-view window for questions that they were uncomfortable answering, which also supports the hypothesis that people with a self-view window who report greater comfort will have looked at the self-view more. With regards to gaze aversion, self-view respondents looked away less during sensitive question-answer sequences (and, correspondingly, looked at the screen more). There was no difference in gaze based on question type for respondents without a self-view window. Taken together, the findings provide good evidence that the presence of a self-view window changes where survey respondents look particularly for sensitive questions.

General Discussion

Taken together, the findings presented here suggest that the self-view window can create a distinct interviewer-respondent dynamic—potentially increasing respondents' engagement with the interviewer and reducing their worry about self-presentation. The fact that the self-view window does not reduce disclosure of sensitive information across a sample—and perhaps may even facilitate it in some cases—suggests that there is much more to explore about how live video might best be deployed for survey interviewing.

On the one hand, the fact that an aspect of the interface setup not under the researcher's control can have significant impacts may be worrying, in that survey respondents might choose settings on their computer or mobile device that harm the quality of survey data they provide. (Our respondents were randomly assigned to having a self-view window or not, but in most platforms that video respondents would be likely to use in field settings, respondents can choose whether to see the self-view by enabling or disabling it, or turning their camera feed on or off.) Given the substantial variability in how often our respondents looked at the self-view window, our findings suggest that effects might be different for different subgroups of respondents—for example, that the effects might proceed through different mechanisms for those who avoid looking at themselves vs. those who fixate.

On the other hand, these findings open the door to potential strategic and sophisticated deployment of features of video platforms for interviewing, once more is known about which respondents who have access to the technology might benefit from or even prefer video interviews to in-person or telephone interviews (Schober et al., 2023) and what people's reasons are for feeling more or less uncomfortable in a live video survey than in other survey modes (Okon et al., 2025). Beyond the display of the self-view or not, current video platforms could allow selectively turning on and off the camera feeds to change the experience of being seen, even as an interview proceeds, so that more sensitive sections could change the respondent's experience of privacy.

Our studies tested effects of self-view in one particular implementation, with a relatively small self-view window in the lower right corner of a desktop screen superimposed on a full-screen view of the interviewer, and no representation of the self in the no-self-view conditions. Self-view can, of course, be instantiated in many different ways on different platforms and devices, from smartphones to multi-screen setups; the view of the self can be as large as the view of the interlocutor, and juxtaposed with the view of the other in many different ways. It is also now common for a placeholder for the self-view image to be visible during a video session that includes the participant's name if the camera is turned off. Beyond how video is implemented on a device, people's attentiveness to what is on the screen can vary in different task settings, for example if they are multitasking (which of course was not an option in our lab study), and what is visible can vary under different camera placements and lighting conditions. Interviewers may also vary in how and when they look at the respondent, and how much "mutual" gaze is an option.

How these variations might change effects of self-view on survey data quality and respondents' experience is of course unknown. As video platforms and norms of usage have evolved, especially post-pandemic, how exactly the findings reported here (data were collected in 2012 and 2017) will apply now also remains to be seen. It is also unknown how respondents' chosen self-view preferences in an interview (e.g., disabling the self-view, having a larger one, turning the camera off partway through) will affect data quality and their experience: whether or not the respondent-selected settings that make them most comfortable will lead to improved data quality, following the evidence and logic that allowing respondents to select a preferred interview mode can improve data quality (e.g., Conrad et al., 2017). In any case, it is clear that effects of respondent-selected video settings in the field will need to be much better understood.

References

Abramova, O., Gladkaya, M., & Krasnova, H. (2021). An unusual encounter with oneself: Exploring the impact of self-view on online meeting outcomes. *ICIS 2021 Proceedings*, 16. https://aisel.aisnet.org/icis2021/is_future_work/is_future_work/16

Austin, M. R., Fogler, K. A. J., & Daniel, D. B. (2022). Seeing self and others on-screen does not negatively impact learning in virtual classrooms. *Scholarship of Teaching and Learning in Psychology, 8*(4), 368–373. https://doi.org/10.1037/stl0000303

Bailenson, J. N. (2021). Nonverbal overload: A theoretical argument for the causes of Zoom fatigue. *Technology, Mind, and Behavior, 2*(1). https://doi.org/10.1037/tmb0000030

Balogova, K., & Brumby, D. (2022). How do you Zoom?: A survey study of how users configure video-conference tools for online meetings. *Proceedings of the 1st Annual Meeting of the Symposium on Human-Computer Interaction for Work*, Article 5, 1–7. https://doi.org/10.1145/3533406.3533408

Carver, C. S., & Scheier, M. F. (1978). Self-focusing effects of dispositional self-consciousness, mirror presence, and audience presence. *Journal of Personality and Social Psychology, 36*(3), 324–332. https://doi.org/10.1037/0022-3514.36.3.324

Centeno, L., Kelley, J., Arrue, J., Edwards, B., Hubbard, R., & Dulaney, R. (2023, July 17–21). *Video interviewing in full production: A new mode is here to stay* [Conference presentation]. European Survey Research Association Conference, Milan, Italy.

Clark, D. M., & Wells, A. (1995). A cognitive model of social phobia. In R. G. Heimberg, M. R. Liebowitz, D. A. Hope, & F. R. Schneier (Eds.), *Social phobia: Diagnosis, assessment, and treatment* (pp. 69–93). The Guilford Press.

Conrad, F. G., Schober, M. F., Antoun, C., Yan, H. Y., Hupp, A. L., Johnston, M., Ehlen, P., Vickers, L., & Zhang, C. (2017). Respondent mode choice in a smartphone survey. *Public Opinion Quarterly, 81*(S1), 307–337. https://doi.org/10.1093/poq/nfw097

Conrad, F. G., Schober, M. F., Hupp, A. L., West, B. T., Larsen, K. M., Ong, A. R., & Wang, T. (2023). Video in survey interviews: Effects on data quality and respondent experience. *Methods, Data, Analyses, 17*(2), 135–170. https://doi.org/10.12758/mda.2022.13

Corkrey, R., & Parkinson, L. (2002). A comparison of four computer-based telephone interviewing methods: Getting answers to sensitive questions. *Behavior Research Methods, Instruments, & Computers, 34*(3), 354–363. https://doi.org/10.3758/BF03195463

Doherty-Sneddon, G., Anderson, A. H., O'Malley, C., Langton, S. R. H., Garrod, S. C., & Bruce, V. (1997). Face-to-face and video mediated communication: A comparison of dialogue structure and task performance. *Journal of Experimental Psychology: Applied, 3*(2), 105–125. https://doi.org/10.1037/1076-898X.3.2.105

Endres, K., Hillygus, D. S., DeBell, M., & Iyengar, S. (2023). A randomized experiment evaluating survey mode effects for video interviewing. *Political Science Research and Methods, 11*(1), 144–159. https://doi.org/10.1017/psrm.2022.30

Fail, S., Schober, M. F., & Conrad, F. G. (2021). The time it takes to reveal embarrassing information in a mobile phone survey. *International Journal of Social Research Methodology, 24*(2), 249–264. https://doi.org/10.1080/13645579.2020.1824629

Fenigstein, A. (1979). Self-consciousness, self-attention, and social interaction. *Journal of Personality and Social Psychology, 37*(1), 75–86. https://doi.org/10.1037/0022-3514.37.1.75

Fowler, F. J., & Mangione, T. W. (1990). *Standardized survey interviewing: Minimizing interviewer-related error.* SAGE Publications. https://doi.org/10.4135/9781412985925

Glenberg, A. M., Schroeder, J. L., & Robertson, D. A. (1998). Averting the gaze disengages the environment and facilitates remembering. *Memory & Cognition, 26*(4), 651–658. https://doi.org/10.3758/BF03211385

Hanson, T., Briceno-Rosas, R., & Mitchell, J. (2023, July 17–21). *Video interviewing on the European Social Survey* [Conference presentation]. European Survey Research Association Conference, Milan, Italy.

Horley, K., Williams, L. M., Gonsalvez, C., & Gordon, E. (2003). Social phobics do not see eye to eye: A visual scanpath study of emotional expression processing. *Journal of Anxiety Disorders, 17*(1), 33–44. https://doi.org/10.1016/s0887-6185(02)00180-9

Joinson, A. N. (2001). Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity. *European Journal of Social Psychology, 31*(2), 177–192. https://doi.org/10.1002/ejsp.36

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

Kuhn, K. M. (2022). The constant mirror: Self-view and attitudes to virtual meetings. *Computers in Human Behavior, 128,* 107110. https://doi.org/10.1016/j.chb.2021.107110

Lind, L. H., Schober, M. F., Conrad, F. G., & Reichert, H. (2013). Why do survey respondents disclose more when computers ask the questions? *Public Opinion Quarterly, 77*(4), 888–935. https://doi.org/10.1093/poq/nft038

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics* (pp. 105–142). Academic Press.

Miller, M. K., Mandryk, R. L., Birk, M. V., Depping, A. E., & Patel, T. (2017). Through the looking glass: The effects of feedback on self-awareness and conversational behaviour during video chat. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 5271–5283. https://doi.org/10.1145/3025453.3025548

Morency, L.-P., Christoudias, C. M., & Darrell, T. (2006). Recognizing gaze aversion gestures in embodied conversational discourse. *Proceedings of the 8th International Conference on Multimodal Interfaces*, 287–294. https://doi.org/10.1145/1180995.1181051

Mott, F. L. (1985). *Evaluation of fertility data and preliminary analytical results from the 1983 (5th Round) Survey of the National Longitudinal Surveys of Work Experience of Youth*. Center for Human Resource Research, Ohio States University.

Neiger, D., Phillips, B., Lester, G., Slamowicz, S., Farrell, E., Gerlach, K., & Carmo, P. (2023, July 17–21). *Video-assisted live interviewing in comparison to other survey methods in Australia* [Conference presentation]. European Survey Research Association Conference, Milan, Italy.

Okon, S., Dolgin, R. S., & Schober, M. F. (2025). Participants' reported discomfort with live video as a mode for answering a sensitive survey question. *Public Opinion Quarterly.* https://doi.org/10.1093/poq/nfaf031

Ooms, K., Dupont, L., Lapon, L., & Popelka, S. (2015). Accuracy and precision of fixation locations recorded with the low-cost Eye Tribe tracker in different experimental set-ups. *Journal of Eye Movement Research, 8*(1), Article 1. https://doi.org/10.16910/jemr.8.1.5

Pfund, G. N., Hill, P. L., & Harriger, J. (2020). Video chatting and appearance satisfaction during COVID-19: Appearance comparisons and self-objectification as moderators. *International Journal of Eating Disorders, 53*(12), 2038–2043. https://doi.org/10.1002/eat.23393

Romano Bergstrom, J. C., Olmsted-Hawala, E. L., & Jans, M. E. (2013). Age-related differences in eye tracking and usability performance: Website usability for older adults. *International Journal of Human–Computer Interaction, 29*(8), 541–548. https://doi.org/10.1080/10447318.2012.728493

Sanchez, C., Brown, M., Cole, K., & Taylor, K. (2023, July 17–21). *Experiences of video interviewing in two UK national cohort studies* [Conference presentation]. European Survey Research Association Conference, Milan, Italy.

Scheier, M. F., & Carver, C. S. (1985). The Self-Consciousness Scale: A revised version for use with general populations. *Journal of Applied Social Psychology, 15*(8), 687–699. https://doi.org/10.1111/j.1559-1816.1985.tb02268.x

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), e0128337. https://doi.org/10.1371/journal.pone.0128337

Schober, M. F., Conrad, F. G., Dijkstra, W., & Ongena, Y. P. (2012). Disfluencies and gaze aversion in unreliable responses to survey questions. *Journal of Official Statistics, 28*(4), 555–582. https://www.proquest.com/scholarly-journals/disfluencies-gaze-aversion-unreliable-responses/docview/1266766372/se-2

Schober, M. F., Conrad, F. G., Ehlen, P., & Fricker, S. S. (2003). How web surveys differ from other kinds of user interfaces. In *Proceedings of the Survey Research Methods Section* (pp. 190–195). American Statistical Association. http://www.asasrms.org/Proceedings/y2003/Files/JSM2003-000336.pdf

Schober, M. F., Conrad, F. G., Hupp, A. L., Larsen, K. M., Ong, A. R., & West, B. T. (2020). Design considerations for live video survey interviews. *Survey Practice, 13*(1). https://doi.org/10.29115/SP-2020-0014

Schober, M. F., Okon, S., Conrad, F. G., Hupp, A. L., Ong, A. R., & Larsen, K. M. (2023). Predictors of willingness to participate in survey interviews conducted by live video. *Technology, Mind, and Behavior, 4*(2). https://doi.org/10.1037/tmb0000100

Seitz, J., Benke, I., Heinzl, A., & Maedche, A. (2024). The impact of video meeting systems on psychological user states: A state-of-the-art review. *International Journal of Human-Computer Studies, 182*, 103178. https://doi.org/10.1016/j.ijhcs.2023.103178

Shin, S. Y., Ulusoy, E., Earle, K., Bente, G., & Van Der Heide, B. (2022). The effects of self-viewing in video chat during interpersonal work conversations. *Journal of Computer-Mediated Communication, 28*(1), zmac028. https://doi.org/10.1093/jcmc/zmac028

Shockley, K. M., Gabriel, A. S., Robertson, D., Rosen, C. C., Chawla, N., Ganster, M. L., & Ezerins, M. E. (2021). The fatiguing effects of camera use in virtual meetings: A within-person field experiment. *Journal of Applied Psychology, 106*(8), 1137–1155. https://doi.org/10.1037/apl0000948

Thórólfsson, Æ., Jónsdóttir, G. A., Hjaltason, Á. B., & Guðmundsson, H. (2023, July 17–21). *Mode effects of video interviewing as a proxy of face-to-face interviewing in R10 of ESS in Iceland* [Conference presentation]. European Survey Research Association Conference, Milan, Italy.

Tien, I. S., Imundo, M. N., & Bjork, E. L. (2023). Viewing oneself during synchronous online learning increases appearance anxiety and decreases memory for lecture content. *Applied Cognitive Psychology, 37*(2), 443–451. https://doi.org/10.1002/acp.4048

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

Zavala-Rojas, D., Korkut Keles, O., Schwarz, H., & Romanin, E. (2023, July 17–21). *Measurement invariance of social trust and attitudes towards immigration across face-to-face and video-interviews in ESS Round 10* [Conference presentation]. European Survey Research Association Conference, Milan, Italy.

Appendix A

Survey Interview Questions

1. Would you say that in general your health is excellent, very good, good, fair, poor, don't know, or you're not sure?

- ☐ Excellent
- ☐ Very good
- ☐ Good
- ☐ Fair
- ☐ Poor
- ☐ Don't know
- ☐ Not sure

2. A routine checkup is a general physical exam, not an exam for a specific illness, injury, or condition. When was the last time you saw a doctor for a routine checkup?

- ☐ Within the past 6 months (any time less than 6 months ago)
- ☐ Within the last year (more than 6 months but less than 12 months ago)
- ☐ Within the past 2 years (more than 1 year but less than 2 years ago)
- ☐ Within the past 5 years (more than 2 years but less than 5 years ago)
- ☐ 5 or more years ago
- ☐ Don't know
- ☐ Not sure
- ☐ Never

3. How would you describe your weight?

- ☐ Very underweight
- ☐ Slightly underweight
- ☐ About the right weight
- ☐ Slightly overweight
- ☐ Very overweight

4. How many of the past seven days did you do exercise that made you sweat and breathe hard, such as running, swimming, dancing, or other physical activity?

☐ 7 out of 7 days
☐ 6 out of 7 days
☐ 5 out of 7 days
☐ 4 out of 7 days
☐ 3 out of 7 days
☐ 2 out of 7 days
☐ 1 out of 7 days
☐ 0 out of 7 days

5. Thinking about nutrition, about how many total servings of fruit and/or vegetables did you eat yesterday? A serving would equal one medium apple, a handful of broccoli, or a cup of carrots.

_________ servings

6. Have you smoked at least 100 cigarettes in your entire life?

☐ Yes
☐ No
☐ Don't know/Not sure

7. Have you ever been tested for HIV? Do not count tests you may have had as part of a blood donation?

☐ Yes
☐ No
☐ Don't know/Not sure

8. How many sex partners have you had in the last 12 months?

_________ partners

9. Have your sexual partners in the past 12 months been exclusively male, exclusively female, or both male and female?

☐ Exclusively male
☐ Exclusively female
☐ Both male and female
☐ I've had no sex partners in the last 12 months

10. About how often did you have sex in the past 12 months?

☐ Not at all
☐ Once or twice
☐ Once a month
☐ 2 – 3 times a month
☐ Weekly
☐ 2 – 3 times per week
☐ 4 or more times per week

11. Now thinking about the time since your eighteenth birthday, how many female partners have you had sex with?

________ female partners

12. Now thinking about the time since your eighteenth birthday, how many male partners have you had sex with?

________ male partners

13. Which term best describes how you think of yourself?

☐ Heterosexual or straight
☐ Homosexual, gay, or lesbian
☐ Bisexual

14. A drink of alcohol is one can or bottle of beer, one glass of wine, one can or bottle of wine cooler, one cocktail, or one shot of liquor. During the past 30 days, how many days per week did you have at least one drink of any alcoholic beverage?

________ day(s)

15. A drink of alcohol is one can or bottle of beer, one glass of wine, one can or bottle of wine cooler, one cocktail, or one shot of liquor. During the past 30 days, how many days per month did you have at least one drink of any alcoholic beverage?

________ day(s)

16. Considering all types of alcoholic beverages, how many times during the past 30 days did you have more than 5 drinks on one occasion?

________ time(s)

17. Thinking back over the last 12 months, about how regularly did you drink alcoholic beverages?

- ☐ Never in those 12 months
- ☐ 1 to 3 times in 12 months
- ☐ 4 to 7 times in 12 months
- ☐ 8 to 11 times in 12 months
- ☐ 1 to 3 times a month
- ☐ Once or twice a week
- ☐ 3 to 4 times per week
- ☐ 5 times a week or more

18. Again, as you think back over the last 12 months, how many drinks would you have on a typical day when you drank?

________ drink(s)

19. During the past 12 months, have you read novels, short stories, poems, or plays, other than those required by work or school?

- ☐ Yes
- ☐ No
- ☐ Don’t know/Not sure

20. How often do you read the newspaper? By ‘newspaper’ we mean the content no matter whether it appears in print, online or on a mobile device. Count only newspapers from recognized journalistic outlets. Every day, a few times a week, once a week, less than once a week, or never?

- ☐ Every day
- ☐ A few times a week
- ☐ Once a week
- ☐ Less than once a week
- ☐ Never

21. On the average day, about how many hours do you personally watch television?

________ hour(s)

22. I'll now ask about some different kinds of television shows. Would you tell me how often you watch prime-time drama or situation comedy programs? Would you say every day, several times a week, several times a month, rarely, or never?

☐ Every day
☐ Several times a week
☐ Several times a month
☐ Rarely
☐ Never

23. How often do you watch world or national news programs?

☐ Every day
☐ Several times a week
☐ Several times a month
☐ Rarely
☐ Never

24. How often do you attend religious services? At least once a week, almost every week, about once a month, seldom or never?

☐ At least once a week
☐ Almost every week
☐ About once a month
☐ Seldom
☐ Never

25. Which of the following statements best describes your current employment status?

☐ Employed for wages
☐ Self employed
☐ Out of work more than one year
☐ Out of work less than one year
☐ A homemaker
☐ A student
☐ Retired
☐ Unable to work

26. Now I have some questions about credit cards and charge cards. Do you have any credit cards or charge cards? Please do not include debit cards.

☐ Yes

☐ No

☐ Don’t know/Not sure

27. Are any of the cards you have any type of Visa, Master Card, Discover, or American Express cards you can pay off over time? (Do not include regular American Express charge cards that must be paid in full.)

☐ Yes

☐ No

☐ Don’t know/Not sure

28. How many do you have? Please do not count duplicate cards for the same account or any business or company accounts.

☐ card(s)

29. On your last bills, roughly how much were the new charges made to these accounts?

_________ $.00

30. After the last payments were made on these accounts, roughly what was the balance still owed on these accounts?

____

31. Thinking only about Visa, Master Card, Discover, American Express cards you can pay off over time, and store cards, do you almost always, sometimes, or hardly ever pay off the total balance owed on the account each month?

☐ Almost always

☐ Sometimes

☐ Hardly ever

32. Now I'd like to ask you some questions about your attitudes about savings. People have different reasons for saving, even though they may not be saving all the time. Which of the following statements comes closest to describing your saving habits?

☐ Don't save – usually spend more than income
☐ Don't save – usually spend about as much as income
☐ Save whatever is left over at the end of the month – no regular plan
☐ Save income of one family member, spend the other
☐ Spend regular income, save occasional other income
☐ Save regularly by putting money aside each month

33. Over the past year, would you say that your spending exceeded your income, that it was about the same as your income, or that you spent less than your income?

☐ Spending exceeded income
☐ Spending equaled income
☐ Spending was less than income

34. Some people seem to follow what's going on in government and public affairs most of the time, whether there's an election going on or not. Others aren't that interested. Would you say you follow what's going on in government and public affairs most of the time, some of the time, only now and then, or hardly at all?

☐ Most of the time
☐ Some of the time
☐ Only now and then
☐ Hardly at all

35. In 2012, you may remember that Barack Obama ran for President on the Democratic ticket against Mitt Romney for the Republicans. Did you vote in that election?

☐ Yes
☐ No
☐ Not eligible to vote in that election
☐ Not sure

36. What about local elections – do you always vote in those, do you sometimes miss one, do you rarely vote, or do you never vote?

☐ Always vote
☐ Sometimes miss one
☐ Rarely vote
☐ Never vote

37. In the past three or four years, have you attended any political meetings or rallies?

☐ Yes
☐ No
☐ Not sure

38. In the past three or four years, have you contributed money to a political party or candidate or to any other political cause?

☐ Yes
☐ No
☐ Not sure

39. During the past 12 months, how often have you donated blood?

☐ More than once a week
☐ Once a week
☐ Once a month
☐ At least two or three times a year
☐ Once in the past year
☐ Not at all in the past year
☐ Don't know/Not sure

40. During the past 12 months, how often have you given food or money to a homeless person?

☐ More than once a week
☐ Once a week
☐ Once a month
☐ At least two or three times a year
☐ Once in the past year
☐ Not at all in the past year
☐ Don't know/Not sure

41. During the past 12 months, how often have you done volunteer work for a charity?

- ☐ More than once a week
- ☐ Once a week
- ☐ Once a month
- ☐ At least two or three times a year
- ☐ Once in the past year
- ☐ Not at all in the past year
- ☐ Don’t know/Not sure

42. During the past 12 months, how often have you given money to a charity?

- ☐ More than once a week
- ☐ Once a week
- ☐ Once a month
- ☐ At least two or three times a year
- ☐ Once in the past year
- ☐ Not at all in the past year
- ☐ Don’t know/Not sure

43. During the past 12 months, how often have you offered your seat on a bus or subway, or in a public place to a stranger who was standing?

- ☐ More than once a week
- ☐ Once a week
- ☐ Once a month
- ☐ At least two or three times a year
- ☐ Once in the past year
- ☐ Not at all in the past year
- ☐ Don’t know/Not sure

Appendix B

Annotation and Coding

In order to prepare survey and gaze data for analysis, paid research assistants and additional student volunteers from The New School were recruited to mark the start times of each Q-A sequence in the video-recorded interviews (as unfortunately, Camtasia video timestamps, Qualtrics timestamps, and gaze timestamps were not sufficiently synchronized to allow direct merging across the data sources). The research assistants first recorded each video's start time and then, using Adobe Premiere, they marked the start of each of the 43 questions. Assistants were told that the start of each question should be identified by the start of the interviewer's utterance (i.e., including fillers, disfluencies, deviations from question verbiage "so, um"). That is, the Q-A sequence goes from the start of the question to the beginning of the next question. The end of the last question was also marked. That was defined as the moment the interviewer enters the participant's response.[8] The markers were then exported to an Excel document that documented the time elapsed (to the millisecond) since start of video, for each marker. From this Excel sheet, we were able to generate a time stamp for the start of each question (and therefore, the Q-A sequence) using SPSS's date/time computation wizard. The timestamps needed to be linked with the stream of gaze behavior from the eye tracker, which generated x and y coordinates of gaze at a rate of 30 frames per second.

A computer programmer from Parsons School of Design used Python to create a file that connected each row of gaze data with the appropriate question for each participant. The resolution of gaze regions needed for this study's analytic purposes was not high (looking at self-view region, looking at interviewer/screen, looking elsewhere), and so establishing the linkage, once the file was complete, was relatively straightforward given the tracker's capabilities.

8 Note that with this annotation procedure, the final Q-A sequence is therefore shortened compared to the rest, but this is exactly the same across both experimental groups.

Appendix C

Classification of Survey Interview Questions As Sensitive

Using a more inclusive categorization that classifies questions as sensitive if other prior studies have identified mode differences in response distributions (e.g., Tourangeau et al., 1997), we examined another categorization of sensitive vs. nonsensitive questions (Column 1, other studies). This yielded a total of 21 sensitive questions and 22 nonsensitive and included the question of religious service attendance based on the findings from Study 1. Results show a similar pattern for gaze elsewhere (and thereby gaze at screen), but no effect on gaze at the self-view window. Specifically, using this classification, respondents with a self-view window averted their gaze less during sensitive than nonsensitive questions (and they gazed more at the screen during sensitive than nonsensitive questions).

Thus far, the categorization of questions as sensitive or not was based on prior findings of interview mode effects on socially desirable responding (e.g., FTF vs. ACASI). However, other questions in this survey not identified through this categorization may well be embarrassing to be asked, and it is plausible that they could lead to socially desirable responding (e.g., frequency of sex). In order to explore this further, we carried out another analysis this time using judgments by the Qualtrics Panel to classify questions as sensitive or nonsensitive.

As a first pass, we classified questions as sensitive if more than 50% of the Qualtrics panel members rated a question as one that they thought people would find "somewhat" or "very" embarrassing to be asked. We also compared alternate classifications using cutoffs of 40%, 35%, and 30% of the Qualtrics panel rating a question as somewhat or very embarrassing to answer. Different cutoffs yield different classifications of the questions. Under all these different cutoffs, there was no evidence that respondent gaze (at self, screen, or elsewhere) differed for sensitive vs. nonsensitive questions.

A classification table of questions as sensitive based on these different strategies can be found below.

Table A1 Survey interview question sensitivity

Interview question	Other studies	Lind et al.	Panel judgments			
			at 50%	at 40%	at 35%	at 30%
1. Would you say that in general your health is excellent, very good, good, fair, poor, don't know, or you're not sure?					X	X
2. A routine checkup is a general physical exam, not an exam for a specific illness, injury, or condition. When was the last time you saw a doctor for a routine checkup?				X	X	X
3. How would you describe your weight?	X		X	X	X	X
4. How many of the past seven days did you do exercise that made you sweat and breathe hard, such as running, swimming, dancing, or other physical activity?	X		X	X	X	X
5. Thinking about nutrition, about how many total servings of fruit and/or vegetables did you eat yesterday? A serving would equal one medium apple, a handful of broccoli, or a cup of carrots.			X	X	X	X
6. Have you smoked at least 100 cigarettes in your entire life?	X	X	X	X	X	X
7. Have you ever been tested for HIV? Do not count tests you may have had as part of a blood donation?	X		X	X	X	X
8. How many sex partners have you had in the last 12 months?	X		X	X	X	X
9. Have your sexual partners in the past 12 months been exclusively male, exclusively female, or both male and female?			X	X	X	X
10. About how often did you have sex in the past 12 months?			X	X	X	X
11. Now thinking about the time since your eighteenth birthday, how many female partners have you had sex with?	X	X	X	X	X	X
12. Now thinking about the time since your eighteenth birthday, how many male partners have you had sex with?	X	X	X	X	X	X
13. Which term best describes how you think of yourself?				X	X	X

Interview question	Other studies	Lind et al.	Panel judgments			
			at 50%	at 40%	at 35%	at 30%
14. A drink of alcohol is one can or bottle of beer, one glass of wine, one can or bottle of wine cooler, one cocktail, or one shot of liquor. During the past 30 days, how many days per week did you have at least one drink of any alcoholic beverage?	X			X	X	X
15. A drink of alcohol is one can or bottle of beer, one glass of wine, one can or bottle of wine cooler, one cocktail, or one shot of liquor. During the past 30 days, how many days per month did you have at least one drink of any alcoholic beverage?	X			X	X	X
16. Considering all types of alcoholic beverages, how many times during the past 30 days did you have more than 5 drinks on one occasion?	X		X	X	X	X
17. Thinking back over the last 12 months, about how regularly did you drink alcoholic beverages?	X	X	X	X	X	X
18. Again, as you think back over the last 12 months, how many drinks would you have on a typical day when you drank?	X			X	X	X
19. During the past 12 months, have you read novels, short stories, poems, or plays, other than those required by work or school?						
20. How often do you read the newspaper? By "newspaper" we mean the content no matter whether it appears in print, online or on a mobile device. Count only newspapers from recognized journalistic outlets. Every day, a few times a week, once a week, less than once a week, or never?	X	X				
21. On the average day, about how many hours do you personally watch television?	X	X				
22. I'll now ask about some different kinds of television shows. Would you tell me how often you watch prime-time drama or situation comedy programs? Would you say every day, several times a week, several times a month, rarely, or never?	X	X				

Interview question	Other studies	Lind et al.	Panel judgments			
			at 50%	at 40%	at 35%	at 30%
23. How often do you watch world or national news programs?						
24. How often do you attend religious services? At least once a week, almost every week, about once a month, seldom or never?	X				X	X
25. Which of the following statements best describes your current employment status?					X	X
26. Now I have some questions about credit cards and charge cards. Do you have any credit cards or charge cards? Please do not include debit cards.	X	X				X
27. Are any of the cards you have any type of Visa, Master Card, Discover, or American Express cards you can pay off over time? (Do not include regular American Express charge cards that must be paid in full.)						X
28. How many do you have? Please do not count duplicate cards for the same account or any business or company accounts.					X	X
29. On your last bills, roughly how much were the new charges made to these accounts?					X	X
30. After the last payments were made on these accounts, roughly what was the balance still owed on these accounts?				X	X	X
31. Thinking only about Visa, Master Card, Discover, American Express cards you can pay off over time, and store cards, do you almost always, sometimes, or hardly ever pay off the total balance owed on the account each month?				X	X	X
32. Now I'd like to ask you some questions about your attitudes about savings. People have different reasons for saving, even though they may not be saving all the time. Which of the following statements comes closest to describing your saving habits?	X	X		X	X	X

Interview question	Other studies	Lind et al.	Panel judgments at 50%	at 40%	at 35%	at 30%
33. Over the past year, would you say that your spending exceeded your income, that it was about the same as your income, or that you spent less than your income?				X	X	X
34. Some people seem to follow what's going on in government and public affairs most of the time, whether there's an election going on or not. Others aren't that interested. Would you say you follow what's going on in government and public affairs most of the time, some of the time, only now and then, or hardly at all?				X	X	X
35. In 2012, you may remember that Barack Obama ran for President on the Democratic ticket against Mitt Romney for the Republicans. Did you vote in that election?					X	X
36. What about local elections – do you always vote in those, do you sometimes miss one, do you rarely vote, or do you never vote?					X	X
37. In the past three or four years, have you attended any political meetings or rallies?						X
38. In the past three or four years, have you contributed money to a political party or candidate or to any other political cause?						X
39. During the past 12 months, how often have you donated blood?						X
40. During the past 12 months, how often have you given food or money to a homeless person?	X	X				X
41. During the past 12 months, how often have you done volunteer work for a charity?	X				X	X
42. During the past 12 months, how often have you given money to a charity?				X	X	X
43. During the past 12 months, how often have you offered your seat on a bus or subway, or in a public place to a stranger who was standing?	X	X				X

Appendix D

Table A2 Qualtrics panel ratings of sensitive survey interview responses (> 40%)

Question and response options	Sensitive? (Y/N)
1. Would you say that in general your health is excellent, very good, good, fair, poor, don't know, or you're not sure?	
Excellent	N
Very good	N
Good	N
Fair	Y
Poor	Y
Don't know	Y
Not sure	Y
2. A routine checkup is a general physical exam, not an exam for a specific illness, injury, or condition. When was the last time you saw a doctor for a routine checkup?	
Within the past 6 months (any time less than 6 months ago)	N
Within the last year (more than 6 months but less than 12 months ago)	N
Within the past 2 years (more than 1 year but less than 2 years ago)	Y
Within the past 5 years (more than 2 years but less than 5 years ago)	Y
5 or more years ago	Y
Don't know	Y
Not sure	Y
Never	Y
3. How would you describe your weight?	
Very underweight	Y
Slightly underweight	Y
About the right weight	N
Slightly overweight	Y
Very overweight	Y
4. How many of the past seven days did you do exercise that made you sweat and breathe hard, such as running, swimming, dancing, or other physical activity?	
7 out of 7 days	N
6 out of 7 days	N
5 out of 7 days	N
4 out of 7 days	N
3 out of 7 days	Y
2 out of 7 days	Y
1 out of 7 days	Y
0 out of 7 days	Y

Question and response options	Sensitive? (Y/N)
5. Thinking about nutrition, about how many total servings of fruit and/or vegetables did you eat yesterday? A serving would equal one medium apple, a handful of broccoli, or a cup of carrots.	
0	Y
1–2	Y
3+	N
6. Have you smoked at least 100 cigarettes in your entire life?	
Yes	Y
No	N
7. Have you ever been tested for HIV? Do not count tests you may have had as part of a blood donation?	
Yes	Y
No	Y
Don't know/Not sure	Y
8. How many sex partners have you had in the last 12 months?	
0	Y
1	Y
2+	Y
9. Have your sexual partners in the past 12 months been exclusively male, exclusively female, or both male and female?	
Exclusively male	Y
Exclusively female	Y
Both male and female	Y
I've had no sex partners in the last 12 months	Y
10. About how often did you have sex in the past 12 months?	
Not at all	Y
Once or twice	Y
Once a month	Y
2–3 times a month	Y
Weekly	Y
2–3 times per week	Y
4 or more times per week	Y
11. Now thinking about the time since your eighteenth birthday, how many female partners have you had sex with?	
0	Y
1–5	Y
6+	Y

Question and response options	Sensitive? (Y/N)
12. Now thinking about the time since your eighteenth birthday, how many male partners have you had sex with?	
0	Y
1 – 5	Y
6+	Y
13. Which term best describes how you think of yourself?	
Heterosexual or straight	N
Homosexual, gay, or lesbian	Y
Bisexual	Y
14. A drink of alcohol is one can or bottle of beer, one glass of wine, one can or bottle of wine cooler, one cocktail, or one shot of liquor. During the past 30 days, how many days per week did you have at least one drink of any alcoholic beverage?	
0	N
1 – 2	N
3+	Y
15. A drink of alcohol is one can or bottle of beer, one glass of wine, one can or bottle of wine cooler, one cocktail, or one shot of liquor. During the past 30 days, how many days per month did you have at least one drink of any alcoholic beverage?	
0	N
1 – 5	Y
6+	Y
16. Considering all types of alcoholic beverages, how many times during the past 30 days did you have more than 5 drinks on one occasion?	
0	N
1 – 2	Y
3+	Y
17. Thinking back over the last 12 months, about how regularly did you drink alcoholic beverages?	
Never in those 12 months	N
1 to 3 times in 12 months	N
4 to 7 times in 12 months	Y
8 to 11 times in 12 months	Y
1 to 3 times a month	Y
Once or twice a week	Y
3 to 4 times per week	Y
5 times a week or more	Y
18. Again, as you think back over the last 12 months, how many drinks would you have on a typical day when you drank?	
0	N
1	Y
2+	Y

Question and response options	Sensitive? (Y/N)
19. During the past 12 months, have you read novels, short stories, poems, or plays, other than those required by work or school?	
Yes	N
No	Y
20. How often do you read the newspaper? By "newspaper" we mean the content no matter whether it appears in print, online or on a mobile device. Count only newspapers from recognized journalistic outlets. Every day, a few times a week, once a week, less than once a week, or never?	
Every day	N
A few times a week	N
Once a week	N
Less than once a week	Y
Never	Y
21. On the average day, about how many hours do you personally watch television?	
0	N
1 – 3	N
4+	Y
22. I'll now ask about some different kinds of television shows. Would you tell me how often you watch prime-time drama or situation comedy programs? Would you say every day, several times a week, several times a month, rarely, or never?	
Every day	N
Several times a week	N
Several times a month	N
Rarely	N
Never	N
23. How often do you watch world or national news programs?	
Every day	N
Several times a week	N
Several times a month	N
Rarely	N
Never	Y
24. How often do you attend religious services? At least once a week, almost every week, about once a month, seldom or never?	
At least once a week	N
Almost every week	N
About once a month	Y
Seldom	Y
Never	Y

Question and response options	Sensitive? (Y/N)
25. Which of the following statements best describes your current employment status?	
Employed for wages	N
Self employed	N
Out of work more than one year	Y
Out of work less than one year	Y
A homemaker	Y
A student	N
Retired	N
Unable to work	Y
26. Now I have some questions about credit cards and charge cards. Do you have any credit cards or charge cards? Please do not include debit cards.	
Yes	N
No	Y
27. Are any of the cards you have any type of Visa, Master Card, Discover, or American Express cards you can pay off over time? (Do not include regular American Express charge cards that must be paid in full.)	
Yes	N
No	N
28. How many do you have? Please do not count duplicate cards for the same account or any business or company accounts.	
0	N
1–2	N
3+	Y
29. On your last bills, roughly how much were the new charges made to these accounts?	
$0	N
$1–249	Y
$250+	Y
30. After the last payments were made on these accounts, roughly what was the balance still owed on these accounts?	
$0	N
$1–499	Y
$500+	Y
31. Thinking only about Visa, Master Card, Discover, American Express cards you can pay off over time, and store cards, do you almost always, sometimes, or hardly ever pay off the total balance owed on the account each month?	
Almost always	N
Sometimes	Y
Hardly ever	Y

Question and response options	Sensitive? (Y/N)
32. Now I'd like to ask you some questions about your attitudes about savings. People have different reasons for saving, even though they may not be saving all the time. Which of the following statements comes closest to describing your saving habits?	
Don't save – usually spend more than income	Y
Don't save – usually spend about as much as income	Y
Save whatever is left over at the end of the month – no regular plan	Y
Save income of one family member, spend the other	Y
Spend regular income, save occasional other income	N
33. Over the past year, would you say that your spending exceeded your income, that it was about the same as your income, or that you spent less than your income?	
Spending exceeded income	Y
Spending equaled income	Y
Spending was less than income	N
34. Some people seem to follow what's going on in government and public affairs most of the time, whether there's an election going on or not. Others aren't that interested. Would you say you follow what's going on in government and public affairs most of the time, some of the time, only now and then, or hardly at all?	
Most of the time	N
Some of the time	Y
Only now and then	Y
Hardly at all	Y
35. In 2012, you may remember that Barack Obama ran for President on the Democratic ticket against Mitt Romney for the Republicans. Did you vote in that election?	
Yes	N
No	Y
Not eligible to vote in that election	Y
36. What about local elections – do you always vote in those, do you sometimes miss one, do you rarely vote, or do you never vote?	
Always vote	N
Sometimes miss one	N
Rarely vote	Y
Never vote	Y
37. In the past three or four years, have you attended any political meetings or rallies?	
Yes	N
No	Y
38. In the past three or four years, have you contributed money to a political party or candidate or to any other political cause?	
Yes	N
No	N

Question and response options	Sensitive? (Y/N)
39. During the past 12 months, how often have you donated blood?	
More than once a week	N
Once a week	N
Once a month	N
At least two or three times a year	N
Once in the past year	Y
Not at all in the past year	Y
40. During the past 12 months, how often have you given food or money to a homeless person?	
More than once a week	N
Once a week	N
Once a month	N
At least two or three times a year	N
Once in the past year	Y
Not at all in the past year	Y
41. During the past 12 months, how often have you done volunteer work for a charity?	
More than once a week	N
Once a week	N
Once a month	N
At least two or three times a year	N
Once in the past year	Y
Not at all in the past year	Y
42. During the past 12 months, how often have you given money to a charity?	
More than once a week	N
Once a week	N
Once a month	N
At least two or three times a year	Y
Once in the past year	Y
Not at all in the past year	Y
43. During the past 12 months, how often have you offered your seat on a bus or subway, or in a public place to a stranger who was standing?	
More than once a week	N
Once a week	N
Once a month	N
At least two or three times a year	N
Once in the past year	Y
Not at all in the past year	Y

methods, data, analyses | Vol. 20(1), 2026, pp. 179-208 DOI: 10.25521/mda.663

Live Video Interviewing as a Complementary Mode to In-Person Interviews: Evidence from the European Social Survey

Tim Hanson[1], Roberto Briceno-Rosas[2], Mare Ainsaar[3] & Hanna Kerstina Kartau[3]

[1] *City St George's, University of London*
[2] *GESIS – Leibniz Institute for the Social Sciences*
[3] *University of Tartu*

Abstract

Live video interviewing emerged as a method for collecting survey data during the COVID-19 pandemic, having rarely been used for survey data collection prior to this. There is now a need to assess experiences and outcomes from studies that utilised video interviewing, partly with a view to informing the future feasibility of the method in different contexts. This paper reports on the experience of the European Social Survey (ESS) with video interviewing, having used this approach as a complementary method to in-person interviewing at its 10th round (2020–2022). The ESS can provide a unique perspective, being the first cross-national survey to use video interviews. In total, 16 countries offered video interviewing alongside in-person interviewing at ESS Round 10. In this paper, we present a range of results based on ESS Round 10 in two main categories. We first look at the effectiveness of the implementation of video interviewing and then compare quality between video interviews and in-person interviews across various indicators, including interviewer effects. The results show that the prevalence of video interviews varied widely between countries, likely relating to national contextual factors. However, in countries where a large share of video interviews was carried out, we found that the interview experience was rated positively, and quality indicators were closely comparable with in-person interviews. These results suggest that future use of video interviewing may be more feasible in some countries that others, but in certain contexts it has the potential to offer an effective complementary option to in-person interviewing.

Keywords: live video interviewing, cross-national surveys, European Social Survey, interviewer effects

© The Author(s) 2025. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

The COVID-19 pandemic prompted several surveys to use live video interviewing for quantitative survey data collection for the first time. This included the European Social Survey (ESS), which used live video interviews as a complementary method to in-person interviewing in several countries for its 10th Round, carried out between 2020 and 2022. This paper reports on the experiences and outcomes of live video interviews for ESS[1].

Video Interviewing for Surveys

While use of video interviewing for surveys was rare prior to the pandemic, it had long been foreseen as a potentially promising method. Anderson (2008) considered the potential of video-mediated surveys as a future approach and saw several benefits. These included the scope for greater rapport and engagement with respondents compared with telephone and web surveys, and the potential to appeal to certain population sub-groups that are sometimes underrepresented in surveys (e.g., younger people). However, challenges were also foreseen. These included the risk that the presence of an interviewer, even remotely, may reduce candour in reporting sensitive behaviours compared to a self-administered mode like web surveys, and challenges with some population groups accessing video platforms, including older people and those less technologically experienced. There could also be challenges with low internet connectivity in certain regions. Anderson concluded that "research on [video's] detailed impacts on real survey interviews is urgently needed. Pilot studies of this kind should be undertaken before any widespread adoption is planned" (Anderson, 2008, p. 115).

While video platforms have been used for qualitative research for some time (e.g., Irani, 2019), they were rarely used for quantitative survey data collection prior to the pandemic. This resulted in some studies adopting video interviewing as an approach at speed, and sometimes without the extensive piloting that Anderson (2008) recommended. While publications relating to video interviewing are still rare, there are now several examples of its use that provide early indicators of experiences and outcomes associated with the method.

There is evidence of a good take-up rate of video interviews in some contexts, particularly for longitudinal studies (Dulaney et al., 2023; Sanchez et al., 2023). A study by Conrad et al. (2023) found live video respondents were less likely to give non-differentiated responses and reported higher satisfaction than web respon-

1 In the remainder of this paper, we generally use 'video interviews' when referring to 'live video interviews' for brevity. In a few places we refer to 'live video interviews' to differentiate from pre-recorded video interviews.

Direct correspondence to
Tim Hanson, City St George's, University of London, London, United Kingdom
E-mail: tim.hanson@citystgeorges.ac.uk

dents, suggesting the presence of a (remote) interviewer can keep respondents motivated and conscientious. Carr et al. (2023) also reported positive feedback on the experience of video interviews from both respondents and interviewers, with technical issues with the video platform being quite rare. Some studies have observed sample composition differences between video and in-person modes, which may point to the potential for this mode to bring in underrepresented groups (e.g., Dulaney et al., 2023; Þórólfsson et al., 2023).

There are so far relatively few studies that compare measurement between in-person and video modes, but there are some encouraging early findings. Kelley et al. (2023) found no measurement effects from introducing video interviewing as a mode of data collection, while Zavala-Rojas et al. (2023) looked at two concepts measured in the European Social Survey and found generally consistent relationships between variables for video and in-person modes.

Regarding interviewer-respondent interactions, Sun et al. (2021) found no significant difference in respondents' rapport ratings between video-mediated and in-person interviews, suggesting that rapport is just as well established through remote video interviewing. Kelley et al. (2023) found that video interviewing was akin to in-person interviewing, with video interviewers being adept at maintaining the meaning of questions. West et al. (2022) found little evidence of significant interviewer effects for either live or pre-recorded video interviews. They recommended that future studies should compare interviewer effects between video interviews and in-person interviews.

Endres at al. (2023) randomised respondents to either interviewer-administered video or interviewer-administered in-person modes after completing a self-administered online survey wave. They found that video interviewing is more comparable to in-person interviewing than online interviewing across multiple measures of satisficing, social desirability and respondent satisfaction. This is of particular relevance to this paper, since ESS used video interviewing as a complementary mode to in-person interviewing, and combined interviews from each approach in its published data set.

While the above examples provide reassurance and support for use of video interviewing, other studies raise some concerns and limitations. Despite different sample compositions between video and in-person interviewing, video may not lead to a better response rate or improved sample composition compared with a solely in-person approach (Þórólfsson et al., 2023). Take-up of video interviews, when offered alongside other modes, can sometimes be low (Sanchez et al., 2023). Some groups of respondents or households may be less open to video interviewing, while technical aspects of administering video interviews can be more challenging for interviewers (Centeno et al., 2023). Carr et al. (2023) found examples of differential reporting between video and other modes. And Conrad et al. (2023) found higher levels of rounding for numerical questions and more socially desirable answers for live video respondents compared with web respondents. It's clear from the studies completed to date that evidence is mixed, and that there are different ways that the success, or otherwise, of video interviewing can be assessed.

Contributing to the Existing Evidence

This paper aims to investigate the potential use of video interviewing as a complementary mode to in-person interviewing by analysing data from the 10th Round of the European Social Survey (ESS).

The ESS is conducted as a cross-sectional biennial survey across over 30 countries. Until Round 10, all countries were required to collect fully in-person samples based on a central specification. As with other surveys that rely on in-person data collection, the COVID-19 pandemic posed several challenges to the ESS's usual approach. This resulted in some modifications being made to the specification for ESS Round 10 (carried out in 2020-2022; Hanson et al., 2022), including, for the first time in the ESS's history, allowing video interviewing as a complementary mode to in-person interviewing.

The specific use of video interviewing as a complementary mode to in-person interviewing is an important feature of our study, in two respects. First, there was no attempt to switch to a full video-interviewing approach and replace in-person interviewing. It was not expected that a video-only approach could deliver a representative sample for a cross-sectional, general population survey such as the ESS. The use of video was therefore reliant on in-person interviewing being possible. And second, our approach used two interviewer-administered modes (in-person and video) and was not part of a broader mixed-mode strategy that might include non-interviewer modes (e.g., self-completion web). We can therefore only assess how effectively video interviewing complemented in-person interviewing and not make judgments about the role of video in a wider mixed-mode approach.

There are good reasons why the combination of video and in-person interviewing may present a suitable complementary approach. The sampling design, contact and cooperation processes can remain unchanged. Both modes require interviewers to read out questions and record responses as given by the respondent. The in-person ESS interviews use showcards that display response lists to respondents throughout the interview, and these can be shared with the respondent through the video interview platform. But there are also differences; rather than sharing the same physical space, the interviewer and respondent engage through a remote online platform.

The relative cost-effectiveness of video interviewing compared with in-person interviewing makes it a potentially attractive option. If a large share of interviews that took part in person can move to a remote video approach, this will save significantly on travel costs. Video interviewing may therefore offer significant potential beyond the pandemic for surveys that traditionally rely on in-person interviewing. To inform this, more evidence is needed on the experience of video interviewing, and its comparability with in-person interviewing. The ESS was one of the first studies to introduce video interviewing as a complementary approach to in-person interviewing in a full production survey and so can provide evidence on this point.

To our knowledge, the ESS is also the only cross-national study that has used video interviewing. The challenge of achieving equivalence is greatly magnified for cross-national surveys (Jowell et al., 2007). Adding a new data collection mode may present particular challenges and share new insight in this context. For example, there may be differences in the take-up of video interviewing between countries, which may be linked to contextual factors (e.g., sample frames, contact methods, levels of internet use). Experiences of video interviewing may differ due to variations in approaches used. It's also crucial to understand how video and in-person interviewing compare across quality indicators between countries. If, for example, there are differences in the comparability of modes between countries, and different rates of video interviews are observed between countries, this may compromise the comparability of the cross-national data produced.

Expectations for Video Interviewing as a Complementary Approach to In-Person Interviewing

Determining the effectiveness and worth of introducing a complementary mode in a cross-national survey context requires reflection on several key aspects. A good complementary mode must meet certain minimum requirements to be considered beneficial for implementation. These considerations include understanding the characteristics that define a successful complementary mode and identifying the essential criteria it must fulfil to enhance the survey process effectively across different countries. We consider the two central criteria as an approach to our analysis: (I) the effectiveness of implementations, and (II) the quality and comparability of measurements.

For the implementation of video interviews to be effective as a complementary mode to in-person interviewing, it needs to (1) have a substantial use (in terms of the number and/or share of video interviews conducted) and/or enhance survey response, and (2) provide a comparable interview experience to respondents.

For measurement quality to be comparable between video and in-person interviews, the two methods need to (1) provide similar interviewing conditions (e.g., video replicating the in-person approach in terms of being interviewer-administered and using showcards, but through a different medium), (2) generate similar behaviours to answering the questionnaire by the target respondent (e.g., item nonresponse, level of straightlining), and (3) produce similar interviewer effects on answers. The expectation is that the platform of communication should not significantly influence respondents' answers or the integrity of the data collected.

Approach and Research Questions

The analyses presented in this paper is divided into two parts, both of which are crucial in understanding the effectiveness and suitability of video interviewing. We first look at the *implementation and experience of video interviewing* (see Research Question 1, below). Here we share the prevalence of video interviews conducted in each country, present results from respondent and interviewer 'interview experience' questions, and share findings on technical issues reported.

We then look at the *quality and comparability of video and in-person interviews across modes* (Research Question 2). Here we will make comparisons across two quality metrics: indicators of satisficing, and interviewer effects.

Cross-country comparability is a dimension that goes across both analytical focuses of the paper (Research Question 3). This third dimension allows us to provide a more relative perspective on the findings of video interview as a complementary mode across multiple country-specific contexts and survey conditions.

Through this analysis, our paper seeks to answer the following research questions:

1. How effective is video interviewing in complementing in-person interviewing for large cross-national surveys like the ESS?
2. How did the quality of the interviewing process via video compare to the interviewing process via in-person interviewing in the ESS?
3. How does the assessment of video interviewing vary between countries regarding its implementation, quality, and comparability with in-person interviewing (based on both interviewer and respondent perspectives)?

Data

We use data from the European Social Survey Round 10 (ESS ERIC, 2023), carried out between 2020 and 2022. This was the first ESS round in which video interviewing was offered as a complementary mode to in-person interviewing. This reflected concerns that even where in-person interviewing was possible, some target respondents would be unable or unwilling to take part in an in-person interview due to the COVID-19 pandemic. This may be due to needing to shield themselves for health reasons, for concerns over being infected, or for other reasons.

Since the video interviewing approach was offered as an 'emergency' measure for Round 10, limited development work was carried out to test the method in the ESS context. A small amount of user testing was carried out among ESS team members and in some countries to assess and refine the approach. The approach was informed by best practice guidance based on experience from other studies (e.g., Schober et al., 2020).

In total, 16 countries offered video interviewing. When describing the prevalence of video interviews, we include all 16 countries in our analysis. However, for the remaining analysis, we focus on only the six countries that conducted the largest number of video interviews (Estonia, Finland, Iceland, Italy, Netherlands and Norway). In these countries, between 240 and 491 video interviews were conducted. In all other countries, the number of video interviews was below 100.

ESS's Approach to Video Interviewing

ESS's data collection model is based on decentralized fieldwork contracting and data collection. National teams are appointed in each country, and they either organize fieldwork in-house (where they have an interviewer fieldforce) or contract this activity to a survey agency. For the Round 10 video interviewing approach, each national team was asked to describe their approach, in response to centrally produced guidelines. Their approach was then reviewed and approved by ESS's central coordination team.

The main process for contacting target respondents was unchanged from the usual fully in-person interviewing approach. In the majority of countries, interviewers attempted in-person contact, sometimes following delivery of an advance letter. A video interview could then be offered as an alternative to an in-person interview. A small number of countries had access to named person samples that included telephone numbers for sample members. In these cases, they could attempt first contact by telephone and offer a video interview at that point (meaning in these cases a video interview could be conducted without any in-person contact).

National teams were allowed to select between the following approaches for carrying out video interviews:

- Option A: Establish a specialist team of video interviewers, separate from those carrying out in-person interviews,
- Option B: Allow all their in-person interviewers to also carry out video interviews.

– Option C: Allow a subset of their in-person interviewers to also carry out video interviews.

There were pros and cons with each approach. Option A allowed for more central control and management but led to a more complex flow of steps between the in-person interviewer and the video interviewer. There may also be concerns with interviewer effects if a small number of interviewers were carrying out a large number of interviews—which applies to option A and to some extent C—as the impact of these effects increases with the size of workload per interviewer (West & Blom, 2017). Increasing workload of interviewers would contradict recommendations to reduce interviewer effects in in-person surveys (Schnell & Kreuter, 2005). Option B (and to some extent C) was a more seamless process as the same interviewer was responsible for both the in-person contact and for administering the video interview. However, particularly for option B, there are challenges with training, equipping and monitoring a large interviewer fieldforce to take on the video interviewing task. A little over half of the countries adopted option A: establishing a small specialist team of video interviewers.

National teams were advised to use Microsoft Teams or Zoom to carry out the video interviews. They were required to use a licensed version of the platforms. Respondents were not required to have a Teams/Zoom account in order to take part, or to download any software. In a small number of cases, countries used alternative platforms, subject to these platforms meeting GDPR requirements and tested with respondents in advance of fieldwork.

Interviewers were instructed to use two screens to carry out the video interviews. One screen included the CAPI questionnaire, which was not shared with the respondent. This allowed the interviewer to read the questions and enter responses. The other screen featured the video call, including the interviewer and respondent videos. This screen also included showcards, which are used throughout the ESS questionnaire and were screen-shared with the respondent. The interviewer moved on to the correct card as required throughout the interview.

Respondents could take part using any internet-enabled device, including smartphones. It was recommended that they used a larger screen device, if available, to allow for clearer display of the video questionnaire and showcards. Other features and requirements regarding ESS's video interviewing are described in Appendix A (Table A1).

Methods

We conduct the analysis in two stages: (I) an assessment of the effectiveness of the implementation of the video interviews as a complementary approach to in-person interviewing, and (II) an assessment of the comparability of data quality between in-person and video modes.

We assess the *effectiveness* of video interviewing in the ESS (Stage I) in two main steps.

1. *Prevalence of video interviews:* The first step to assessing the effectiveness of video interviews is to establish the level of demand for them (as opposed to in-person interviews) across ESS countries. If only a small number of respondents opt for video interviews, it may not be worthwhile offering this approach in future as impact on the data and potential cost savings would be limited. We look at the proportion of video interviews carried out in each of the 16 countries that offered this approach. Through this, we assess the different contexts and factors present in each country and consider how this may have impacted on the prevalence of video interviews conducted.

2. *Experience of video interviews:* It is important to gain feedback from both interviewers and respondents on the experience of taking part in a video interview, and assess how this compares to in-person interviews. If the experience of a video interview is judged to be worse than an in-person interview, this may point to underlying issues with the video approach in the context of ESS, or perhaps a need for future improvements in how it is provided. To measure this, we compare responses to an 'interview experience' question that was asked of both respondents and interviewers in video and in-person modes ("*How would you rate the overall experience of taking part in this survey? Please answer on a scale from 0 to 10 where 0 is very negative and 10 is very positive*"). The same question was included for both in-person and video interviews, to allow for a comparison of the experience between modes. For the respondent version, the question was asked to the respondent by the interviewer at the end of the interview. For the interviewer version, the interviewer was asked to record their experience score in a short 'interviewer questionnaire' included at the end of the interview. Clearly, there may be interviewer and respondent level effects that influence the interview experience beyond the mode difference. However, by including these questions we intended to gain a broad impression of whether taking part in a video interview seemed to represent a worse (or improved) interviewing experience compared with in-person interviewing. Two sample t-tests were performed to compare experience scores in the in-person interviewing and video interviewing groups. For video interviews, both respondents and interviewers were asked to report any technical issues experienced during the interview. We also present these results.

Similarly, we will compare video and in-person interviews in different ways, with the objective of assessing the *quality and comparability* (Stage II) of the new video method in comparison to ESS's long-standing in-person interviewing approach.

1. *Satisficing behavior:* Satisficing response behaviour is typical in situations when the respondent is fatigued or unmotivated to answer truthfully. As part of our analysis, we looked at two metrics relating to satisficing behaviour: item non-response and straightlining. Respondents who are less engaged in the interview or less motivated are likely to have higher rates of item non-response and straightlining. We carried out chi-squared tests to assess if these indicators of satisficing differ between in-person and video interviews. For both item non-response and straightlining, we analysed the 21-item Schwartz Human Values Scale[2]. This scale is located at the end of the ESS questionnaire and has been deliberately chosen, since we might expect satisficing behaviours to be more common in the latter part of a long (approx. 1 hour) interview. Respondents selecting the same answer to at least 15 of the 21 human values items are coded as exhibiting straightlining behaviour. The same approach has been previously used by Ainsaar et al. (2013). As these questions ask about the respondent's level of similarity to many different contradicting values, it is unlikely that the respondent would identify with most of the values to the same degree. To compensate for the differences between respondents, we use only data from persons with relatively similar backgrounds: a person living with a partner who has at least a secondary or higher education and who is younger than 60 years old. This allows for in-person sample sizes between 302 and 1,283 per country, and video interviewing sample sizes between 214 and 421 per country.

2. *Interviewer effects:* The communication platform can potentially influence the extent to which interviewers affect responses. The shift to digital communication introduces unique challenges, raising questions about how video-based dynamics, compared to in-person interactions, might alter an interviewer's influence on respondents' answers. West et al. (2022) discuss mechanisms that typically generate interviewer effects in video interviews—such as interviewer behaviour, visible characteristics, and responsiveness to different question types—and how these might vary between live and prerecorded video interviews. These ideas can be applied when comparing video interviews with in-person interviews. For example, real-time interaction in video settings may limit certain nonverbal cues, potentially reducing engagement, although, so far, there is no evidence that video mediated communication changes how interviewers might affect respondents (see Sun & Conrad, 2026 and Sun et al., 2021).

2 Items in the human values scale include: "It is important to her/him to be humble and modest. She/He tries not to draw attention to herself/himself" and "Being very successful is important to her/him. She/He hopes that people will recognise her/his achievements".

Comparing live video interviews to in-person interviews allows us to observe to what extent each mode impacts interviewer effects. For video interviews to serve as an effective complement to in-person interviews, interviewer effects should be small and similar across modes. While standardized protocols aim to maintain consistency and reliability across modes, they may not fully mitigate variability introduced by individual interviewers. Thus, we examine how interviewers may differently affect responses in video versus in-person formats, estimating the variance attributable to interviewer clustering across these modes. To assess possible differences in interviewer effects, we estimate how much of the variance from all responses of each single item can be explained by the clustering within interviewers from in-person interviews versus video interviews. Interviewer effects on single items are estimated by taking the average of the intra-interviewer correlations (ICC) of numeric and ordinal items in the questionnaire. ICCs are estimated from linear models with an interviewer-level random effect for all numeric and ordinal items measured on a scale of at least 4-points in the ESS Round 10 questionnaire. A total of 175 items from the ESS questionnaire were selected as suitable for the analysis. To control for similarities between respondents arising from area effects rather than interviewer effects, the geographical region and self-reported degree of urbanization of respondents' domicile are included in the models[3].

A high ICC indicates that responses from respondents interviewed by the same interviewer are more similar than otherwise would be expected and are suggestive of differences between interviewers in the way they interact with or affect respondents during the interview. Large difference between in-person and video interviews in the ICCs would suggest that the impact of interviewers on responses depends on the interview mode. ICCs are also expected to correlate with other indicators of interviewer behavior, such as speed of interviews (Vandenplas et al., 2019). Due to the small sample size of video interviews and the low number of interviewers conducting video interviews across participating countries, estimates were suppressed for items administered in video interviews if they had fewer than 10 interviewers (cluster units) and a ratio lower than 4 interviews per interviewer (instead of the recommended ratio of 5).[4] It should be noted that this reduced number of clusters might affect the accuracy of the estimates and these results should be read with the necessary caution.

3 Given the lack of random assignment, interviewer and area effects cannot be fully disentangled, and some (presumably small) portion of the intra-interviewer correlations may be attributable to area effects.

4 This is below the typically recommended number of above 30 clusters (Hox, 2010) but within the parameters considered by other authors as the minimum (Hadler, 2004).

Results

Effectiveness of Implementing Video Interviewing

Prevalence of Video Interviews

The number and proportion of video interviews varied substantially between countries. Table 1 shows the prevalence of video interviews across countries, alongside other information on national approaches to video interviewing and contextual details.

Table 1 Prevalence of interviews across countries and other contextual information

Country	Percentage of all interviews done by video	Number of video interviews	Percentage of respondents who use internet at least most days[a]	Who carried out video interviews?	Sampling unit[b]	Contact approach with target respondents
Iceland	37.0%	333	95.3%	All face-to-face interviewers	Individual	Telephone, in-person
Norway	34.8%	491	94.3%	Subset of face-to-face interviewers	Individual	Telephone, in-person
Italy	17.3%	457	69.9%	All face-to-face interviewers	Individual	In-person
Netherlands	16.9%	248	94.1%	Subset of face-to-face interviewers	Individual	In-person
Estonia	15.6%	240	80.5%	Specialist team of video interviewers	Individual	In-person
Finland	15.2%	240	87.0%	Specialist team of video interviewers	Individual	Telephone, in-person
Croatia	6.0%	95	69.7%	Specialist team of video interviewers	Individual	In-person
United Kingdom	4.8%	55	88.5%	Specialist team of video interviewers	Household	In-person
Switzerland	3.3%	50	88.0%	Specialist team of video interviewers	Individual	In-person

Table 1 (continued)

Country	Percentage of all interviews done by video	Number of video interviews	Percentage of respondents who use internet at least most days[a]	Who carried out video interviews?	Sampling unit[b]	Contact approach with target respondents
France	2.3%	46	79.3%	All face-to-face interviewers	Individual	In-person
Belgium	1.2%	16	84.1%	Specialist team of video interviewers	Individual	In-person
Greece	0.8%	23	70.9%	Specialist team of video interviewers	Household	In-person
Portugal	0.4%	8	66.7%	Subset of face-to-face interviewers	Household	In-person
North Macedonia	0.3%	4	71.9%	Specialist team of video interviewers	Household	In-person
Slovakia	0.0%	0	61.9%	Specialist team of video interviewers	Household	In-person
Slovenia	0.0%	0	73.4%	Specialist team of video interviewers	Individual	In-person

[a] Based on responses to Item A2 in the ESS Round 10 questionnaire: *People can use the internet on different devices such as computers, tablets and smartphones. How often do you use the internet on these or any other devices, whether for work or personal use?*

[b] Where a household-based sample frame is used, interviewers were required to randomly select one adult aged 15+ for interview within the household. The selection was carried out via an electronic contact form. Countries could choose between three different selection methods: a random selection via the electronic questionnaire (replicating a Kish grid), the Rizzo method, or next/last birthday.

The countries can broadly be split into three groups based on their prevalence of video interviews. In the first group, two countries—Iceland and Norway—conducted more than a third of all their interviews by video. In the second group, four countries—Italy, Netherlands, Estonia and Finland—conducted around 1 in 6 of their interviews by video (between 15% and 17%). And finally, the remaining 10 countries conducted a much smaller share of video interviews. The range in this last group varies from Croatia, where 6% of interviews were done by video, to Slovakia and Slovenia, where no video interviews were carried out despite this option being offered.

A few features stand out among the countries that carried out a large compared to a smaller share of video interviews.

First, the countries with the largest share of video interviews include those with the highest level of internet use (daily or most days): Iceland, Norway, and the Netherlands. In contrast, levels of internet use were markedly lower among most countries where video interviewing was less productive (Greece, Portugal, North Macedonia, Slovakia, Slovenia). There was a strong positive correlation between the level of daily internet use and the percentage of video interviews conducted in that country ($r = .65$, $p < .01$).

Second, all countries that carried out a relatively large share of video interviews used samples of individuals. This meant they could contact the target respondents by name in advance of an interviewer visiting—in an advance letter and/or, in some cases, a telephone call. Other countries used household samples, which required the interviewer to visit the household and make a person selection to identify the target respondent. It might be the case that by being able to identify and contact the target respondent earlier in the process, the option of conducting the interview by video can be introduced sooner in the recruitment process.

Third, and expanding on the last point, the two countries with the largest proportion of video interviews, Iceland and Norway, were able to make initial contact with sample members by telephone. It seems logical that the video interviewing option is more likely to be selected by the sample member in such cases as it can be offered without needing to send an interviewer to make any in-person contact. Where an interviewer visit is required, there may be a greater tendency among both the interviewer and respondent to agree to an in-person interview, since the interviewer is already there.

There appears to be no clear relationship between the type of interviewer team and the prevalence of video interviews. Of the six countries to include the largest share of video interviews, two used a specialist team of video interviewers (Estonia and Finland), two selected a subset of their face-to-face interviewers to carry out video interviews (Norway and the Netherlands), and two required all of their face-to-face interviewers to carry out video interviews (Iceland and Italy).

We should also note that there are likely to be other factors behind different rates of video interviewing between countries. Italy, for example, has a relatively low level of internet use and did not have the option of making telephone contact, but had the third highest percentage of video interviews. It was evident in discussions with national teams that some were more positive about the video option and put in place greater resources to support it. Other national teams were more neutral (sometimes sceptical) about video interviewing and set it up on a much smaller scale as a last-resort option. Such differences in expectations and operational planning may also partly explain some of the differences between countries.

Sample Composition

Appendix B (Table A2) includes a comparison of sample compositions between the in-person and video interviewing approaches in the six countries where the highest number of video interviews were achieved. We found that, compared with those interviewed in-person, the video interview sample tends to be younger, from smaller households, were more likely to be doing paid work, and tended to have higher levels of education. There is no clear trend regarding difference in sex, legal partnership status, or immigration background.

Interview Duration

Appendix C (Table A3) consists of a table showing the average interview duration between in-person and video interviews in each of the six countries. Differences in duration could emerge due to different interviewers being used for each mode, technical problems, differences in respondent profiles, or for other reasons linked to the speed of administration for each mode. Interview duration can impact the effectiveness of implementing each mode. For example, if video interviews are found to take significantly longer than in-person interviews, this can impact interviewing costs.

In most countries, we found there was no difference between the length of video and in-person interviews. The only exceptions are Estonia and Italy, where the video interviews on average had longer durations than in-person interviews. It is unclear why the video interviews had a longer duration in these two countries.

Interview Experience and Technical Issues

Table 2 shows the mean scores given by respondents and interviewers to the experience question between in-person and video interviews.

Respondent experience scores were mixed between the two modes. Those interviewed in-person in Finland and Norway had a higher mean experience score, i.e. a more positive experience, compared with those interviewed by video, while the reverse was true in Italy. In the other three countries, there was no significant difference between the scores. There was no difference when the experience scores were combined across all six countries.

Based on interviewer-reported experience scores, there were significant differences between modes in five of the six countries. In three countries—Estonia, Iceland and Italy—the mean experience score was higher for video interviews. In two countries—Netherlands and Norway—it was higher for in-person interviews. Across all six countries, the mean interviewer experience score for video interviewing was significantly higher compared with in-person interviewing.

While there were some (generally small) differences between modes, what is evident from these results is that the interview experience scores are high for both modes among respondents and interviewers (averaging at more than 8 out of 10), and there is no consistent evidence that the experience of video interviews is worse compared with in-person interviewing.

Table 2 Interview experience in in-person and video interviews (0 – very negative … 10 – very positive) (sample sizes in parentheses after experience scores)

Country	Respondent					Interviewer				
	In-person interviews		Video interviews		*p*-value	In-person interviews		Video interviews		*p*-value
Estonia	8.04	(1,301)	8.22	(240)	.156	8.75	(1,302)	8.95	(240)	.026*
Finland	8.51	(1,332)	8.05	(240)	<.001***	8.76	(1,335)	8.88	(240)	.134
Iceland	8.54	(563)	8.36	(332)	.060	8.55	(551)	8.82	(325)	.006**
Italy	7.75	(2,163)	7.97	(457)	.009**	7.84	(2,136)	8.37	(456)	<.001***
Netherlands	8.37	(1,215)	8.40	(247)	.757	8.44	(1,220)	8.26	(247)	.014*
Norway	8.51	(918)	8.34	(491)	.047*	9.11	(914)	8.85	(491)	<.001***
All 6 countries	8.19	(7,492)	8.22	(2,007)	.467	8.47	(7,458)	8.68	(1,999)	<.001***

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Respondents and interviewers were also asked to report experiences of any technical issues with the video interviews. Tables 3 (respondents) and 4 (interviewers) show these results by country.

In the majority of cases, both respondents and interviewers reported no technical issues with video interviews. Taking all countries together, the most commonly reported issues were problems starting the video call (reported by respondents in 6.7% of interviews, and interviewers in 7.2% of interviews), with internet connections (6.1%/5.3%), and with audio being unclear (7.5%/5.8%).

The results did vary somewhat between countries. The most notable differences were between Estonia and Italy; respondents reported at least one technical issue in 35% of interviews in Estonia compared with 10% of interviews in Italy.

It was noted earlier that respondents were advised, where possible, to use a device with a larger screen to optimise display of the showcards. However, even among those who completed on smartphones ($n = 450$), only 2.2% of respondents reported issues with seeing or reading the showcards (compared with 1.5% using a desktop or laptop PC and 1.9% using a tablet).

The above findings show that technical issues only occurred in a minority of cases, which may be seen as a positive outcome for video interviewing. Furthermore, since these data are based on complete interviews, the issues were likely resolved by the respondent or interviewer to the extent that the interview

Table 3 Percentage of technical issues experienced (reported by respondents)[5]

Technical issue	Estonia	Finland	Iceland	Italy	Nether-lands	Norway	All 6 countries
Issues with starting video call	14.6	7.5	6.6	2.2	9.3	5.3	6.7
Issues with internet connection affect-ing video call	12.1	7.9	5.4	2.8	8.1	4.7	6.1
Issues with seeing/reading showcards on screen	2.9	0.8	0.6	2.4	1.2	1.8	1.7
Audio not being clear	9.2	13.3	9.3	1.5	4.8	9.6	7.5
Video display not being clear	1.3	2.1	0.9	1.5	2.4	0.8	1.4
Other issues	1.3	1.7	5.1	1.5	5.6	5.3	3.5
No technical issues	65.0	70.0	74.5	89.7	73.4	75.6	76.4

Table 4 Percentage of technical issues experienced (reported by interviewers)

Technical issue	Estonia	Finland	Iceland	Italy	Nether-lands	Norway	All 6 countries
Issues with starting video call	10.8	10.8	6.9	2.2	8.1	7.9	7.2
Issues with internet connection affect-ing video call	10.4	6.7	4.2	2.8	6.9	4.5	5.3
Issues with display-ing showcards on screen	2.9	1.3	0.9	2.6	2.0	1.8	1.9
Audio not being clear	5.8	7.9	6.6	1.5	4.0	9.0	5.8
Video display not being clear	1.3	2.1	2.1	2.0	2.8	1.6	1.9
Other issues	5.0	5.4	4.8	2.0	5.6	5.7	4.6
No technical issues	70.8	72.9	77.8	89.1	75.4	74.7	77.9

5 The denominator for the percentages reported in Tables 3 and 4 is all respondents to be interviewed by video in each country. The sum of column percentages may exceed 100 as some respondents/interviewers reported multiple technical issues.

could be completed. But there were other cases where video interviews could not be started or finished due to technical challenges. We do not have any data on the scale of these problems, but have received anecdotal feedback from national teams that this did occur on occasion.

Comparison of Video Interviews and In-Person Interviews Across Different Quality Metrics

Satisficing

Table 5 shows the proportion of missing responses (don't know, refusal, other missing answer) per respondent between modes for each country based on the human values items at the end of the ESS questionnaire. The rate of missing values is low for both in-person and video modes. Where there are differences, in most cases the rate of missing values was higher for in-person interviews compared with video interviews. When results are combined across all six countries, there was a higher rate of missing responses for in-person compared with video interviews.

Table 5 Average number of missing responses per respondent across countries and interview modes and chi-squared test results

Country	In-person	Video	*p*-value
Estonia	0.00	0.08	<.001***
Finland	0.12	0.01	.036*
Iceland	0.36	0.15	.016*
Italy	0.31	0.08	<.001***
Netherlands	0.07	0.07	.806
Norway	0.08	0.04	.015*
All 6 countries	0.15	0.07	.029*

*$p < .05$, **$p < .01$, ***$p < .001$.

Table 6 shows the proportion of respondents identified as straightlining in the human values scale, broken down by country and mode.

Across countries, a significant difference in straightlining between modes is only present in Estonia, with less straightlining in video interviewers. When results are combined across all six countries, there is significantly less straightlining in video than in-person interviews.

Table 6 Percentage of respondents straightlining by countries and modes (Schwartz basic human value question) and chi-squared test results

Country	In-person	Video	*p*-value
Estonia	4.0	0.4	.005**
Finland	1.4	0.0	.063
Iceland	1.4	0.6	.263
Italy	10.5	11.6	.486
Netherlands	3.1	1.6	.196
Norway	1.0	0.6	.474
All 6 countries	4.7	3.1	.002**

*$p < .05$, **$p < .01$, ***$p < .001$.

Interviewer Effects

Across all six countries, it was possible to conduct a total of 509 comparisons of interviewers' intra-class correlations (ICCs) between video and in-person interviews corresponding to a total of 175 items from the ESS questionnaire (see Appendix D, Table A4 for a list of the variable names for these items). From the 175 items included in the analysis, we were able to estimate the ICC for 101 items in Estonia, 48 items in Finland, 86 items in Iceland, 112 items in Italy, 77 items in the Netherlands, and 85 items in Norway. The remaining items had nonconvergent models due to singular fit.

For each country, Figure 1 presents the distribution of the differences in size of the intra-class correlation estimates per item between video and in-person interviews. The x-axis shows the differences between the ICCs in video interviews and the ICCs for in-person interviews, while the y-axis represents the number of items grouped by 5-percentage intervals based on the ICC difference. Each figure has been centered at zero, which represents no difference between the ICCs from video and in-person interview. Following previous empirical studies on the expected size of ICCs (Beullens & Loosveldt, 2016), the differences in ICCs are grouped into color-coded ranges to indicate the magnitude of the differences and support the interpretation of their substantive relevance. Items with differences in their ICCs of plus/minus .5 points (a range of ±5% explained variance from zero) are colored in grey and are considered to be practically equal; while differences larger than .5 are in different colors and are considered substantial (Beullens & Loosveldt, 2016). Differences ranging from ±5% to ±+10% (orange bars) are considered substantial but small differences, while difference larger than ±10% or ±20% explained variance (red and purple bars) can be considered as showing large and very large differences respectively. Bars with positive values on the x-axis (right side) indicate larger interviewer effects in video interviews compared to in-person interviews for those items.

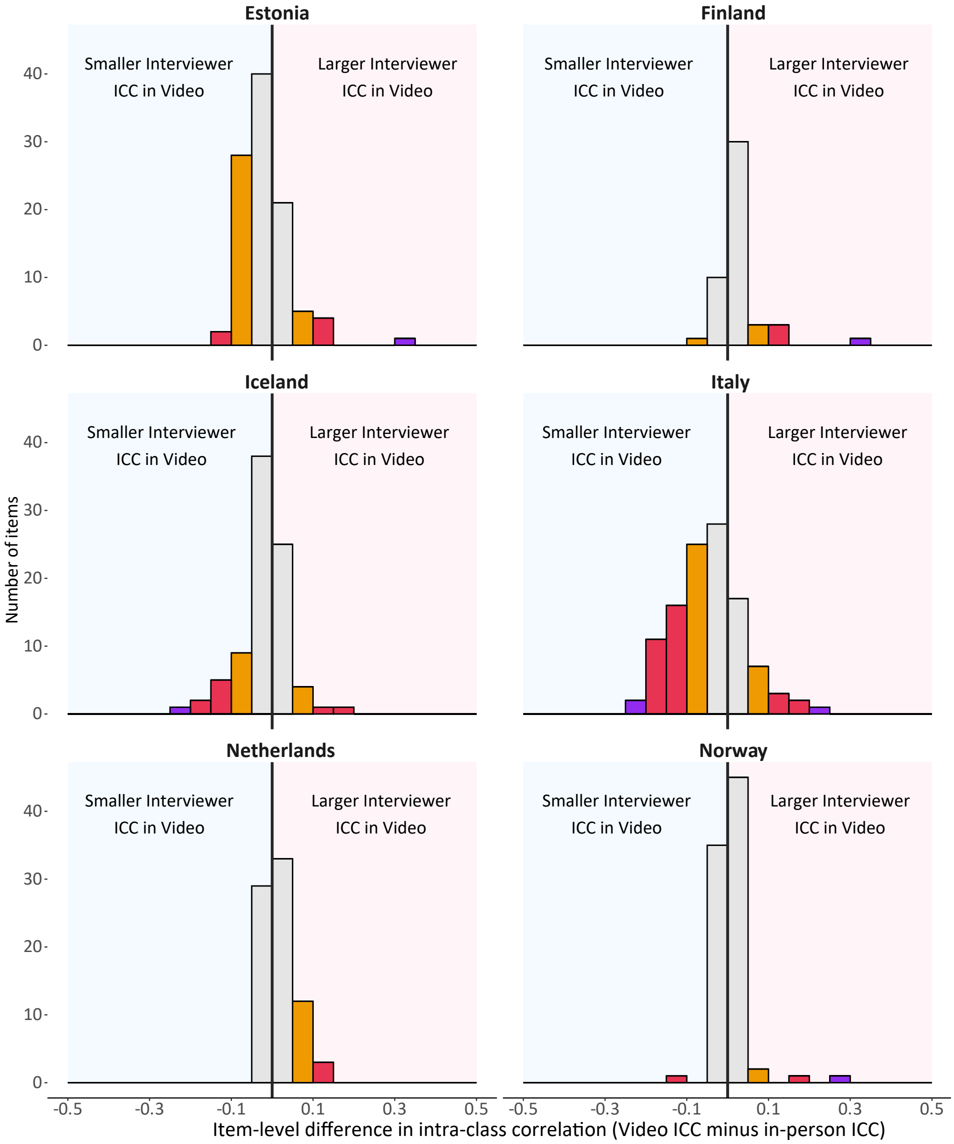


Figure 1 Difference in interviewer effects between video and in-person interviews

Firstly, we focus on the extent that ICCs differ between communications platforms, and secondly, we interpret the results regarding the direction of the observed differences. On average across all countries, 72% of the items have ICCs for video and in-person interviewers that are less than. 5 points from each other (less than 5% difference in variance explained). This indicates a general tendency of interviewer effects across these communication platforms to be more similar than dissimilar to each other. There are also differences in the interviewer effects across countries. In Norway, Finland, and the Netherlands, only a small number of items had difference in the ICCs larger than ±.5 points of explained variance (respectively 6%, 17%, and 19% of the total items). The lowest was Norway with only 5 items outside the .05 range around zero. For these countries, the observed differences between video and in-person interview can be considered negligible. In Iceland and Estonia, we observe a higher number of items with differences above the 5% range around zero. However, most of the differences observed are within the 5% to 10% range around zero difference in ICCs. The largest differences are observed in Italy, where more than half of the items have an ICC difference larger than ±.5 points. In Italy, a total of 67 items have an ICC difference larger than ±.05 (59.8% of all items in the analysis), from which 35 items (about one third of all items) have a difference larger than ±.1 points. It should be noted this analysis is not focused on the absolute sizes of the ICCs. It is not a measure of how large the interviewer effects are in video or in-person interview for these countries, but to what extent they differ by the interview mode.

Focusing on the direction of the differences in the ICCs, we observed that in countries with larger differences (like Italy, Iceland and Estonia) there are more items with larger interviewer effects in in-person interviews than in video interviews. In relative terms, the interviewer effects in Italy, Estonia and Iceland tend to be smaller in video interviews, while in Finland, the Netherlands, and Norway, the distribution skews towards higher ICCs in video interviews. If we focus only on the right side of the figure (larger ICCs in video interviews), we observe that the distributions are very similar across all countries. All countries have a small number of items with larger ICCs in video interviews (i.e., 5% or 10% above their in-person counterpart). Iceland and Norway show the smallest number of items with larger ICCs in video interviews.

In sum, interviewer effects are generally similar across video and in-person interviews, indicating that interviewers influence survey responses in comparable ways across both modes. In the countries with the largest observed differences in interviewer effects (Italy, Iceland, and Estonia) the tendency was for interviewer effects to be smaller in video interviews compared to in-person interviews, suggesting potential advantages of video platforms for data quality in certain contexts.

Discussion

This paper has taken a first look at the use of video interviewing as a complementary mode to in-person interviewing during Round 10 of the European

Social Survey. This has enabled, for the first time, a comparison of the experiences of video and in-person interviewing between countries in a cross-national survey. We sought to evaluate both experience of the implementation of video interviewing, and assess its comparability to in-person interviewing, including the quality of the data collected via video.

Regarding implementation, experiences varied widely between countries. Of the 16 countries to offer video interviewing, the proportion of respondents participating by video only exceeded 6% of the total achieved sample in six countries: Iceland, Norway, Italy, Netherlands, Estonia and Finland. Nevertheless, in these six countries a substantial proportion of video interviews were conducted, presumably lowering costs due to fewer in-person visits and reduced travel expense, as well as the possible inclusion of some respondents who may not have taken part in an in-person interview.

The reasons for differences in prevalence between countries are not fully known. However, our results suggest that high internet penetration in a country, access to samples of named individuals as opposed to households, and the ability to make prior telephone contact with target respondents, may contribute to higher yields of video interviews. Countries with some or all of those characteristics may therefore be especially well suited to video interviewing.

In those six countries where a substantial share of video interviews was achieved, we also observed positive interview experience ratings from respondents and interviewers, in most cases at least as good as those reported for in-person interviews. Based on cases where interviews were completed, relatively few technical problems were experienced, though there were some differences between countries. These experience results broadly suggest that in countries where there is demand for video interviews, video interviews can be implemented effectively and result in an experience comparable to that of in-person interviewing.

Regarding response quality, the findings for video interviewing provide some reassurance about its comparability with in-person interviewing. Only small differences between modes were observed for missingness and straightlining. Where there were differences, in most cases missingness and straightlining were lower for video compared with in-person interviews. The interviewer effects in video and in-person interviews tended to be more similar than dissimilar to each other; for most single items, the intra-interviewer correlations in video interviews differed by no more than ±5 percentage points from those observed in in-person interviews.

Overall, considering all measures, our results suggest that the interviewing process is equally effective for video interviewing compared with in-person interviewing. However, it does not necessarily follow that video interviewing should be introduced as a complementary option to in-person interviewing in all surveys. There are costs and practical concerns with the set up and implementation of video interviewing that may not always be justified if the number of interviews conducted by video may be very small. This is a decision that will need to be taken based on national and survey contexts.

The primary objective with ESS's video approach was to support the implementation of ESS Round 10 in challenging circumstances due to the pandemic. This meant that the video approach was not designed experimentally to compare with in-person interviewing in the way that would usually be done for a relatively untested method. Future research should therefore seek to randomly allocate cases to video or in-person modes. The same interviewers could also be used for both modes, with cases randomly assigned to either in-person or video mode, to reveal more about any differences in interviewer effects between modes. Where possible, sample sizes should be large enough to detect small differences between modes.

As we reported, different countries used different approaches for selecting interviewers to collect data via video. Some established a small team of specialist video interviewers, some equipped all their in-person interviewers to carry out video interviewers, and others used a subset of their in-person interviewers. Since the approach varied between countries, and was not tested experimentally, we cannot say which approach might work best in terms of factors such as uptake, practicality, and quality, or how this may vary between different contexts. This would be a fruitful area for future studies to investigate.

The nature of ESS Round 10 data collection during the pandemic may have also presented a non-typical picture. For example, demand for video interviews may have been higher at that point than might be the case in 'normal' circumstances. It will be important to continue to assess the effectiveness and comparability of video interviewing beyond the pandemic. ESS will be able to provide more insight in this regard, since several countries implemented video interviewing, alongside in-person interviewing, for the 11th Round of the survey in 2023–24.

References

Ainsaar, M., Lilleoja, L., Lumiste, K., & Roots, A. (2013). *ESS mixed mode experiment results in Estonia (CAWI and CAPI mode sequential design)*. Tartu University. https://europeansocialsurvey.org/sites/default/files/2024-04/ess_dace_mixed_mode_ee_report.pdf

Anderson, A. H. (2008). Video-mediated interactions and surveys. In Conrad, F. G., & Schober, M. F. (Eds.), *Envisioning the survey interview of the future* (pp. 95–118). John Wiley & Sons. https://doi.org/10.1002/9780470183373

Beullens, K., & Loosveldt, G. (2016). Interviewer effects in the European Social Survey. *Survey Research Methods*, *10*(2), 103–118. https://doi.org/10.18148/srm/2016.v10i2.6261

Carr, C., McDaniel, K., Guyer, H., Reed, W., Smith, T., Geiger, P., Carpenter, L., & Suresh, R. (2023, July 17–21). *Video interviewing in household and nonhousehold settings for the mental and substance use disorders prevalence study (MDPS)* [Conference presentation]. 10th Conference of the European Survey Research Association, Milan, Italy.

Centeno, L., Kelley, J., Arrue, J., Edwards, B., Hubbard, R., & Dulaney, R. (2023, July 17–21). *Video interviewing in full production* [Conference presentation]. 10th Conference of the European Survey Research Association, Milan, Italy.

Conrad, F. G., Schober, M. F., Hupp, A. L., West, B. T., Larsen, K. M., Ong, A. R., & Wang, T. (2023). Video in survey interviews: Effects on data quality and respondent experience. *Methods, Data, Analyses, 17*(2), 135–170. https://doi.org/10.12758/mda.2022.13

Dulaney, R., Kelley, J., Arrue, J., & Edwards, B. (2023, July 17–21). *Who is choosing video interviewing? A look into respondent characteristics of video adopters* [Conference presentation]. 10th Conference of the European Survey Research Association, Milan, Italy.

Endres, K., Hillygus, D. S., DeBell, M., & Iyengar, S. (2023). A randomized experiment evaluating survey mode effects for video interviewing. *Political Science Research and Methods, 11*(1), 144–159. https://doi.org/10.1017/psrm.2022.30

European Social Survey European Research Infrastructure (ESS ERIC). (2023). *ESS10 – Integrated file, edition 3.2* [Data set]. Sikt – Norwegian Agency for Shared Services in Education and Research. https://doi.org/10.21338/ess10e03_2

Hadler, M. (2004). Die Mehrebenen-Analyse. Ihre praktische Anwendung und theoretische Annahmen [Multilevel analysis: Practical application and theoretical assumptions]. *Österreichische Zeitschrift Für Soziologie, 29*(1), 53–74. https://doi.org/10.1007/s11614-004-0003-9

Hanson, T., Fitzgerald, R., Ghirelli, N., & O'Muircheartaigh, S. (2022, May 11–13). *Delivering the European Social Survey during COVID-19: Reflections and future implications* [Conference presentation]. 77th Annual conference of the American Association for Public Opinion Research, Chicago, IL, United States.

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). Routledge. https://doi.org/10.4324/9780203852279

Irani, E. (2019). The use of videoconferencing for qualitative interviewing: Opportunities, challenges, and considerations. *Clinical Nursing Research, 28*(1), 3–8. https://doi.org/10.1177/1054773818803170

Jowell, R., Roberts, C., Fitzgerald, R., & Eva, G. (Eds.). (2007). *Measuring attitudes cross-nationally: Lessons from the European Social Survey*. Sage Publications. https://doi.org/10.4135/9781849209458

Kelley, J., Edwards, B., Dulaney, R., & Arrue, J. (2023, July 17–21). *Evaluating potential mode effects in video interviewing* [Conference presentation]. 10th Conference of the European Survey Research Association, Milan, Italy.

Sanchez, C., Brown, M., Taylor, K., Spencer, S., Englefield, L., & Brayden, P. (2023, July 17–21). *Experience of video interviewing in two UK national cohort studies* [Conference presentation]. 10th Conference of the European Survey Research Association, Milan, Italy.

Schnell, R. & Kreuter, F. (2005). Separating interviewer and sampling-point effect. *Journal of Official Statistics, 21*(3), 389–410. https://www.proquest.com/scholarly-journals/separating-interviewer-sampling-point-effects/docview/1266792195/se-2

Sun, H., & Conrad, F. G. (2026). How does video interviewing affect the interviewers' and respondents' paralinguistic behaviors? A first exploration. *Methods, Data, Analyses, 20*(1), 302–316. https://doi.org/10.25521/mda.581.

Sun, H., Conrad, F. G., & Kreuter, F. (2021). The relationship between interviewer-respondent rapport and data quality. *Journal of Survey Statistics and Methodology, 9*(3), 429–448. https://doi.org/10.1093/jssam/smz043

Þórólfsson, Æ., Jónsdóttir, G. A., Hjaltason, Á. B., & Guðmundsson, H. (2023, July 17–21). *Mode effects of video interviewing as a proxy of face-to-face interviewing in R10 of ESS in Iceland* [Conference presentation]. 10th Conference of the European Survey Research Association, Milan, Italy.

Vandenplas, C., Beullens, K., & Loosveldt, G. (2019). Linking interview speed and interviewer effects on target variables in face-to-face surveys. *Survey Research Methods, 13*(3), 249–265. https://doi.org/10.18148/srm/2019.v13i3.7321

West, B. T., & Blom, A. G. (2017). Explaining interviewer effects: A research synthesis. *Journal of Survey Statistics and Methodology, 5*(2), 175–211. https://doi.org/10.1093/jssam/smw024

West, B. T., Ong, A. R., Conrad, F. G., Schober, M. F., Larsen, K. M., & Hupp, A. L. (2022). Interviewer effects in live video and prerecorded video interviewing. *Journal of Survey Statistics and Methodology, 10*(2), 317–336. https://doi.org/10.1093/jssam/smab040

Zavala-Rojas, D., Keles, O. K., Schwarz, H., & Romanin, E. (2023, July 17–21). *Measurement invariance of two concepts across face-to-face and video interview mode in ESS R10* [Conference presentation]. 10th Conference of the European Survey Research Association. Milan, Italy.

Appendix A

Key Features of ESS's Video Interviewing Approach

Table A1 Summary of key features of ESS's video interviewing approach

Design element	ESS's approach
Contact with target respondents	The process for contacting target respondents was unchanged from the usual fully in-person interviewing approach. In the majority of countries, interviewers would attempt in-person contact, sometimes following delivery of an advance letter. A video interview could then be offered based on the requirements at that stage of fieldwork (see below). A small number of countries had access to named person samples that included telephone numbers for a large majority of the sample members. In these cases, they could attempt first contact by telephone and offer a video interview at that point (meaning in these cases a video interview could be achieved without any in-person contact).
Structure of video interviewing team	National teams were allowed whether to either a) establish a small specialist team of video interviewers, or b) allow all of their in-person interviewers to also carry out video interviews. There were pros and cons with each approach. Option a) allowed for more central control and management, but led to a more complex flow of steps between the in-person interviewer and the video interviewer. There may also be concerns with interviewer effects if a small number of interviewers were carrying out a large number of interviews. Option b) was a more seamless process as the same interviewer was responsible for both the face-to-face contact and the video interview. However, there were challenges with training, equipping and monitoring a large interviewer fieldforce to take on the video interviewing task. In the event, most countries adopted option a): establishing a small specialist team of video interviewers.
When to offer video interviews	The approach changed over the round. Initially, interviewers were instructed to only offer a video interview in cases where a face-to-face interview was refused. However, feedback was received that it was very hard to convert people to take part in any way after this refusal. Subsequently, interviewers could offer a video interview if they noticed any reluctance to take part in-person. Finally, towards the end of the round, interviewers could offer a video interview from first contact, as an equal option to the in-person interview. This last approach reflected both a need to assist national teams to complete fieldwork in an extremely challenging round, and positive feedback being received based on experiences of video interviews to that point.

Table A1 (continued)

Design element	ESS's approach
Video interviewing platforms	National teams were advised to use Microsoft Teams or Zoom. They were required to use a licensed version of the platforms. Respondents would not be required to have a Teams/Zoom account in order to take part, or to download any software. In a small number of cases, countries were permitted to use alternative platforms, subject to these platforms meeting GDPR requirements and tested with respondents in advance of fieldwork.
Devices – interviewer and respondent	Interviewers should have access to two devices or screen. One would include the CAPI questionnaire, which was not shared with the respondent. This would allow the interviewer to read the questions and enter responses. The other screen would feature the video call, including the interviewer and respondent videos. This screen also included showcards, which would be screen-shared with the respondent. The showcards include response lists for many survey questions and are presented to respondents throughout ESS interviews. The interviewer moved on to the correct card as required throughout the interview. Respondents could take part using any internet-enabled device, including smartphones. It was recommended that they used a larger screen device, if available, to allow for clearer display of the video questionnaire and showcards.
Pre-testing	All countries were required to carry out at least 10 video interviews as part of a pre-test before their main stage fieldwork.
Interviewer briefings	All interviewers engaged in the video interviewing task needed to be briefed by national teams prior to fieldwork. Briefing materials were prepared by the central ESS coordinating team. National teams would need to adapt these materials in some cases to reflect their national approach.

Appendix B

Sample Composition Comparison of In-Person and Video Interviews

Table A2 Sample composition comparison between in-person and video interviews across countries

	Mode of data collection	Mean age	% of men	% legal partner-ship	% uni-versity education	% at work	% Not born in country	Mean house-hold size
Estonia	In-person	53.9**	44.9	40.5	29.4**	59.2**	14.9**	2.4**
	Video	39.5**	45.4	41.3	60.8**	82.1**	7.1**	3**
Finland	In-person	54.7**	50.2	48.1	32.9**	49.1**	3.7	2.2**
	Video	40.9**	45.4	41.7	57.4**	72.9**	2.9	2.6**
Iceland	In-person	54.8**	48.7	49.4	32.5**	57.0**	8.5	2.6**
	Video	42.2**	47.4	49.2	45.0**	82.3**	7.2	3.3**
Italy	In-person	52.4**	47.1	48.2	15.1**	46.7**	7.5	2.5**
	Video	47.9**	49.5	53.6	24.7**	58.6**	6.8	2.9**
Nether-lands	In-person	50.2**	51.8	52.7	37.7**	62.8**	8.4	2.6**
	Video	40.9**	46.8	49.6	54.8**	80.2**	6.9	3.1**
Norway	In-person	50.3**	52.5	48.3	40.8**	62.1**	10.7	2.6**
	Video	41.8**	48.3	42.2	50.3**	77.4**	10.2	2.9**

Notes: Comparisons based on 2-sample t-tests.
$^{**}p < .01$.

Appendix C

Comparison of Interview Duration Between In-Person and Video Interviews

Table A3 Length of the main interview (in minutes) by different modes

Country	In-person	Video	*p*-value
Estonia	57.3	66.1	.001***
Finland	60.6	60.9	.421
Iceland	60.7	60.7	.486
Italy	47.6	58.1	.001***
Netherlands	65.6	67.1	.264
Norway	63.6	63.6	.497

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

To compensate for the differences between interviewees, we use only data of persons with a relatively similar background to compare the length of the interviews: a person living with a partner who has at least a secondary or higher education and who younger than 60 years old. This allows for in-person sample sizes between 302 and 1,283 per country, and video interviewing sample sizes between 214 and 421 per country.

Appendix D

Variable Names of Items Included in Intra-Interviewer Correlation

Table A4 Variables used in intra-interviewer correlation

Items compared for ICCs in at least one country	ACTROLGA, AESFDRK, AGEPNT, ANCTRY2, ATCHCTR, ATCHERP, C19MCO12, C19MCPNT, C19SPWRK, C19WHACC, C19WHOME, CHLDO12, CHPLDMC, CHPLDMI, CLOSEPNT, COLCOM, COLHLP, COLPHONE, COLPROP, COLSCRN, COLSPEAK, COMO12, COMPNT, CPTPPOLA, CTTRESA, CTTRESAC, DCSFWRKA, DFPRTAL, DFPRTALC, EDULVLB, EDULVLFB, EDULVLMB, EDULVLPB, EDUYRS, ESTSZ, EUFTF, FAIRELC, FAIRELCC, FAMADVS, FAMPDF, FAMPREF, FREEHMS, GINCDIF, GOVMONPB, GOVPRIPH, GPTPELC, GPTPELCC, GRDFINC, GRDFINCC, GVBALC19, GVCONC19, GVCTZPV, GVCTZPVC, GVELDC19, GVFAMC19, GVHANC19, GVIMPC19, GVJOBC19, HAPPY, HEALTH, HHLIO12, HHMMB, HINCFEL, HINCTNTA, HMSACLD, HMSFMLSH, HSCOPC19, IMBGECO, IMDFETN, IMPCNTR, IMPDIFF, IMPENV, IMPFREE, IMPFUN, IMPRICH, IMPSAFE, IMPTRAD, IMSMETN, IMUECLT, IMWBCNT, INPRDSC, IORGACT, IPADVNT, IPBHPRP, IPCRTIV, IPEQOPT, IPFRULE, IPGDTIM, IPHLPPL, IPLYLFR, IPMODST, IPRSPOT, IPSHABT, IPSTRGV, IPSUCES, IPUDRST, ISCO08, JBPRTFP, KEYDEC, KEYDECC, LRSCALE, MANCOM, MANHLP, MANSCRN, MANSPEAK, MANSUPP, MANWRKPL, MCCLOSE, MCCOORD, MCINTER, MCMSINF, MCPRIV, MCWRKHOM, MEDCRGV, MEDCRGVC, NETUSOFT, NETUSTM, NJBSPV, NWSPOL, PANCLOBO, PANFOLRU, PANMONPB, PANPRIPH, PANRESMO, PFMFDJBA, POLINTR, PPLFAIR, PPLHLP, PPLTRST, PRAY, PRTDGCL, PSPPIPLA, PSPPSGVA, RGHMGPR, RGHMGPRC, RLGATND, RLGDGR, SCIDECPB, SCLACT, SCLMEET, SCRNO12, SECGRDEC, SPEAKO12, SPEAKPNT, STFDEM, STFECO, STFEDU, STFGOV, STFHLTH, STFLIFE, STFMJOB, TEAMFEEL, TRDAWRK, TRSTEP, TRSTLGL, TRSTPLC, TRSTPLT, TRSTPRL, TRSTPRT, TRSTUN, TTMINO12, VIEPOL, VIEPOLC, VOTEDIR, VOTEDIRC, WKDCORGA, WKHCT, WKHTOT, WKHTOTP, WPESTOP, WPESTOPC, WRKEXTRA, WRKHOME, WRKLONG, YRBRN

methods, data, analyses | Vol. 20(1), 2026, pp. 209-236 DOI: 10.25521/mda.662

Measurement Invariance of Survey Data across Face-to-face and Video Interview Modes

Diana Zavala-Rojas[1,2], Oguz Korkut Keles[2] & Hannah Schwarz[2]

[1] *European Social Survey ERIC*

[2] *Universitat Pompeu Fabra*

Abstract

The COVID-19 pandemic posed challenges to the traditional mode of administration of surveys. In most European countries it was not feasible to conduct standardized in-person interviews due to restrictions. These events accelerated the trend towards alternative modes of survey data collection, such as live video interviewing. In this paper, we investigate whether survey data collected with two different interviewing modes, face-to-face and video interview, are comparable. Measurement invariance is a prerequisite to make meaningful comparisons across groups. Using data from the European Social Survey Round 10, we assess if the measurement invariance condition is met for two concepts, social trust and attitudes towards immigration, in six European countries: Estonia, Finland, Iceland, Italy, the Netherlands and Norway. The results are encouraging: Full metric invariance is established for all countries for social trust and for three out of six countries (Iceland, Italy and Norway) for attitudes towards immigration. For those cases, analysts can aggregate data across modes to investigate relationships among variables, e.g., using regression analysis. Scalar invariance holds in fewer cases. For the concept social trust, full scalar invariance is established for Finland, Italy, the Netherlands and Norway while for the concept attitudes towards immigration, it is established for Iceland and Italy. For those cases, data can be aggregated across modes to compare observed means at the country-language level. Generally, where measurement equivalence is not reached, we recommend excluding observations from the video interview mode before analysis.

Keywords: Measurement invariance, video mode, face-to-face mode, European Social Survey

© The Author(s) 2026. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Conducting video interviews for various purposes is becoming increasingly popular as the required technology becomes ever more accessible (Conrad et al., 2023). Video interviewing offers potential advantages compared to in-person interviewing, such as cost reduction and time efficiency. A trend towards video interviewing was already discernible before the COVID-19 pandemic (Joshi et al., 2020). With the COVID-19 pandemic posing numerous challenges to survey research relying on in-person (i.e., face-to-face) interviews, its relevance peaked as conducting in-person interviews became infeasible for entities conducting survey interviews in many countries, especially for surveys also covering respondents in health risk groups such as elderly persons. Given the apparent success of video interviews, survey researchers have continued studying this method beyond the pandemic period. The mode of interview administration typically has effects on the responses that people provide to survey questions. Recent literature investigates the possible implications the video interview mode might have on different aspects of surveys, for example in terms of interviewer effects (West et al., 2022), implementation challenges (Durrant et al., 2024), participation (Schober et al., 2023) and data quality (Conrad et al., 2023). Furthermore, survey mode effects across in-person and video interview mode have been studied (Endres et al., 2023).

These studies suggest that differences between modes may influence how respondents process questions and provide answers—through mechanisms such as interviewer behavior or perceived confidentiality. Our work contributes to this body of research by investigating the comparability of survey data across in-person and video interview mode by formally testing for measurement equivalence (Meredith, 1993). We assess measurement invariance to examine whether constructs are measured equivalently across modes, that is, whether differences in responses reflect true differences in the underlying concepts. We use measurement equivalence testing because respondents have not been randomly assigned to interview modes in the survey data we use. In this context, measurement invariance testing[1] provides a statistical means to evaluate comparability across groups while acknowledging that some residual bias from sampling differences may remain. It enables researchers to assess if the compara-

[1] *Measurement equivalence or measurement invariance* is assessed by testing whether certain parameters of a measurement model are identical across groups (Davidov et al., 2015).

Acknowledgement
We thank Carlos Poses for his comments on an early version of this study and Laura Font Rigueiro who collaborated as a research assistant, as well as the anonymous reviewers for their helpful comments.

Direct correspondence to
Hannah Schwarz, Edificio Mercè Rodoreda, Despatx 24.407, Ramón Trías Fargas 25-27, 08005 Barcelona, Spain
E-mail: hannah.schwarz@upf.edu

bility of data is not hindered by group membership (here: mode groups). Only if measurement invariance is established between in-person and video interview data, can measurements across groups be analyzed jointly in conventional statistical analysis.

There are different levels of tests for measurement invariance with increasing restrictions in the parameters of the models (Meredith, 1993; Vandenberg & Lance, 2000). Configural invariance means that the same survey items (i.e., indicators) load on the same latent construct (i.e., factor) across groups. Sometimes, this is also referred to as the measurement model. It is interpreted as a construct having similar theoretical content across groups. Because the factor model relies on only three observed variables (statistically this is known as a just identified model), we do not test for configural invariance. This situation is common with very short scales, particularly in survey research where three-indicator single-factor models are not uncommon and where the identification constraints needed to set the metric of the latent variable exhaust all available degrees of freedom. Prior methodological work (e.g., Cheung & Rensvold, 2002; Wu & Estabrook, 2016) therefore treats configural invariance in such cases as satisfied, noting that no empirical misfit can be detected at this step. In other words, the theoretical models are not tested but assumed[2].

Loading invariance, also known as metric invariance, means that the items' factor loadings are equal across groups. Where this is the case, statistical relationships of each item of a construct can be meaningfully compared across groups. In other words, the relative weight of each item is the same across groups.

Intercept invariance, also known as scalar invariance, means that not only the items' factor loadings but also the intercepts are equal across groups. When metric and scalar invariance are established across groups, observed means can be compared. In the remainder of this paper, we use the terms metric and scalar invariance.

The above also represents an ascending rank ordering of desirability of types of invariance for analysts. In order to conduct meaningful analyses across groups, here mode groups, establishing scalar invariance results in most ease for analysts as it allows them to meaningfully compare observed means across the mode groups. Where only metric but not scalar invariance is established, only statistical relationships can be meaningfully compared across mode groups. If analysts still wish to compare means having only established metric invariance, they have to resort to the strategy of comparing latent means to do so.

---

[2] Our approach in this article of assuming the theoretical model in a case of only three items is in line with other published work such as Nickel and Weber (2024); Pirralha and Weber (2020); Van de Vijver (2019); Nießen et al. (2020). We include in the Appendix the commonly reported global fit statistics for our scalar models purely for the sake of transparency, not because we actually rely on them for model evaluation.

One, furthermore, distinguishes between partial and full invariance where partial invariance means that a given type of invariance can be established for only some items constituting a construct, not all (Byrne et al., 1989). For the case of finding partial invariance, a strategy analysts can resort to is, once again, comparing latent means instead of observed means (Pokropek et al., 2019).

If the interactions between respondents and interviewers were the same in both modes (West et al., 2022), we would expect scalar invariance. However, we will argue that the social dynamics in the two modes differ due to factors such as perceptions of confidentiality and personal proximity. Furthermore, interviewers have less control over the context of the interview in video mode such that distractions through multitasking on the side of the respondent are more likely than in in-person mode (Deakin & Wakefield, 2014). Additionally, other people might be present during the interview unbeknownst to the interviewer who only sees the part of the room captured by the camera. More generally, a relevant argument for this study concerns potential differences in social desirability across modes and the concepts selected for the measurement invariance tests. Social desirability bias tends to be less pronounced in modes where interpersonal proximity between respondent and interviewer is lower (Heerwegh, 2009). This would be the case in video interviews as both of them are not physically present in the same place, but the interaction happens by the means of a screen.

For all these reasons, we expect that data collected across the two modes may not be fully comparable, that is, we do not expect full scalar invariance.

Using data from the European Social Survey (ESS) Round 10, we test if the measurement invariance condition is met for two concepts, social trust and attitudes towards immigration, in six different European countries: Estonia, Finland, Iceland, Italy, the Netherlands and Norway.

Despite having found differences across countries and concepts in the levels of invariance, if in-person interviews are the benchmark, our results are encouraging for video interviews: Full metric invariance is established for all countries for social trust and for three out of six countries (Iceland, Italy and Norway) for attitudes towards immigration. For those cases, analysts can aggregate data across modes to investigate relationships among variables, for example, using regression analysis. Full scalar invariance is established for fewer cases but still for four countries for social trust (Finland, Italy, the Netherlands and Norway) and for two countries for attitudes towards immigration (Iceland and Italy). For those cases, data can be aggregated across modes to compare observed means at the country level. Generally, where measurement equivalence is not reached, we recommend excluding observations from the video interview mode before analysis or using latent variable approaches, to account for measurement error (Saris & Gallhofer, 2014). In this article, we argue that differences in invariance levels across countries and concepts may be attributable to differences in social desirability across modes, concepts and cultures. However, future research will be necessary to elucidate the exact causes of non-invariance.

The remainder of the paper is structured as follows: First, we introduce the relevant literature and position our research within it. Then, we define the concepts that are examined across different modes, namely social trust and attitudes towards immigration. Next, we present the measurement instruments (i.e., survey questions) used to measure them. Then, in the methods section, we elaborate on the data we use, on measurement invariance testing more generally and on our analytical approach. Finally, we present our results followed by a discussion and conclusion.

Survey Modes and Potential Implications for Equivalence

Our research contributes to the growing literature that tests for measurement invariance to elucidate potential issues with data equivalence across groups. The necessity of testing for measurement invariance has been emphasized in several studies as it ensures meaningful, valid, and unbiased comparisons across data gathered in different groups (e.g., modes, countries, languages; Meuleman et al., 2023). This is particularly critical in cross-cultural studies, where invariance testing is essential to validate comparisons (Davidov et al., 2014; Steinmetz et al., 2009).

More specifically, our research contributes to the literature on measurement invariance across modes of data collection. Previous research on this includes, for instance, Boal et al. (2017) who assess measurement invariance across two different data collection modes, telephone and web self-administration. Their hypothesis is that there may be differences because the survey is presented primarily visually in web self-completion mode and primarily aurally in telephone mode, and additionally, because phone administration involves an interviewer, whilst web administration does not. Their results show configural, metric and partial scalar invariance for the Posttraumatic Stress Disorder (PTSD) Checklist across the two mode groups. The authors conclude that, in practical terms (i.e., for PTSD diagnosis) the use of both modes is unproblematic.

There is, furthermore, literature on mode effects, some of which we will review here because it helps us develop our hypotheses. For example, Buerger et al. (2016) assess mode effects between different self-completion modes hypothesizing that they could exist due to factors such as item layout, input devices, scrolling, and the likelihood of revisiting of past responses. Some of these factors might also differ between video and in-person interviews. For example, input devices and/or item layout (depending on the screen size) might differ between the two modes. Furthermore, the physical presence of an interviewer in in-person interviews may induce pressure on respondents to continue with the interviewing process swiftly which could lead to respondents in in-person interviews being less inclined to revisit previous questions than those in video interviews.

More generally, a relevant argument for this study concerns potential differences in social desirability across modes. Social desirability bias tends to be less pronounced in modes where the interpersonal proximity between respondent and interviewer is lower (Heerwegh, 2009). While this argument is often made concerning the distinction between interviewer-administered and self-completion modes (Heerwegh, 2009; Schnell & Kreuter, 2005), it similarly applies to the comparison of a context where the interviewer is physically present in a room with respondents as compared to a context where respondent and interviewer communicate via videochat (Moallem, 2015). If this is the case, we could expect differences across the concepts. As attitudes towards immigration shows high social desirability bias (Creighton et al., 2019), it could be the case that in video interview respondents would not feel as pressured to give a socially expected answer as they would when the interviewer is present in the same place as them.

Closely linked to this is the literature exploring the varying impact of interpersonal interactions across interview modes. We argue that the type of relationship established between the respondent and the interviewer during the interview process may differ due to factors such as their perceptions of confidentiality, willingness to share information, social presence, personal proximity, and/or the frequency of interactions or dialogues (e.g., there might be less probing in one mode; Hoehne et al., 2024; de Villiers et al., 2022). The quality of the interaction between respondent and interviewer is an important aspect of the mode effect (Conrad et al., 2023) and may also be important for the comparability of the data across cultures (Van de Vijver & Tanzer, 2004). Even if a live interviewer is present in both assessed modes, there is evidence that a successful interaction through videochat requires experience (de Villiers et al., 2022). Further research demonstrates that rapport is stronger in in-person interviews or, more generally, in-person communication than in video interviews (Faucett et al., 2017; Fernandez et al., 2021; Gordon et al., 2020; Joseph et al., 2017; Sherman et al., 2013). Yet, a weaker social presence of the interviewer in video mode may also be beneficial for topics in which the respondent perceives a high expectation of a socially desirable answer.

Another aspect to consider is that video interviews can take place in a larger variety of environments than in-person interviews which are usually conducted at a respondent's home. An important consequence is the higher potential for disruptive background factors in video interviews which might lead to distraction or multitasking behavior on the side of the respondent (Deakin & Wakefield, 2014).

Furthermore, in the literature a distinction is made between two types of video interviews: pre-recorded and live interviews. Schober et al. (2020) argue that differences between live video interviews and pre-recorded video interviews that may induce different answering behavior across the two modes due to the fact live video interviews involve interaction and pre-recorded video interviews do not.

Concepts

We test for measurement invariance of the concepts of social trust and attitudes towards immigration using the country cases of the ESS Round 10 which have data available for video as well as in-person interview mode. We specifically chose these concepts because they differ in how prone they are to social desirability bias, that is, the tendency to provide an answer that the respondent thinks aligns with a socially expected and accepted idea. The measurement of attitudes towards immigration is argued to suffer from high social desirability bias (Rinken et al., 2021; Piekut, 2021) while this is not the case for social trust. Furthermore, video and in-person interviewing should induce differing levels of social presence of the interviewer (Moallem, 2015) which, in turn, should affect how strongly answers are impacted by social desirability bias. Hence, we expect variation in patterns of answering behavior between the two modes across the two concepts.

Furthermore, we consider the selected concepts especially relevant as both are widely studied in the social sciences. There is a broad consensus among social scientists about the importance of the concept of social trust, from a political and societal perspective. For example, it has been argued that social trust is central for the functioning of democracies (Inglehart, 1999). However, even if social trust has been routinely used in cross-country research (Inglehart, 1999; Delhey & Newton, 2005; Letki & Evans, 2005; Rothstein & Uslaner, 2005; Adam, 2008; Herreros & Criado, 2008; Kaasa & Parts, 2008; Zmerli & Newton, 2008), the measurement of social trust is not necessarily comparable across different groups (Van der Veld & Saris, 2011; Reeskens & Hooghe, 2008). Studying the concept of attitudes towards immigration has also been a relevant matter of research in the social sciences for some decades (see, e.g., Ceobanu & Escandell, 2010). The determinants of these attitudes are also a matter of study, both at the individual and the country level. For a review see for instance Davidov and Semyonov (2017). Furthermore, this concept is frequently studied using cross-nationally comparative datasets (Ceobanu & Escandell, 2010), where, among many other factors, the modes of data collection vary (Smith et al., 2011).

For both concepts, cross-cultural measurement equivalence has been studied with ESS data (see, e.g., Davidov et al., 2015; Nickel & Weber, 2024; Reeskens & Hooghe, 2008). Furthermore, Revilla (2013), assesses measurement equivalence of the two concepts across in-person and web self-completion mode in survey data from the Netherlands and finds scalar invariance for both concepts. Moreover, punctual experiments have been conducted in different ESS Rounds which assessed mode effects across a variety of modes. Results from these show, for example, mode differences in attitudinal questions in general, and more specifically for the concept of attitudes towards immigration in in-person interviews versus various alternative modes (Villar & Fitzgerald, 2017).

In this study, each of the concepts is measured with three indicators. The wording of the three corresponding ESS items is shown in Table 1 and Table 2, respectively. All six survey questions are presented with 11-point item-specific scales (0-10). Figure 1 illustrates the configural model, that is, the relationship between the latent variables and the indicators for the concept of social trust, as an example. As shown by the three unidirectional arrows, the responses to the observed variables (y_{ij}) are determined by the latent construct (ξ_j). The corresponding equations to this graph are shown in Section 4.2.

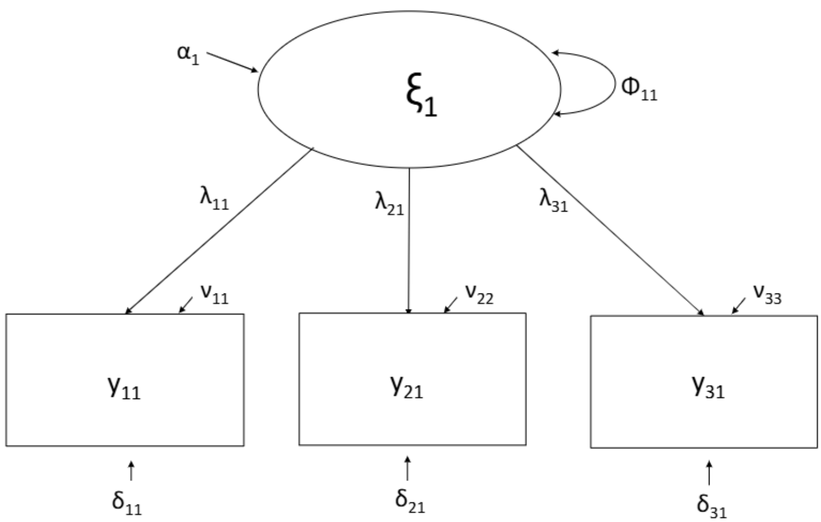


Figure 1 Path diagram of the configural model

Table 1 ESS survey instrument for measuring concept social trust

Question wording	Variable name	Response scale
Using this card, generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people? Please tell me on a score of 0 to 10, where 0 means you can't be too careful and 10 means that most people can be trusted.	ppltrst	0 (You can't be too careful) to 10 (Most people can be trusted)
Using this card, do you think that most people would try to take advantage of you if they got the chance, or would they try to be fair?	pplfair	0 (Most people try to take advantage of me) to 10 (Most people try to be fair)
Would you say that most of the time people try to be helpful or that they are mostly looking out for themselves?	pplhlp	0 (People mostly look out for themselves) to 10 (People mostly try to be helpful)

Table 2 ESS survey instrument for measuring concept attitudes towards immigration

Question wording	Variable name	Response scale
Would you say it is generally bad or good for [country]'s economy that people come to live here from other countries?	imbgeco	0 (Bad for the economy) to 10 (Good for the economy)
And, using this card, would you say that [country]'s cultural life is generally undermined or enriched by people coming to live here from other countries?	imueclt	0 (Cultural life undermined) to 10 (Cultural life enriched)
Is [country] made a worse or a better place to live by people coming to live here from other countries?	imwbcnt	0 (Worse place to live) to 10 (Better place to live)

Methods

Data

We use data from the ESS Round 10 (ESS ERIC, 2023). Due to locally varying contact restrictions in the context of the COVID-19 pandemic, some ESS countries opted for adding a video-interviewing option to their Round 10 fieldwork protocol while others did not (Ghirelli et al., 2024). In this paper, we include all six country cases for which a video interviewing option was offered to respondents as an alternative to in-person interviewing, namely Estonia, Finland, Iceland, Italy, the Netherlands and Norway. Fieldwork in these countries was conducted between June 2021 and May 2022. The ESS reports the following response rates: 47.2% (Estonia), 41.1% (Finland), 33.6% (Iceland), 49.8% (Italy), 35.7% (the Netherlands) and 37.9% (Norway). These response rates are calculated as the number of complete and valid interviews over the number of issued eligible sample units (for more detailed information, see Ghirelli et al., 2024). Distributions of the variables of main interest are presented in Tables 3 and 4. These Tables also show that the proportions of cases missing due to item nonresponse are low with a mean of 0.7% for the concept of social trust and somewhat higher, with a mean of 3.2%, for the concept of attitudes towards immigration. This is unsurprising given that the latter concept can be regarded as more sensitive, frequently being at the center of political debate and given that previous studies have found higher item nonresponse for this concept (for a more extensive discussion, see, e.g., Piekut, 2021).

Table 3 Descriptive statistics of variables measuring social trust (ppltrst, pplfair, pplhlp) by country and mode

Country	Mode	Valid *N*	*N*	ppltrst *M (SD)*	pplfair *M (SD)*	pplhlp *M (SD)*
Estonia	in-person	1296	1302	5.70 (2.17)	5.88 (2.13)	5.16 (2.15)
	video	239	240	6.10 (2.01)	6.44 (1.84)	5.76 (2.02)
Finland	in-person	1333	1337	6.93 (1.87)	7.10 (1.84)	6.37 (1.92)
	video	239	240	7.00 (1.62)	7.18 (1.58)	6.37 (1.64)
Iceland	in-person	560	567	6.53 (2.11)	7.14 (1.69)	6.73 (1.82)
	video	331	333	6.71 (1.99)	7.20 (1.59)	6.55 (1.71)
Italy	in-person	2166	2183	4.96 (2.24)	4.97 (2.04)	4.56 (2.04)
	video	447	457	5.21 (2.12)	5.22 (1.95)	4.69 (2.02)
Netherlands	in-person	1216	1221	6.50 (1.81)	6.76 (1.57)	5.89 (1.75)
	video	245	248	6.76 (1.71)	6.93 (1.34)	6.02 (1.51)
Norway	in-person	917	920	6.90 (1.87)	7.09 (1.74)	6.37 (1.87)
	video	488	491	6.69 (1.81)	6.95 (1.67)	6.22 (1.79)

As respondents self-selected into either the in-person or the video interviewing mode, it can be expected that the sociodemographic composition of these two groups differs. We provide information on the distributions of variables age, sex and university education in Table 5. We find respondents who self-selected into the video interviewing mode to be, on average, younger and more highly educated. We find no strong or consistent differences in terms of sex.

Table 4 Descriptive statistics of variables measuring attitudes towards immigration (imbgeco, imueclt, imwbcnt) by country and mode

Country	Mode	Valid *N*	*N*	imbgeco *M (SD)*	imueclt *M (SD)*	imwbcnt *M (SD)*
Estonia	in-person	1261	1302	5.25 (2.40)	5.22 (2.42)	5.16 (2.25)
	video	238	240	6.41 (2.27)	6.48 (2.16)	5.76 (1.91)
Finland	in-person	1314	1337	6.12 (2.19)	7.29 (1.93)	6.37 (2.00)
	video	236	240	6.40 (2.07)	7.40 (1.96)	6.37 (1.98)
Iceland	in-person	544	567	7.04 (1.99)	7.39 (1.96)	6.73 (1.98)
	video	323	333	7.14 (1.95)	7.62 (1.98)	6.55 (1.84)
Italy	in-person	2089	2183	5.00 (2.39)	5.13 (2.55)	4.56 (2.24)
	video	442	457	5.60 (2.46)	5.86 (2.59)	4.69 (2.27)
Netherlands	in-person	1156	1221	5.91 (1.90)	6.56 (1.97)	5.89 (1.76)
	video	230	248	6.18 (1.59)	6.93 (1.65)	6.02 (1.68)
Norway	in-person	898	920	6.20 (1.97)	6.68 (2.21)	6.37 (1.97)
	video	483	491	6.06 (1.92)	6.73 (2.08)	6.22 (1.95)

Table 5 Descriptive statistics of sociodemographic variables by country and mode

Country	Mode	Age *M (SD)*	Proportion women	Proportion university education
Estonia	in-person	53.88 (18.63)	.55	.29
	Video	39.53 (12.61)	.55	.61
Finland	in-person	54.70 (19.40)	.50	.33
	Video	40.98 (14.05)	.55	.58
Iceland	in-person	54.83 (18.86)	.51	.33
	Video	42.15 (15.79)	.53	.45
Italy	in-person	52.37 (19.01)	.53	.15
	Video	47.87 (16.61)	.51	.24
Netherlands	in-person	50.22 (18.81)	.48	.37
	Video	40.87 (14.69)	.53	.55
Norway	in-person	50.28 (18.84)	.48	.41
	Video	41.76 (15.37)	.52	.50

Measurement Invariance

Measurement invariance testing is suited for situations, such as ours, where respondents self-select into groups that differ in observed and unobserved characteristics. It tests whether the latent constructs are recovered equivalently despite such possible compositional differences (Meredith, 1993; Vandenberg & Lance, 2000; Davidov et al., 2014). Several families of measurement invariance tests are available (e.g. Comparative Fit Index/Root Mean Square Error of Approximation rules, Bayesian Multi-Group Confirmatory Factor Analysis, alignment, Multiple Indicator Multiple Cause model), but their performance varies with sample size among groups and model identification strategies. Because our mode groups are highly unbalanced (as Tables 3 and 4 show) and each construct is measured with only three indicators (yielding just-identified configural models), fit-index–based criteria risk inflated Type II error in detecting non-invariance (Chen, 2007) and Bayesian priors cannot be incorporated. We therefore adopt the bias-corrected bootstrap factor-ratio test proposed by Cheung and Lau (2012) within multi-group confirmatory factor analysis (MG-CFA), which remains robust when group sizes differ substantially.

The model represented in Figure 1 is equivalent to Equation 1, which specifies the configural model (Bollen, 1989; Meredith, 1993).

Equation 1 specifies a model for the relationship between three observed variables $y_{ij}^{g}, i = 1, 2, 3,$ the answers to the survey items and a latent (unobserved) variable, $\xi_j^g, j = 1, 2,$ representing the concepts we measure by the survey questions, namely, 1) social trust and 2) attitudes towards immigration. As multiple groups are involved, a multi-group confirmatory factor analysis (MG-CFA) is used, in our case, this is represented by $g = 1, 2,$ the mode groups: in-person or video interview, respectively. In this model, ν_{ij}^{g}, represents the intercepts, λ_{ij}^{g}, represents the factor loadings between the ith observed variable and the jth latent variable and δ_{ij}^{g} represents the disturbance terms. It is assumed that the disturbance terms have a mean of zero, they are uncorrelated with each other and with ξ_j^g as shown by (4) to (6).

$$y_{ij}^{g} = \nu_{ij}^{g} + \lambda_{ij}^{g}\xi_j^{g} + \delta_{ij}^{g} \tag{1}$$

For the latent means, α_j, we establish that

$$\alpha_j = E(\xi_j) \tag{2}$$

Furthermore, for the variance of ξ_j, we establish that

$$\phi_{jj} = var(\xi_j) \tag{3}$$

Moreover, for all models we make the following assumptions:

$$E(\delta_{ij}) = 0; \tag{4}$$

$$E(\delta_{ij}\xi_1) = 0; \tag{5}$$

$$E(\delta_{ij}\delta_{ij'}) = 0 \; or \; i \neq i' \tag{6}$$

Measurement invariance holds where the parameters (ν_{ij}^{g}, λ_{ij}^{g}) of the factor model for ξ_j^g are equal across groups. Metric invariance is established when λ_{ij}^{g} are equal for $g = 1, 2$ and scalar invariance is established when ν_{ij}^{g} are equal for $g = 1, 2$. Furthermore, for metric and scalar invariance, partial invariance can be achieved. Partial metric invariance refers to the loadings for some but not all items being equal. Partial scalar invariance refers to loadings and intercepts only being equal for some of the items, not all (Byrne et al., 1989).

When equality constraints are imposed on loadings and intercepts, the constrained model is usually compared with the unconstrained (configural) model using a likelihood-ratio test based on the change in chi-square, $\Delta\chi^2$. Yet $\Delta\chi^2$ is highly sample-size sensitive and may inflate type-I error even for minor parameter differences (Cheung & Rensvold, 2002). Consequently, to mitigate this problem, researchers also inspect other fit indices, such as comparative fit index (ΔCFI), root mean square error of approximation (ΔRMSEA) and standardized root mean squared residual (ΔSRMR), but these statistics lose power when group sizes are unbalanced, increasing type-II error risk and potentially masking non-invariance (Chen, 2007).

As dependency of global fit measures on sample size is a long historical challenge of structural equation modelling (SEM), other approaches have been suggested, for instance, to evaluate the models combining the fit index (ΔFI) with the detection of misspecifications using local fit measures (Van der Veld & Saris, 2011). However, in both global and local fit approaches, the employed cut-off values for defining misspecifications and thresholds for concluding that one model has a worse fit than another has been criticized as arbitrary (Cheung & Rensvold, 2002; Kline, 2016; Millsap, 2011). Recently, other more flexible alternatives to MG-CFA such as Bayesian MG-CFA (Muthén & Asparouhov, 2017) have been suggested. However, they are unsuitable for just-identified models, as there are no degrees of freedom for the incorporation of priors. Additionally, the degree of the sensitivity of Bayesian MG-CFA to sample size is being debated (Gelman, 2006) hence we do not consider this a suitable approach.

Another consideration is that invariance testing, and more generally, structural equation models, require using an item as a reference to define the unit of measurement and identify the models, that is, the loading of an arbitrarily chosen item is fixed to 1, e.g. $\lambda_{21} = 1$ and its intercept to zero, e.g. $\nu_{21} = 0$. Cheung and Rensvold (1999) argue that, with this choice, the selected item is implicitly assumed to be invariant. This implies that the choice of the item is critical, and if a non-invariant item is selected, the conclusions about the invariance of other items are likely to be incorrect (Johnson et al., 2009).

All the considerations outlined above led us to work with the bias-corrected (BC) bootstrap confidence interval approach (Cheung & Lau, 2012). This approach uses a single model as opposed to the likelihood ratio test (LRT) and Δ FI approaches, which estimate and compare different models. It offers a solution to test for measurement invariance using data with very unequal sample sizes. It uses maximum likelihood estimation with bootstrapping to evaluate the models, in combination with a factor-ratio test, that allows overcoming the challenges of the reference item, especially relevant in our data as the models have only three items each. The specific implementation of the BC bootstrap confidence interval approach is described in the next section.

Analytical Strategy

To test for metric and scalar measurement invariance following the BC bootstrap confidence interval approach (Cheung & Lau, 2012), we constrained the estimates of the parameters λ_{ij}^{g} for both mode groups, $g = 1$ for in-person interviews, $g = 2$, for video interviews to equality, within each country model. The factor-ratio test uses three constraints which are created using each item as reference. Equation 7 to Equation 9 illustrate the test for the concept of social trust. When y_{11}^{g} is the referent item, the hypotheses to test for metric invariance are:

$$\frac{\lambda_{21}^{1}}{\lambda_{11}^{1}} - \frac{\lambda_{21}^{2}}{\lambda_{11}^{2}} = 0 \tag{7}$$

$$\frac{\lambda_{31}^{1}}{\lambda_{11}^{1}} - \frac{\lambda_{31}^{2}}{\lambda_{11}^{2}} = 0 \tag{8}$$

When y_{2j}^{g} is the referent item, an additional hypothesis is tested:

$$\frac{\lambda_{31}^{1}}{\lambda_{21}^{1}} - \frac{\lambda_{31}^{2}}{\lambda_{21}^{2}} = 0 \tag{9}$$

As there are only three items, there are no other hypotheses to test when y_{3j}^{g} is the referent item. If zero falls outside of a confidence interval, the null hypothesis of invariance can be rejected. Should zero fall outside of the confidence interval for only one of these constraints, we can establish partial invariance. Using the same procedure, a factor-ratio test is defined to test for scalar invariance, and confidence intervals are computed and evaluated, using the same rules: Should zero fall outside of the confidence interval for one of these constraints, we can establish partial invariance and should zero fall outside of the confidence interval for at least two of these constraints, this shows that there is non-invariance[3]. We estimate confidence intervals using the bias-corrected bootstrapping approach, for which 10,000 bootstrap samples are created, parameters are estimated for each bootstrap sample and calculations are made to the bootstrap distribution of the parameters to form the confidence intervals.

Results

Table 6 summarizes our findings. We show results by country, concept and type of invariance. We show the point estimates and the confidence intervals of the factor-ratio tests in a series of graphs. In order to facilitate the interpretation of the graphs, we simplify the notation of the equations from Section 4.3 in the Y-axis of the graphs.

Below, we present the original labels from our factor-ratio tests along with explanations.

1. $\lambda_{2.1}^{g1} - \lambda_{2.1}^{g2}$ represents the difference in the factor loading for item 2 between group 1 (in-person) and group 2 (video), when item 1 is used as the reference (i.e., its loading is fixed). Under measurement invariance, the relative contribution of item 2 compared to item 1 should be identical across groups.

Y-axis label: *Loading Diff. for Item 2 (Ref: Item 1)*

2. $\lambda_{3.1}^{g1} - \lambda_{3.1}^{g2}$ this is the analogous difference for item 3's loading, again with item 1 as the reference. It tests whether the relative loading of item 3 is equivalent across groups.

Y-axis label: *Loading Diff. for Item 3 (Ref: Item 1)*

3. $\frac{\lambda_{3.1}^{g1}}{\lambda_{2.1}^{g1}} - \frac{\lambda_{3.1}^{g2}}{\lambda_{2.1}^{g2}}$ this ratio compares the loading of item 3 relative to item 2 across groups.

It serves as an alternative anchoring check (using item 2 as the reference) to ensure that invariance holds regardless of which item is selected as the anchor.

Y-axis label: *Loading Ratio Diff. for Item 3 vs. Item 2*

3 For a detailed mathematical derivation of these ratio-based constraints, see Cheung and Lau (2012), Appendices B and D.

4. $\nu_2^{g1} - \nu_2^{g2}$ this indicates the difference in the intercept for item 2 between groups when item 1 is used as the reference (with its intercept fixed). Under scalar invariance, the baseline (intercept) of item 2 should be identical across groups.

Y-axis label: *Intercept Diff. for Item 2 (Ref: Item 1)*

5. $\nu_3^{g1} - \nu_3^{g2}$ this is the difference in the intercept for item 3 between groups, with item 1 as the reference. It reflects whether the baseline response level for item 3 is comparable across modes.

Y-axis label: *Intercept Diff. for Item 3 (Ref: Item 1)*

6. $\Delta_{\text{int},3} - \left(\frac{\lambda_{3,1}}{\lambda_{2,1}}\right)\Delta_{\text{int},2}$ here, $\Delta_{\text{int},3}$ is the difference in intercept for item 3 and $\Delta_{\text{int},2}$ is the difference in intercept for item 2 (both computed with item 1 as the reference). Multiplying $\Delta_{\text{int},2}$ by the ratio $\lambda_{3,1}/\lambda_{2,1}$ re-scales the intercept difference for item 2 to reflect the relationship between items 3 and 2. This adjusted difference tests whether, when item 2 is used as the reference, the intercept for item 3 remains invariant.

Y-axis label: *Adjusted Intercept Diff. for Item 3 (Ref: Item 2)*

Table 6 Overview of findings by country, concept and type of invariance

	Social trust		Attitudes towards immigration	
	Metric	Scalar	Metric	Scalar
Estonia	invariance	partial invariance	non-invariance	–
Finland	invariance	invariance	partial invariance	–
Iceland	invariance	non-invariance	invariance	invariance
Italy	invariance	invariance	invariance	invariance
Netherlands	invariance	invariance	non-invariance	–
Norway	invariance	invariance	invariance	partial invariance

The graphs show differences in the parameters of interest, point estimates of loadings for metric invariance models and point estimates of intercepts for scalar invariance models. We include their confidence intervals. We assess if the estimates and their confidence intervals include zero or not. Where confidence intervals overlap with zero, we conclude invariance, because there are no significant differences across parameters. Where zero falls outside of the confidence intervals for one parameter comparison, we regard this as partial invariance. Where zero falls outside of the confidence intervals for at least two parameter comparisons, we regard this as non-invariance. In the graphs, for the cases where metric invariance is not established, intervals are not shown for scalar invariance, as it is not even tested for, and the respective cells remain empty in Table 6, as well.

For Italy, we find metric as well as scalar invariance across the two modes for both concepts as Figure 2 does not show statistically significant differences in parameters. For Norway, both metric and scalar invariance is established for the concept of social trust (see Figure 3). For attitudes towards immigration, metric and partial scalar invariance is established. Iceland shows a different result, in that metric and scalar invariance are established for attitudes towards immigration but only metric invariance for the concept of social trust (see Figure 4). For Finland, we find metric and scalar invariance across the two modes for social trust but we conclude only partial metric invariance for the concept of attitudes towards immigration (see Figure 5). For the Netherlands, results show metric and scalar invariance for the concept of social trust, but metric non-invariance for the concept of attitudes towards immigration (see Figure 6). The latter also holds for Estonia (see Figure 7). Lastly, for the concept of social trust, metric and partial scalar invariance could be established for Estonia (see Figure 7).

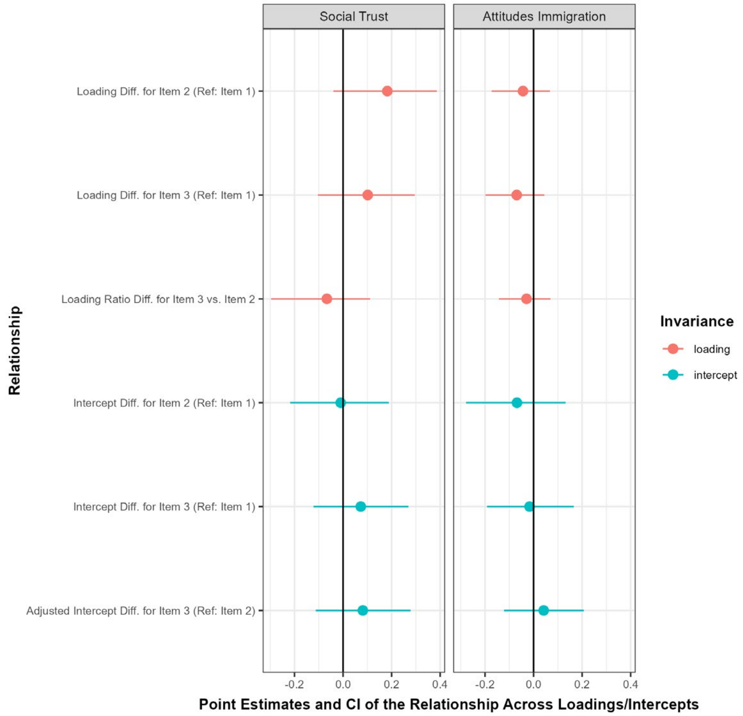


Figure 2 Results invariance testing Italy

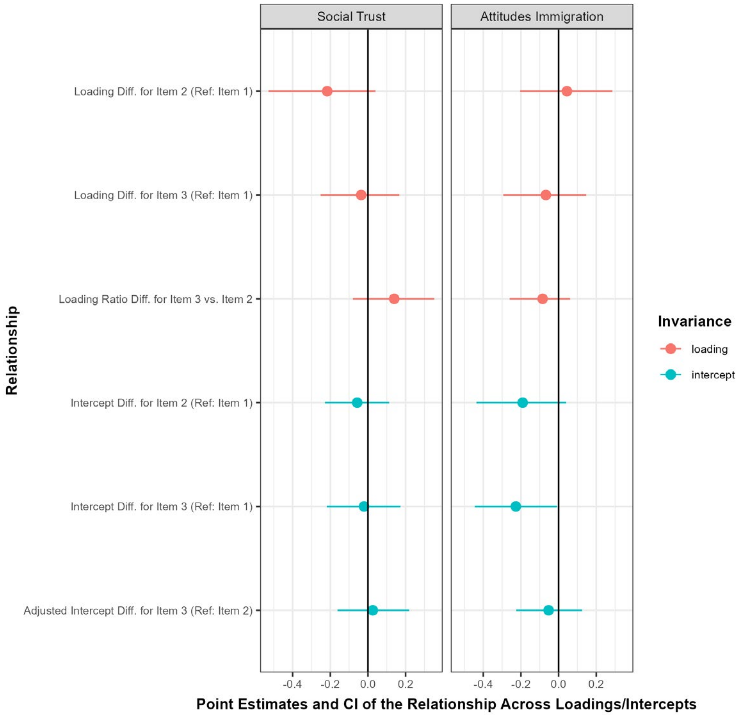


Figure 3 Results invariance testing Norway

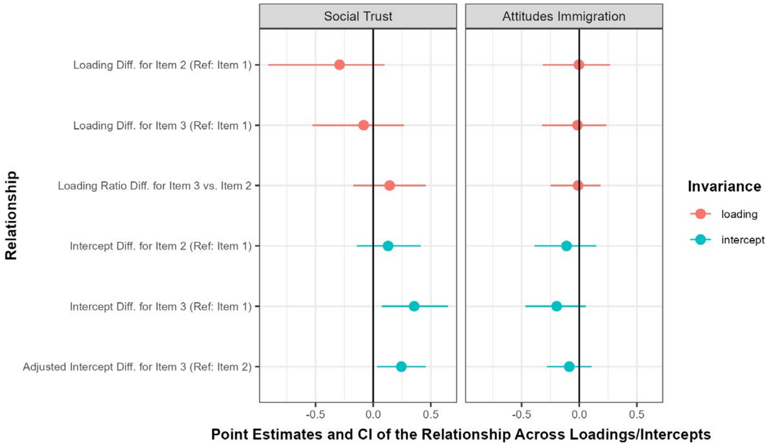


Figure 4 Results invariance testing Iceland

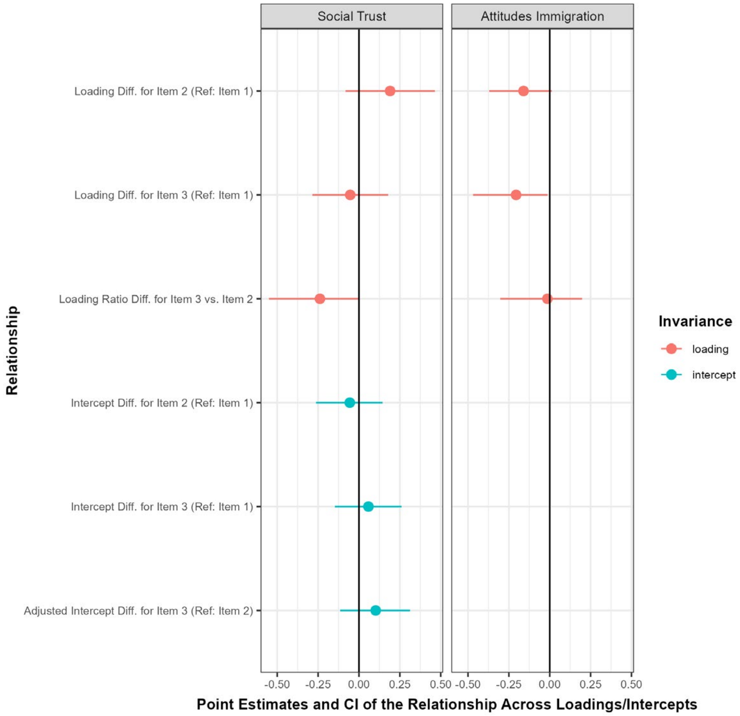


Figure 5 Results invariance testing Finland

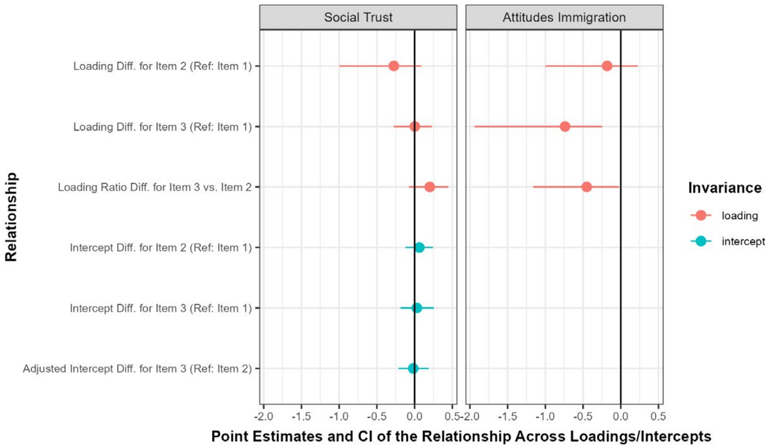


Figure 6 Results invariance testing Netherlands

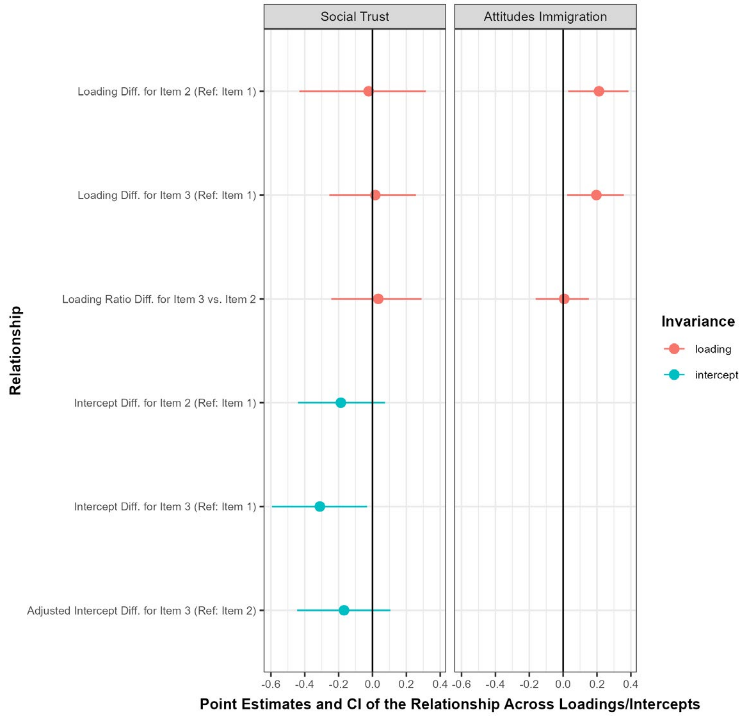


Figure 7 Results invariance testing Estonia

Discussion and Conclusion

The COVID-19 pandemic increased the diffculties of collecting high-quality human data, especially by the means of in-person interviews. New interview formats that do not require respondents and interviewers being in the same place emerged or consolidated, and are nowadays still in use. As noted earlier, ESS Round 10 offered respondents the option of an online video interview as an alternative to an in-person interview (Ghirelli et al., 2024). A body of research also emerged, such as this special issue, evaluating the costs and benefits of the new interview formats and their potential effects on the answers (Deakin & Wakefield, 2014; Oates et al., 2022). We contribute to this research using the ESS Round 10 data and studying the measurement equivalence of data collected via live video and in-person interviewing. As respondents were not randomly allocated to the two groups, but rather self-selected into an interview mode, it cannot be assumed that the data would be equivalent, hence this should be tested (Meredith, 1993). Ensuring the equivalence of the data derived using these modes is crucial for their effective employment in the analysis of social attitudes. This is especially important when the characteristics of the concepts make them prone to bias by response behaviors. The concepts of social trust and attitudes towards immigration differ in the level of social desirability bias among them, with attitudes towards immigration showing larger bias (Creighton et al., 2019).

Measurement invariance between live video and in-person interviews is particularly relevant when considering how different modes of data collection influence response behavior. Video interviewing and in-person interviewing differ regarding the interaction with the administrator and regarding tools used in the interviews, mainly, the intermediation of a screen in video interview. We argue that for concepts where social desirability is present, such as attitudes towards immigration, it is less pronounced in video interviews. Social desirability is lower in modes of data collection where the social presence of an interviewer is weaker (Heerwegh, 2009). The social presence may be perceived as weaker in a context when respondent and interviewer communicate via videochat (Moallem, 2015) compared to a context where the interviewer is physically present in a room with respondents.

Overall, the results can be seen as encouraging. In general, measurement invariance can indeed be established between in-person and video interview mode, across countries and concepts in the ESS Round 10 data. In consequence the data is highly comparable. Particularly for Italy, both metric and scalar invariance are established for both assessed concepts. In practical terms, this means that the data can be aggregated and used as a single dataset for the analysis of the concepts of interest, namely social trust and attitudes towards immigration, and that the dataset can be aggregated as well for the analysis of the means of the concepts.

Measurements for the concept of social trust have shown full metric invariance across in-person and video interviewing modes for all six assessed countries: Estonia, Finland, Iceland, Italy, the Netherlands and Norway (see overview in Table 6). This means that each item contributes consistently to the concept of social trust across modes and countries. The measurements for the concept of social trust are full scalar invariant across video and in- person mode for four out of the six countries (Finland, Italy, the Netherlands and Norway), with Estonia showing partial invariance and Iceland non-invariance. This means that, only for four out of the six countries, the data can be aggregated, and observed mean responses can be computed across the two modes.

For the concept of attitudes towards immigration, the results are somewhat less encouraging. Three out of the six countries show full metric invariance across mode groups. Where full metric invariance is present, the structure of the underlying construct is preserved regardless of the mode of data collection. As a result, analysts can confidently combine data from in-person and video interviewing to study the contribution of each item into the construct in Iceland, Italy and Norway, but not in Estonia, Finland and the Netherlands. In cases where metric invariance is not established, we would advise using latent models, or, as a less optimal solution because it decreases the number of cases, excluding observations from the video interview mode since the relationship of the variables differs across modes. In two out of the three cases showing metric invariance, scalar invariance is also established, namely for Iceland and Italy. For Norway, only partial scalar invariance could be established. Where scalar invariance is not established, we advise against aggregating the data of the two modes to calculate mean values. Instead, analysts could calculate the latent means, taking into account the method variance contribution of both mode groups. This is because the differences in the mean response to items may not purely reflect true differences in the underlying construct across groups, and method variance should be accounted for.

Our findings are neither homogeneous across countries nor across concepts. Measurement invariance across in-person and video interview mode are sensitive to the concept being measured.

It has been found before that measurement invariance is not easily established for attitudes towards immigration if measured in different contexts. Nickel and Weber (2024) study the concept of attitudes towards immigration, and report that metric invariance was established for most ESS countries in Rounds 1 to 9 but scalar invariance only for about half of them. Measurement invariance seems sensitive to contextual influences. Our findings are in line with this result, though the different contexts in our study refer to different modes within countries. Throughout this article, we have presented the argument that it is precisely because of the differences in social desirability between the two concepts that we observe differences in the levels of invariance achieved. On the one hand, attitudes towards immigration presents a larger bias than social trust, on the other, video interviews imply a weaker social presence of the interviewers, than in-person interviews, potentially reducing this bias.

We indeed observe higher levels of invariance for social trust, this means that the data is more comparable than the data for attitudes towards immigration. Our interpretation for this result is that in video mode, respondents feel less pressured to give a socially expected answer, therefore the data is less comparable to an in-person interview. However, more research is needed to explore whether this interpretation is correct.

There are other potential explanations. It could be revealing to investigate questionnaires, interviewer instructions, protocols and similar documentation of the cases where invariance was not consistently reached (especially Estonia, Finland, Iceland and the Netherlands) to see if one can find hints as to how the implementation of these modes might have differed here from those countries where invariance was consistently reached.

Assessing measurement invariance in video and in-person interviews and assessing mode effects are two different subjects of study. In this research, we cannot test simultaneously for mode effects and measurement invariance across modes. Testing for mode effects would require random allocation of respondents into either video or in-person mode. Invariance testing assesses for the comparability of the data in the presence of non-random allocation into groups, in which case characteristics of the survey participants, such as age, gender, employment status, home ownership, and ethnicity, may present differences across interviewing modes (Rowen et al., 2022).

References

Adam, F. (2008). Mapping social capital across Europe: findings, trends and methodological shortcomings of cross-national surveys. *Social Science Information, 47*(2), 159-186. https://doi.org/10.1177/0539018408089077

Boal, A. L., Vaughan, C. A., Sims, C. S., & Miles, J. N. V. (2017). Measurement invariance across administration mode: Examining the posttraumatic stress disorder (PTSD) checklist. *Psychological Assessment, 29*(1), 76–86. https://doi.org/10.1037/pas0000301

Bollen, K. A. (1989). *Structural equations with latent variables*. John Wiley & Sons. https://doi.org/10.1002/9781118619179

Buerger, S., Kroehne, U., & Goldhammer, F. (2016). The transition to computer-based testing in large-scale assessments: Investigating (partial) measurement invariance between modes. *Psychological Test and Assessment Modeling, 58*(4), 597–616.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*(3), 456–466. https://doi.org/10.1037/0033-2909.105.3.456

Ceobanu, A. M., & Escandell, X. (2010). Comparative analyses of public attitudes toward immigrants and immigration using multinational survey data: A review of theories and research. *Annual Review of Sociology, 36*, 309-328. https://doi.org/10.1146/annurev.soc.012809.102651

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Cheung, G. W., & Lau, R. S. (2012). A direct comparison approach for testing measurement invariance. *Organizational Research Methods, 15*(2), 167–198. https://doi.org/10.1177/1094428111421987

Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. J*ournal of Management, 25*(1), 1-27. https://doi.org/10.1177/014920639902500101

Claassen, J., Lenzner, T., Höhne, J. K., & Ziller, C. (2026). A survey mode of the future? Investigating respondents' willingness to participate in self-administered video-based web surveys. *Methods, Data, Analyses, 20*(1), 317–346. https://doi.org/10.12758/mda.2025.10

Conrad, F. G., Schober, M. F., Hupp, A. L., West, B. T., Larsen, K. M., Ong, A. R., & Wang, T. (2023). Video in survey interviews: Effects on data quality and respondent experience. *Methods, Data, Analyses, 17*(2), 135–170. https://doi.org/10.12758/mda.2022.13

Creighton, M. J., Schmidt, P., & Zavala-Rojas, D. (2019). Race, wealth and the masking of opposition to immigrants in the Netherlands. *International Migration, 57*, 245-263. https://doi.org/10.1111/imig.12519

Davidov, E., & Semyonov, M. (2017). Attitudes toward immigrants in European societies. *International Journal of Comparative Sociology, 58*(5), 359-366. https://doi.org/10.1177/0020715217732183

Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review Sociology, 40*, 55-75. https://doi.org/10.1146/annurev-soc-071913-043137

Davidov, E., Cieciuch, J., Meuleman, B., Schmidt, P., Algesheimer, R., & Hausherr, M. (2015). The comparability of measurements of attitudes toward immigration in the European social survey: Exact versus approximate measurement equivalence. *Public Opinion Quarterly, 79*(S1), 244–266. https://doi.org/10.1093/poq/nfv008

de Villiers, C., Farooq, M. B., & Molinari, M. (2022). Qualitative research interviews using online video technology – challenges and opportunities. *Meditari Accountancy Research, 30*(6), 1764–1782. https://doi.org/10.1108/MEDAR-03-2021-1252

Deakin, H., & Wakefield, K. (2014). Skype interviewing: Reflections of two PhD researchers. *Qualitative Research, 14*(5), 603-616. https://doi.org/10.1177/1468794113488126

Delhey, J., & Newton, K. (2005). Predicting cross-national levels of social trust: Global pattern or nordic exceptionalism? *European Sociological Review, 21*(4), 311–327. https://doi.org/10.1093/esr/jci022

Durrant, G., Kocar, S., Brown, M., Hanson, T., Sanchez, C., Wood, M., Taylor, K., Tsantani, M., & Huskinson, T. (2024). *Live video interviewing: Evidence of opportunities and challenges across seven major UK social surveys* (Survey Futures Working Paper No. 1). University of Essex, Institute for Social and Economic Research. https://www.iser.essex.ac.uk/wp-content/uploads/files/working-papers/survey-futures/2024-01.pdf

Endres, K., Hillygus, D. S., DeBell, M., & Iyengar, S. (2023). A randomized experiment evaluating survey mode effects for video interviewing. *Political Science Research and Methods, 11*(1), 144–159. https://doi.org/10.1017/psrm.2022.30

European Social Survey European Research Infrastructure (ESS ERIC). (2023). *ESS10 integrated file*, edition 3.2 [Data set]. Sikt – Norwegian Agency for Shared Services in Education and Research. https://doi.org/10.21338/ess10sce03_2

Faucett, H. A., Lee, M. L., & Carter, S. (2017). I should listen more: Real-time sensing and feedback of non-verbal communication in video telehealth. *Proceedings of the ACM on Human-Computer Interaction, 1*(CSCW), 1–19. https://doi.org/10.1145/3134679

Fernandez, E., Woldgabreal, Y., Day, A., Pham, T., Gleich, B., & Aboujaoude, E. (2021). Live psychotherapy by video versus in-person: A meta-analysis of efficacy and its relationship to types and targets of treatment. *Clinical Psychology & Psychotherapy, 28*(6), 1535–1549. https://doi.org/10.1002/cpp.2594

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis, 1*(3), 515–534. https://doi.org/10.1214/06-BA117A

Ghirelli, N., Lynn, P., Xena, C., Dorer, B., Ambler, M. L., Schwarz, H., Hanson, T., Kappelhof, J., Flore, P., Kessler, G., Lebedev, D., Briceno-Rosas, R., Frank, L.-H., Rød, L.-M., & Øvrebø, O.-P. (2024). *ESS10 overall face-to-face fieldwork and data quality report*. GESIS – Leibniz Institute for the Social Sciences. https://www.europeansocialsurvey.org/sites/default/files/2024-09/ESS10_Quality_Report.pdf

Gordon, H. S., Solanki, P., Bokhour, B. G., & Gopal, R. K. (2020). "I'm not feeling like I'm part of the conversation" Patients' perspectives on communicating in clinical video telehealth visits. *Journal of General Internal Medicine, 35*(6), 1751–1758. https://doi.org/10.1007/s11606-020-05673-w

Gorgievski, M. J., Van der Heijden, B. I. J. M., & Bakker, A. B. (2019). Effort-reward imbalance and work-home interference: a two-wave study among European male nurses. *Work & Stress, 33*(4), 315–333. https://doi.org/10.1080/02678373.2018.1503358

Heerwegh, D. (2009). Mode differences between face-to-face and web surveys: An experimental investigation of data quality and social desirability effects. *International Journal of Public Opinion Research, 22*(1), 111–121. https://doi.org/10.1093/ijpor/edn054

Herreros, F., & Criado, H. (2008). The state and the development of social trust. *International Political Science Review, 29*(1), 53-71. https://doi.org/10.1177/0192512107083447

Inglehart, R. (1999). Trust, well-being and democracy. In M. E. Warren (Ed.), Democracy and Trust (pp. 88–120). chapter, Cambridge: Cambridge University Press.

Johnson, E. C., Meade, A. W., & DuVernet, A. M. (2009). The role of referent indicators in tests of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 16*(4), 642–657. https://doi.org/10.1080/10705510903206014

Joseph C., Garruba M., & Melder A. (2018). Patient satisfaction of telephone or video interpreter services compared with in-person services: A systematic review*. *Australian Health Review, 42*, 168–177. https://doi.org/10.1071/AH16195

Joshi, A., Bloom, D. A., Spencer, A., Gaetke-Udager, K., & Cohan, R. H. (2020). Video interviewing: a review and recommendations for implementation in the era of COVID-19 and beyond. *Academic Radiology, 27*(9), 1316–1322. https://doi.org/10.1016/j.acra.2020.05.020

Kaasa, A., & Parts, E. (2008). Individual-level determinants of social capital in Europe: Differences between country groups. *Acta Sociologica, 51*(2), 145-168. https://doi.org/10.1177/0001699308090040

Kline, R. B. (2016). *Principles and practice of structural equation modeling*. Methodology in the social sciences. Guilford Press.

Letki, N., & Evans, G. (2005). Endogenizing social trust: Democratization in east-central Europe. *British Journal of Political Science, 35*(3), 515–529. https://doi.org/10.1017/S000712340500027X

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4), 525–543. https://doi.org/10.1007/BF02294825

Meuleman, B., Żółtak, T., Pokropek, A., Davidov, E., Muthén, B., Oberski, D. L., Billiet, J., & Schmidt, P. (2022). Why measurement invariance is important in comparative research. A Response to Welzel et al. (2021). *Sociological Methods & Research, 52*(3), 1401-1419. https://doi.org/10.1177/00491241221091755

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. Routledge.

Moallem, M. (2015). The impact of synchronous and asynchronous communication tools on learner self-regulation, social presence, immediacy, intimacy and satisfaction in collaborative online learning. *The Online Journal of Distance Education and e-Learning, 3*(3), 55–77.

Muthén, B., & Asparouhov, T. (2017). Recent methods for the study of measurement invariance with many groups: Alignment and random effects. *Sociological Methods & Research, 47*(4), 637-664. https://doi.org/10.1177/0049124117701488

Nickel, A., & Weber, W. (2024). Measurement invariance and quality of attitudes towards immigration in the European Social Survey. *Methods, Data, Analyses, 18*(2), 213–248. https://doi.org/10.12758/mda.2024.04

Nießen, D., Beierlein, C., Rammstedt, B., & Lechner, C. M. (2020). An English-language adaptation of the interpersonal trust short scale (KUSIV3). *Measurement*

Instruments for the Social Sciences, 2(1), Article e11019. https://doi.org/10.1186/s42409-020-00016-1

Oates, M., Crichton, K., Cranor, L., Budwig, S., Weston, E. J., Bernagozzi, B. M., & Pagaduan, J. (2022). Audio, video, chat, email, or survey: How much does online interview mode matter? *PLoS ONE, 17*(2), Article e0263876. https://doi.org/10.1371/journal.pone.0263876

Piekut, A. (2021). Survey nonresponse in attitudes towards immigration in Europe. *Journal of Ethnic and Migration Studies, 47*(5), 1136–1161. https://doi.org/10.1080/1369183X.2019.1661773

Pirralha, A., & Weber, W. (2020). Correction for measurement error in invariance testing: An illustration using SQP. *PLoS ONE, 15*(10), Article e0239421. https://doi.org/10.1371/journal.pone.0239421

Pokropek, A., Davidov, E., & Schmidt, P. (2019). A Monte Carlo simulation study to assess the appropriateness of traditional and newer approaches to test for measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 26*(5), 724–744. https://doi.org/10.1080/10705511.2018.1561293

Reeskens, T., & Hooghe, M. (2008). Cross-cultural measurement equivalence of generalized trust. Evidence from the European social survey (2002 and 2004). *Social Indicators Research, 85*, 515–532. https://doi.org/10.1007/s11205-007-9100-z

Revilla, M. A. (2012). Measurement invariance and quality of composite scores in a face-to-face and a web survey. *Survey Research Methods, 7*(1), 17–28. https://doi.org/10.18148/srm/2013.v7i1.5098

Rinken, S., Pasadas-del-Amo, S., Rueda, M., & Cobo, B. (2021). No magic bullet: Estimating anti-immigrant sentiment and social desirability bias with the item-count technique. *Quality & Quantity, 55*, 2139–2159. https://doi.org/10.1007/s11135-021-01098-7

Rothstein, B., & Uslaner, E.M. (2005). All for one: Equality, corruption, and social trust. *World Politics 58*(1), 41-72. https://dx.doi.org/10.1353/wp.2006.0022

Rowen, D., Mukuria, C., Bray, N., Carlton, J., Longworth, L., Meads, D., O'Neill, C., Shah, K., & Yang, Y. (2022). Assessing the comparative feasibility, acceptability and equivalence of video-conference interviews and face-to-face interviews using the time trade-off technique. *Social Science & Medicine, 309*, Article 115227. https://doi.org/10.1016/j.socscimed.2022.115227

Saris, W. E., & Gallhofer, I. (2014). *Design, evaluation, and analysis of questionnaires for survey research*. John Wiley & Sons. https://doi.org/10.1002/9781118634646

Schnell, R., & Kreuter, F. (2005). Separating interviewer and sampling-point effects. *Journal of Official Statistics, 21*(3), 389–410.

Schober, M. F., Conrad, F. G., Hupp, A. L., Larsen, K. M., Ong, A. R., & West, B. T. (2020). Design considerations for live video survey interviews. *Survey Practice, 13*(1). https://doi.org/10.29115/SP-2020-0014

Schober, M. F., Okon, S., Conrad, F. G., Hupp, A. L., Ong, A. R., & Larsen, K. M. (2023). Predictors of willingness to participate in survey interviews conducted by live video. *Technology, Mind, and Behavior, 4*(2), 215–229. https://doi.org/10.1037/tmb0000100

Sherman, L. E., Michikyan, M., & Greenfield, P. M. (2013). The effects of text, audio, video, and in-person communication on bonding between friends. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 7*(2), Article 3. https://doi.org/10.5817/CP2013-2-3

Smith, S. N., Fisher, S. D., & Heath, A. (2011). Opportunities and challenges in the expansion of cross-national survey research. *International Journal of Social Research Methodology, 14*(6), 485–502. https://doi.org/10.1080/13645579.2011.611386

Steinmetz, H., Schmidt, P., Tina-Booh, A., Wieczorek, S., & Schwartz, S. H. (2009). Testing measurement invariance using multigroup CFA: Differences between educational groups in human values measurement. *Quality & Quantity, 43*, 599–616. https://doi.org/10.1007/s11135-007-9143-x

Van de Vijver, F., Avvisati, F., Davidov, E., Eid, M., Fox, J. P., Le Donné, N., Lek, K., Meuleman, B., Paccagnella, M., & van de Schoot, R. (2019). *Invariance analyses in large-scale studies* (OECD Education Working Paper No. 201). OECD Publishing. https://doi.org/10.1787/254738dd-en

Van de Vijver, F. J., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology, 54*(2), 119–135. https://doi.org/10.1016/j.erap.2003.12.004

Van der Veld, W. M., & Saris, W. E. (2011). Causes of generalized social trust. In E. Davidov, P. Schmidt, & J. Billiet (Eds.), *Cross-cultural analysis: Methods and applications* (pp. 207–247). Routledge.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4-70. https://doi.org/10.1177/109442810031002

Villar, A., & Fitzgerald, R. (2017). Using mixed modes in survey research: Evidence from six experiments in the ESS. In M. J. Breen (Ed.), *Values and identities in Europe* (pp. 299–336). Routledge.

West, B. T., Ong, A. R., Conrad, F. G., Schober, M. F., Larsen, K. M., & Hupp, A. L. (2022). Interviewer effects in live video and prerecorded video interviewing. *Journal of Survey Statistics and Methodology, 10*(2), 317–336. https://doi.org/10.1093/jssam/smab040

Wu, H., & Estabrook, R. (2016). Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. *Psychometrika, 81*(4), 1014–1045. https://doi.org/10.1007/s11336-016-9506-0

Zmerli, S., & Newton, K. (2008). Social trust and attitudes toward democracy. *Public Opinion Quarterly, 72*(4), 706–724. https://doi.org/10.1093/poq/nfn054

Appendix

Global Fit Indices of Scalar Invariance Models by Concept and Country

Table A1 Global fit indices of scalar invariance models for concept social trust by country

Model	x^2	df	p	rmsea	cfi
Estonia	0.07	2	.96	0.00	1
Finland	4.91	2	.09	0.04	1
Italy	3.84	2	.15	0.03	1
Netherlands	3.33	2	.19	0.03	1
Norway	3.72	2	.16	0.03	1

Table A2 Global fit indices of scalar invariance models for concept attitude towards immigration by country

Model	x^2	df	p	rmsea	cfi
Iceland	0.02	2	.99	0.00	1
Italy	1.75	2	.42	0.00	1
Norway	1.39	2	.50	0.00	1

methods, data, analyses | Vol. 20(1), 2026, pp. 237-268 DOI: 10.25521/mda.661

Live Video Interviewing: Empirical Evidence of Opportunities and Challenges Across Seven Major Social Surveys

Gabriele Durrant[1*], Sebastian Kocar[2*], Matt Brown[3], Tim Hanson[4], Carole Sanchez[3], Martin Wood[5], Kate Taylor[5], Maria Tsantani[5] & Tom Huskinson[6]

[1] *University of Southampton*
[2] *The University of Queensland*
[3] *University College London*
[4] *City St George's, University of London*
[5] *National Centre for Social Research (NatCen)*
[6] *Ipsos*
[*] Joint first authors

© The Author(s) 2026. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Abstract

This paper evaluates the use of video interview (VI) for collecting survey data across seven major UK social surveys that have implemented this mode of data collection, including both cross-sectional and longitudinal surveys. The paper collates rich comparative evidence, on which to evaluate the advancements, opportunities and barriers relating to current and future use of VI in the UK (and elsewhere). The specific aims of the paper are to investigate (a) uptake of, and response rates to, VI compared to other modes; (b) the characteristics of VI respondents to understand representation; and (c) the feasibility of administering particular survey questions and (complex) tasks via VI, such as requesting consent for administrative data linkage and cognitive assessments. Ultimately, the paper aims to evaluate VI's long-term feasibility as a survey data collection mode post-pandemic. The results suggest that, if VI were the only or primary data collection mode, response rates would be notably lower than surveys conducted in person and that this could lead to an increase in sample composition issues and representation bias. There are also encouraging findings. This mode proves to be a suitable approach for obtaining consent for data linkage and conducting cognitive assessments – a key finding because previous research has identified limitations of other remote methods for collecting this kind of data. Overall, the evidence from this study suggests that VI can be a suitable complementary data collection mode in a mixed-mode survey design, offering potentially more cost-efficient fieldwork. We identify particular feasibility advantages for longitudinal surveys.

Keywords: live video interviewing, major UK social surveys, mode selection effects, response rate, representation

Direct correspondence to
Gabriele Durrant, Department of Social Statistics and Demographics, University of Southampton, B58, R2065, University Road, Southampton, SO17 1BJ, United Kingdom
E-mail: g.durrant@southampton.ac.uk

The COVID-19 pandemic led to a rapid move to remote data collection and was a catalyst for further development of both existing and new innovative data collection approaches, including live video interviewing[1] (in the following abbreviated VI; Conrad et al., 2026). In the UK, large-scale social surveys began making use of this method for the first time (e.g., Hanson et al., 2023; Hanson et al., 2026). In addition, in-person[2] surveys have been facing many challenges, including falling response rates, increasing costs and difficulties with recruitment to the interviewer workforce, prompting increased use of remote data collection. However, introducing new survey modes can introduce measurement inequivalence over time, which is particularly problematic for longitudinal studies (Cernat, 2015; Dillman et al., 2009; Dillman & Christian, 2005), or can introduce representation bias if certain sub-groups are less willing to respond via the new mode. Long surveys or those involving specific tasks or complex measures, such as collection of data linkage consents or cognitive assessments, can be difficult to transition to remote modes. Despite emerging research on VI (e.g., Conrad et al., 2023; Endres et al., 2023; Schober et al., 2023; West et al., 2022), our understanding of its use in major social surveys in the UK and elsewhere remains limited. Therefore, it is crucial to further explore the feasibility of VI for data collection post-pandemic.

Aims of the Paper

To address these gaps, this paper reviews evidence and experiences of implementing VI in all of the major social surveys in the UK that have used the mode to date (seven surveys in total). The majority of the surveys are longitudinal, two are cross-sectional and, together, they provide rich evidence on which to evaluate the advancements, opportunities and barriers relating to current and future use of VI in the UK and beyond. The specific aims of this paper are to investigate (a) uptake and response rate to VI, also in comparison to other modes; (b) characteristics of VI respondents in order to explore sample composition and potential impact on representation; and (c) feasibility of administering particular survey questions and tasks via VI, such as asking consent for administrative data linkage and cognitive assessments. The paper is not about UK surveys per se (versus surveys conducted elsewhere) but aims to make, where possible, gen-

[1] In general, three types of VI approaches are being distinguished (only the first option is considered in this paper): live video (personal) interviewing (VI), which resembles in-person interviews but is conducted via a video communication platform; pre-recorded VI, which is a self-interviewing approach using videos embedded in questionnaires (Conrad et al., 2023); and computer-generated agents and/or Artificial Intelligence (AI) supported VI (Höhne, 2024).

[2] In this study, we use the terminology “in-person” interviewing instead of “face-to-face” interviewing when respondent and interviewer meet in the same physical space, since VI is also face-to-face (albeit mediated by a screen).

eralisable contributions to the VI literature. Ultimately, the paper aims to evaluate VI's long-term feasibility as a survey data collection mode.

Research Questions

Given the above gaps in the current literature, the key research questions we aim to address are:

RQ1: *What response rates (uptake) have been achieved with VI, either as the only interview mode or as an alternative to other modes?*

Specifically, we aim to explore whether response rates achieved using VI are comparable with response rates achieved through in-person interviewing. Additionally, we seek to explore how the uptake of VI varies depending on how it is offered, whether as the sole mode of completion or as an option alongside other modes.

RQ2: *What are the characteristics of participants who respond via VI?*

We aim to determine whether VI respondents differ from those who respond via other modes, particularly via in-person interviewing (and if so, to what extent), i.e. we investigate (differences in) sample composition (or representation) between VI and in-person. This is important for assessing potential impacts on representation (due to issues with coverage and/or nonresponse) of survey data collected via VI.

RQ3: *How well does VI perform in administering (complex) survey tasks, in particular obtaining consent for administrative data linkage, as well as administering cognitive assessments?*

These tasks are an important component of many studies but are challenging to implement when interviewers are not physically present (e.g., in web mode). We aim to compare VI consent and completion rates with those achieved in in-person interviews to assess whether VI could be an effective mode for collecting these measures.

All seven UK surveys examined here introduced VI between 2020 and 2023, using it either as the primary mode or as an alternative, either in parallel with other modes or in sequence. All studies initially implemented VI as a result of the COVID-19 pandemic which prevented in-person contact. The goal was to replicate in-person data collection as much as possible. Additional reasons for implementing VI were to increase response rates and attract respondents who might not have participated otherwise. These surveys, each with different designs, topics, and target populations, provide a rich source for exploring VI.

Whilst the method of VI had been explored to a limited degree before the pandemic (e.g., Anderson, 2008; Jeannis et al., 2013; Schober, 2018) – including in qualitative research (e.g., Deakin & Wakefield, 2014; Irani, 2019) – the approach has been evaluated in a number of studies since then (e.g., Conrad et al., 2023; Endres et al., 2023; Phillips et al., 2023). During the pandemic, many traditionally in-person surveys explored VI. Post-pandemic, while some surveys reverted to in-person methods, others continued to offer VI as an option. A key attraction of VI is that VI shares many properties with in-person interviewing (Conrad et al., 2023; Schober et al., 2023), implying measurement differences between the two modes should be minimal (Endres et al., 2023). This mode presents a cost-effective alternative to in-person interviewing with several advantages, potentially increasing overall response rates and representation, which will be explored further in this paper.

Review of Existing Evidence

Primary concerns when comparing VI to other modes include survey completion, recruitment outcomes and data quality (Anderson, 2008). In the following, we review the existing literature, elaborating on the factors associated with (a) (unit) nonresponse rate (uptake of VI), (b) representation, and (c) the collection of relatively “complex” survey tasks, relating to RQ1-3 respectively.

Response Rate and Uptake of VI (RQ1)

Response rates to VI could differ from other modes, such as in-person interviewing for various reasons, including scheduling difficulties (Hupp et al., 2026), technical issues like internet connectivity and usability of the video platform, managing distractions remotely, and privacy concerns (Kunz et al., 2024; Schober et al., 2020; Van Zeeland et al., 2021). Despite these challenges, rapport can be established as effectively in VI as in in-person surveys (Sun et al., 2021), and whilst evidence suggests that respondents still tend to prefer in-person interviews over VI (Selvam et al., 2020), certain subgroups such as those requiring sensory assistance (Schober et al., 2020) may prefer VI. Offering VI alongside other modes can, therefore, potentially increase overall response rates (Jeannis et al., 2013). However, the existing evidence suggests that in cross-sectional surveys the challenges around implementation of VI are more likely to negatively impact unit response rates. Guggenheim and Howell (2021) report that VI, as the first mode in a mixed-mode design, yielded lower responses than web-only or a mixed mode approach involving web and phone. Similarly, Conrad et al. (2023) report substantially lower VI response rates than those achieved in a conventional web survey and “pre-recorded video interviews”, in which video recordings of interviewers asking survey questions are embedded in a web survey, in two nonprobability online panel samples. Additionally, Phillips et al. (2023) reported lower VI response rates compared to five other survey modes. These low VI response rates are likely explained by lack of familiarity with the mode, but as video calls become ever more commonplace in work and day-to-day lives, then willingness to take part via VI could increase. In the UK context, evidence on participation in surveys using VI, for both cross-sectional and longitudinal studies, has been very scarce so far. With respect to RQ1, given that the seven surveys examined in our study were all conducted during a time when VI was emerging as a new mode in the UK we anticipate that VI response rates will be lower than in-person response rates.

Sample Composition and Representation with VI (RQ2)

The discussed variation in VI uptake may result in differing sample compositions between respondents who are able and open to participation via VI and those who are not. VI requires both access to the necessary technology, e.g. internet connection and hardware, and the willingness to participate via that mode (i.e., nonresponse-related issues). Importantly, not everyone has access to video communication technology even if they have access to the internet (Schober et al., 2020), with internet access now being very high (Cabrera-Álvarez et al., 2025). Additionally, internet access and use vary by individual or household characteristics and between geographical regions, at least in a country like the UK (Office for National Statistics, 2021). This implies that individuals from certain socio-demographic groups may face "technological" barriers to participating via VI, despite being willing to do so. To mitigate such representation issues from under-coverage of people without access to the internet or the necessary hardware to communicate via video, providing resources to those without these could be considered. However, this could make data collection more expensive and could risk participants feeling pressured or coerced into taking part in a way they do not prefer which could be an ethical concern (Schober at al., 2020). More specifically, respondents might not be comfortable participating in VI for other reasons including perceived difficulty of use, lack of value or enjoyment, and discomfort disclosing sensitive information via VI (Schober et al., 2023).

On the other hand, some respondents who might be reluctant to let an interviewer into their home (e.g., due to health concerns or safety concerns) might be willing to participate in VI. VI may also offer more flexibility in terms of making appointments (e.g., timing and location); for example, VI interviews could more easily be conducted at a workplace. Certain demographic groups, including those with more technological experience (e.g., younger sample members), could potentially be more attracted to VI (Anderson, 2008), so the mode could potentially be advantageous for drawing in groups otherwise less likely to take part.

Recent evidence on sample composition differences between VI and other survey modes is mixed, which might be associated with the types of surveys that explored the differences, as well as differences between countries. In the U.S., samples often skewed towards more educated individuals (Conrad et al., 2023, p. 145; Dulaney et al., 2023; Guggenheim & Howell, 2021; Martin & Fradier, 2023; Phillips et al., 2023), and younger participants (Dulaney et al., 2023; Martin & Fradier, 2023; Phillips et al., 2023; Thorolfsson et al., 2023). In contrast, Schober et al. (2023) found no socio-demographic predictors for hypothetical VI participation. Evidence on gender representation is mixed. Guggenheim and Howell (2021) reported consistent overrepresentation of women, with this finding also applying across other modes, including in-person, and web. On the other hand, Martin and Fradier (2023, France) and Phillips et al. (2023, Australia) report an overrepresentation of men in their VI samples.

In summary, with respect to RQ2, while it might be challenging to predict any differences by sex, we expect that younger respondents and those holding a university degree will be more likely to participate via VI than older or less educated respondents.

Complex Survey Tasks with VI (RQ3)

In addition, it is important to understand how VI performs in the collection of particular survey questions and tasks, such as requesting consent for administrative data linkage and administering cognitive assessments. VI may offer advantages over other remote methods, such as web, since in VI interviewers are able to explain these complex tasks, tailoring what they say to particular respondent concerns and confusions, which is not feasible in web surveys. There may also be challenges, such as the difficulty of obtaining consent without the interviewer's physical presence.

Obtaining consent in social surveys, such as for linking administrative data, has been extensively studied in both cross-sectional and longitudinal surveys. However, to the best of our knowledge, there is no evidence on consent rates collected via VI. Nevertheless, the existing literature on requesting consent via different modes consistently shows that consent rates tend to be higher in interviewer-administered modes than in self-administered modes, such as mail or online modes (Al Baghal et al., 2020; Jäckle et al., 2021; Jäckle et al., 2022; Jäckle et al., 2023; Sakshaug et al., 2017; Thornby et al., 2018). Evidence on differences in consent rates between interviewer-administered modes is more mixed with some studies reporting higher consent rates in in-person interviews than telephone surveys (Al Baghal et al., 2020; Sakshaug et al., 2012) and others finding no difference (Thornby et al., 2018). Based on this evidence, we could expect that data linkage consent rates will be lower for VI than for in-person respondents, but higher than via web.

The use of VI for collecting survey data that are more complex, such as cognitive tests, open occupation and industry descriptions, and event histories, has not yet been sufficiently explored (although some literature starts to exist, e.g., Fomby et al., 2026, discussing achievement assessment of school children via VI). VI has the potential to offer certain benefits, as existing literature suggests that interviewer-administration is generally more suitable for administering such measures than self-administration (e.g., Emery et al., 2023; Ofstedal et al., 2021; Peycheva et al., 2021). Previous research comparing in-person and web modes has shown measurement mode effects when assessing cognitive ability (Al Baghal, 2019; Gooch, 2015; Ofstedal et al., 2021), psychological functioning (Zager Kocjan et al., 2023), and quality of life (Tsakos et al., 2008). VI may help mitigate some of these issues by retaining the advantages of interviewer administration while allowing data to be collected remotely, although in one test of a complex data collection conducted via VI, Silber et al. (2024) report relatively low response rates and differential nonresponse when collecting complex egocentric network data. Therefore, with respect to RQ3, we expect to find in our study similar evidence of lower completion rates for specific or more complex survey tasks. This expectation particularly applies to tasks that require additional effort from VI respondents to complete, compared to in-person participants (such as a requirement to post a letter cancellation task).

Data

The seven studies used in this study are: the National Child Development Study (NCDS), the 1970 British Cohort Study (BCS70), Next Steps, Children of the 2020s (Cot20s), the English Longitudinal Study of Ageing (ELSA), the UK European Social Survey (ESS) and the Health Survey for England (HSE). Table 1 provides a detailed overview of these seven surveys, their different design features and the different implementations of VI. Implementation of VI across the surveys investigated varied. NCDS, BCS70 and ELSA all used VI as the sole (or primary) mode of data collection for a period of time, whilst the others offered VI as an additional mode to maximise response. In Next Steps, VI was only offered to those who had not responded via web. In some studies, VI was offered to participants living in particular geographical areas where in-person interviewing was not possible (as no in-person interviewers were available) or not permitted due to national restrictions (e.g., Northern Ireland in ESS).

As a result, the number of VI interviews and the share of the total interviews achieved via VI differed across the studies. For instance, in the BCS70 Age 51 Survey, 3,467 interviews were conducted using VI (48% of all interviews). However, in Cot20s, only 99 (about 1%) of all interviews were conducted using VI, and in ESS Round 10, only 55 VI interviews were carried out (about 5% of all UK interviews). It is worth noting that the ESS is a cross-national comparative study, and in some other countries, the share of VI interviews was notably higher, with six countries conducting at least 15% of all interviews via VI, and Iceland having the highest share at 37% (European Social Survey European Research Infrastructure, 2023; Hanson et al., 2026). For more information about study designs, please refer to the Centre for Longitudinal Studies (no date, references, a-d), English Longitudinal Study of Ageing (no date.), National Health Service (no date), and European Social Survey (no date).

Analytical Methods

To examine the feasibility of VI, we explore three key data quality aspects: unit response rate, representation, and completion of specific or complex survey tasks (data linkage consent, cognitive assessments), comparing VI to the other survey modes (mostly in-person). Importantly, the analytical possibilities and relevance were shaped by the characteristics of the available data across the seven studies and the sample sizes of VI respondents. Specifically, the unit response rate analysis (RQ1) can be conducted for all seven studies. However, to address RQ2 and RQ3, composition of sample (representation) and response to particular complex survey tasks can only be explored in the three studies with a sufficient VI sample size (i.e., NCDS, BCS70, and ELSA).

Analysis of Response Rates (Uptake of VI) (RQ1)

To explore RQ1 we investigate the impact of VI on participation rates. NCDS, BCS70 and ELSA all had periods where VI was offered as the sole mode with non-respondents later being invited to take part in person. Once in-person interviewing was re-launched, it became the primary mode again, with VI offered as an alternative mode. We compare response rates (RR) for those initially invited to take part during the VI-only phase with those invited to take part in person (with VI available as alternative mode if requested). For Next Steps, Cot20s, HSE, and ESS, where VI was implemented as a supplementary survey mode with the primary aim to increase overall unit response, we show VI respondents as a proportion of the total interviews achieved (see Figure 1), also in comparison to the other modes offered.

Analysis of Sample Composition and Representation (RQ2)

To address RQ2 we compare the demographic composition of NCDS, BCS70 and ELSA respondents who participated via VI with those who participated via in-person interviewing, even if initially invited to take part in the other mode. Fieldwork for NCDS was initially launched in January 2020 and conducted entirely face-to-face. Data collection was suspended in March 2020 before being

Table 1 Overview and characteristics of the seven UK surveys investigated in this study, employing VI between 2020 and 2023.

Study	Wave/ Round	Sample or cohort information	Study management	Fieldwork agency	Data collection period (VI)	VI designs/conditions	Sample sizes
National Child Development Study (NCDS)	Age 62 (Longitudinal)	The birth cohort in Great Britain was drawn from a singular week of births that occurred in 1958. This cohort represents the general population and was not clustered.	Centre for Longitudinal Studies, University College London	NatCen and Verian (formerly Kantar Public)	January 2020 – November 2023	Survey initially launched in-person interviewing in January 2020 but paused in March due to restrictions. Re-launched with VI-only approach between Autumn 2021 and Spring 2022. VI then offered as alternative to in-person interviewing between Spring 2022 and November 2023.	VI: 2,283 In-person: 5,238
British Cohort Study (BCS70)	Age 51 (Longitudinal)	The birth cohort in Great Britain was drawn from a sin	Centre for Longitudinal Studies, University College London	NatCen and Verian (formerly Kantar Public)	Autumn 2021 – November 2023	Survey launched with VI only approach, Autumn 2021 to Spring 2022. VI then offered as alternative to in-person interviewing between Spring 2022 and November 2023.	VI: 3,467 In-person: 3,721
Next Steps (NS) (previously known as Longitudinal Study of Young People in England, LSYPE)	Age 32 Survey (Longitudinal)	Probability sample with boost of minority ethnic groups. At the start of the survey in 2004, sampled children were 13/14 years of age, and in Wave 9, they were 32 years of age.	Centre for Longitudinal Studies, University College London	Ipsos UK	April 2022 – August 2023	Participants first invited to take part online. VI was one of the modes that was offered to those who did not respond online within 3 weeks.	VI: 8 Other modes (online, in-person, secondary device (tablet), phone): 6,972

Table 1 (continued)

Study	Wave/ Round	Sample or cohort information	Study management	Fieldwork agency	Data collection period (VI)	VI designs/conditions	Sample sizes
Children of the 2020s (Cot20s)	Wave 1 (Longitudinal)	Clustered probability sample of children aged 9 months in England. Interviews in Wave 1 took place as close as possible to the date that children turned nine months old.	University College London	Ipsos UK	June – October 2022	Interviewers were instructed to prioritize in-person interviewing over VI and phone. VI was initially considered as a mode to maximise response rates due to in-person interviewing being restricted or less desired during COVID-19.	VI: 99 Other modes (in-person, telephone): 8,469
English Longitudinal Study of Ageing (ELSA)	Wave 10 (Longitudinal)	Individuals in England, originally selected using a two-step probability sampling method for the Health Survey for England and aged 50 years or older at the start of the survey in 2002 or during the subsequent refreshment waves (3, 4, 6, 7, or 9).	University College London	NatCen	Late 2020 – 2022	VI-only approach was used during the early phase when in-person interviewing was not allowed. Subsequently, VI was a second option (in-person-first).	VI: 1,213 Other modes (in-person, telephone): 5,050
European Social Survey (ESS) (United Kingdom)	Round 10 (UK) (Cross-sectional)	Random probability sampling was carried out to generate a general population cross-sectional sample comprising individuals aged 15 and above.	City, University of London	NatCen	August 2021 – September 2022	Initially, VI could only be offered following a refusal to take part in an in-person interview. Later in fieldwork, VI could be offered at first contact as an equal option to in-person interviewing.	VI: 55 In-person: 1,090
Health Survey for England (HSE)	2021 pilot (Cross-sectional)	HSE is representative of the population living in private households in England. It uses the Postcode Address File as the sampling frame within a multi-stage stratified probability sampling design.	NatCen, University of London	NatCen	Early 2021	VI was employed in a pilot, which adopted an opt-in approach with sample members invited by letter to prompt them to provide contact details via an online form or a freephone number. In the first stage of the pilot, interviewers proactively suggested VI, and in the second stage, participants could choose between telephone or VI.	VI: 51 Telephone: 747

re-launched in November 2021, initially using a video-only approach before in-person interviewing re-commenced in April 2022. The sample for the NCDS analysis comprises those who were invited to take part after the study was re-launched (*N* = 7,649). The sample for the BCS70 includes all those who were invited to take part during the main stage of fieldwork (*N* = 10,824), and the sample for ELSA consists of all in-person and VI participants, excluding a small subsample of CATI respondents. The selected demographic variables, namely sex, highest qualification/education, and economic activity, were included due to their availability and comparability across the three studies. The age variable, which was provided by the fieldwork agency in a categorical format, is only relevant for ELSA, since BCS70 and NCDS are birth cohort studies where all participants were born in the same week. To test differences in sample composition, Chi-Square testing was conducted, and Phi is reported as the effect size measure.

Analysis of Response to Consent Requests and Cognitive Assessments (RQ3)

To investigate RQ3 we explore the performance of VI in conducting specific (or complex) survey tasks. NCDS, BCS70 and ELSA participants who had not consented in previous waves of the study were asked to consent to link their survey data with (a) health records held by the UK National Health Service (NHS)[3] and economic records held by (b) the Department for Work and Pensions (DWP) and (c) Her Majesty's Revenue and Customs data. We use chi-square testing to confirm statistically significant differences in consent rates between modes. NCDS, BCS70 and ELSA also included a series of cognitive assessments: (a) immediate word list recall test, (b) delayed word list recall test, (c) animal naming and (d) letter cancellation[4]. The first three tasks were scored by the interviewer during the interview. The letter cancellation task, on the other hand, was self-completed on a paper sheet during the interview and scored after being returned to the study team. In VI interviews this sheet was posted to respondents in advance and participants were then required to return this by post after the interview. During in-person interviews the sheet was provided and collected by the interviewer who was responsible for posting this to the study team for scoring. Again, Chi-Square testing was conducted to confirm statistically significant differences in completion rates between modes.

3 In ELSA, respondents were asked for consent to link their survey data to the following NHS data sources: Mortality, Hospital Episode Statistics and Primary Care data. They were also asked to agree to wear an activity monitor.

4 In the letter cancellation task, participants are presented with a grid or sheet of letters arranged in random order. They are instructed to scan the grid and mark/cancel out a specific target letter. Performance on the letter cancellation task is typically measured in terms of accuracy and speed.

Results

In this section, we begin by examining VI uptake and associated response rates for all seven UK surveys. We then explore the socio-demographic composition of both VI and non-VI samples to assess their comparability. We conclude this analysis section by discussing how VI performs in collecting survey data that are traditionally more challenging to obtain in survey modes that are not in-person, specifically securing participant consent and conducting complex cognitive assessments.

Analysis of Response Rates and Uptake of VI (RQ1)

We first present the results of the response rate analysis for the studies where VI was used as the primary mode of data collection and resulted in large samples of respondents participating via VI (NCDS, BCS70, ELSA), and then separately discuss the other studies where VI was included as an additional or supplementary mode, leading to smaller samples of VI participants (Next Steps, Cot20s, HSE and ESS).

VI as a Primary Mode

NCDS. Fieldwork for NCDS Age 62 Survey was conducted in batches, and VI was employed in two different ways. Participants allocated to earlier batches of fieldwork were initially only invited to take part via VI, with non-respondents being invited to take part in person. Later in fieldwork, when in-person interviewing could resume due to the lifting of COVID restrictions, in-person interviewing became the primary mode. Participants allocated to these later batches of fieldwork were invited to take part in person with VI offered as an additional option. First, we present uptake by the mode to which participants were allocated, followed by the overall response rate by allocation mode.

Of the 3,001 eligible participants who were allocated to and invited to take part via VI, just under half (1,432 or 48%) participated through that mode (see Figure 1). In contrast, of the 4,394 eligible participants who were allocated to and invited to take part in person, 3,126 (or 71%) took part via that mode. The response rate for the mode to which cohort members were initially assigned was notably lower for the VI batches (VI: 48%, in-person: 71%, $\phi = .237$, $p < .001$).

Additionally, of the 3,001 eligible participants allocated to earlier batches of fieldwork who were invited to take part via VI, 415 (or 14%) participated in person, giving an overall response rate of 62% (i.e. 48% VI + 14% in-person, see Figure 1). In contrast, of the 4,394 eligible participants allocated to later batches (in-person-first), 29 (or 1%) opted to take part via VI, giving an overall response rate of 72% (i.e., 71% in-person + 1% VI). The overall response rate for those first invited to take part via VI (62%) is significantly lower than the comparable overall response rate amongst those first invited to take part in person (72%, ϕ = .108, p < .001). For more information on implementation of VI in NCDS see Table A1 in Appendix A.

BCS70. The Age 51 Survey implemented VI in the same way as NCDS. Initial batches of fieldwork used a video-first approach, and later batches used an in-person-first approach with VI as an alternative option. Of the 6,114 eligible participants who were invited to the VI-first batches, 2,842 (or 46%) participated via VI. In contrast, of the 4,619 eligible participants who were invited to take part in subsequent in-person-first batches of fieldwork, 2,732 (or 59%) participated in person. The response rate for the mode to which cohort members were initially assigned was notably lower for the VI batches (VI: 46%, in-person: 59%, ϕ = .125, p < .001), which is consistent with our findings from NCDS.

Additionally, of the 6,114 eligible participants who were allocated to VI, a further 830 (or 14%) took part in person, giving an overall response rate of 60% (i.e., 46% VI + 14% in-person). In contrast, of the 4,619 eligible participants who were allocated to in-person interviewing, a further 62 (1%) took part via VI. The overall response rate in the in-person-first group was ultimately the same as that of the VI-first group, at 60% (i.e., 59% in-person + 1% VI, see also Appendix A).

ELSA. The Wave 10 Survey took a similar approach to NCDS and BCS70. Initial batches of fieldwork used a VI-first approach, and later batches used an in-person-first approach with VI available as an alternative mode. Unlike NCDS and BCS70, towards the end of fieldwork, non-respondents (from both VI-first and in-person-first groups) were given an opportunity to participate via telephone; for more information see Table A1 in Appendix A.

The results presented in Figure 1 show that in the VI-first group, only 25% (or 1,041 participants) responded via VI. In contrast, amongst those invited to take part in person first, 66% (or 2,177 participants) were interviewed in person. The response rate for the mode to which cohort members were initially assigned was notably lower for the VI-first group, which is consistent with our findings from NCDS and BCS70.

Additionally, amongst those invited to take part via VI, a further 1,432 were interviewed in person (35%) and 81 via telephone (2%), giving an overall response rate of 68% (i.e., 25% VI + 35% in-person, 2% telephone). Amongst those invited to take part in person first, relatively small proportions were interviewed via telephone (33 or 1%) and via VI (14 or 0.4%). The overall response rate was higher for the in-person-first group (68%) than for the VI-first group (62%) (ϕ = .062, p < .001).

This evidence overall shows that response rates are lower for VI compared to in-person interviewing. In the case of NCDS and ELSA, the overall response rates were higher for respondents invited to take part during the in-person-first batches than for those who were invited to take part in the VI-first batches, although this was not the case for BCS70. However, the proportions of those who were assigned to in-person interviewing and participated via that mode were substantially larger in all three studies, including BCS70, than the proportions of those who were assigned to VI and participated via that mode. In other words, we observe both the negative effect on the propensity to respond via VI compared to in-person interviewing, as well as an increase in the likelihood of responding via the alternative mode if assigned to VI. These findings are further supported by the results presented in the next section.

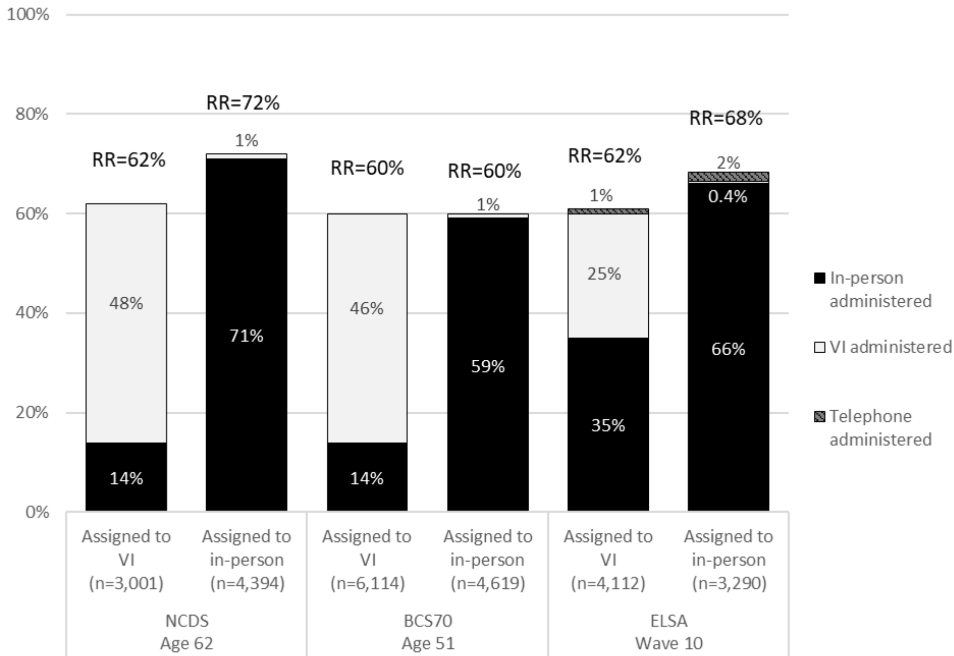


Figure 1 Response rate (RR) in the studies where VI was primary mode – NCDS, BCS70 and ELSA (for during and post-COVID times combined, 2021-2023).

VI as a Supplementary Survey Mode

VI was used as a supplementary mode in Next Steps, Cot20s, ESS and HSE in slightly different ways: Next Steps employed a sequential mixed mode design in which participants were first invited to participate via web. Web non-respondents were then followed up by in-person interviewers who were able to offer VI as an option for completing alongside a standard in-person interview, completion of the web interview via a tablet provided by the survey agency or a telephone interview. In the end, VI contributed only 8 interviews (or 0.1%, see Figure 2), in addition to 6,096 (or 87%) via web (including 159 or 2% on a tablet[5] by completing an online questionnaire), 737 (or 11%) in-person, and 139 (or 2%) via telephone. VI interviews represented a very small proportion of all interviews that were not conducted via the primary survey mode (web).

In Cot20s a concurrent mixed-mode design was used. Interviewers were instructed to prioritise the modes via negotiation with respondents as follows: in-person (top priority), VI (next priority), and telephone (lowest priority). In the end, VI contributed 99 interviews (or 1%), in addition to 7,273 interviews (or 84%) completed via in-person interviewing and 1,256 (or 15%) by telephone. Similar to Next Steps, VI interviews accounted for a small proportion of all achieved interviews.

In the ESS (UK), VI was the only approach offered in Northern Ireland as in-person interviews were not permitted but, in the rest of the UK, an advance letter was followed by an in-person visit, and VI was offered as an alternative mode to in-person interviewing. 55 (5%) participants took part via VI, with the remainder taking part in person.

In the HSE pilot, the fieldwork agency responded to the pandemic-related pause to in-person interviewing by providing two remote options (VI and telephone). For this pilot in early-2021 (January-March)[6], an opt-in approach was adopted for the survey, with sample members invited by postal mail to provide contact information via a portal. In total, 51 individuals participated via VI out of 798 (6%). (For more information see Table A1 in Appendix A.)

The results presented in Figure 2 show that in all of these studies, where VI was offered alongside other modes, the uptake of VI was very low, compared to the other alternative modes (in-person and telephone in NS, telephone in Cot20s). The lowest proportion of respondents participating via VI was in Next Steps (0.1%) and the largest proportion of respondents participating via VI was in the HSE 2021 pilot (6%).

---

5 While the mode of data collection is the same for the standard completion of an online questionnaire and completion on a tablet, there is a notable methodological difference. In Next Steps, tablets were provided and picked up after completion by interviewers, whereas interviewers were not involved in web completion.

6 Following the successful implementation of VI in a pilot for BCS70, an approach was piloted that gave sample members the choice of conducting an interview remotely by telephone or VI. Nevertheless, due to a relatively low take-up and the availability of other remote approaches in HSE, VI was not subsequently used on the main stage of the survey in 2021.

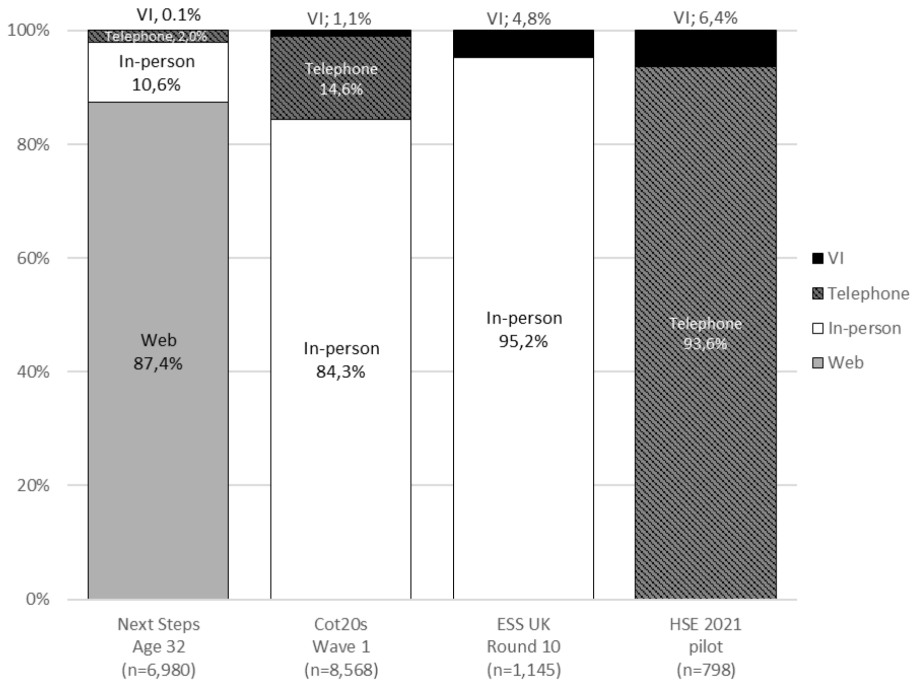


Figure 2 Survey mode in the studies with VI as a supplementary survey mode: uptake (RR) of VI interviews and of other modes (telephone, in-person, web).

Analysis of Composition of VI Samples and Representation (RQ2)

To compare the characteristics of VI and in-person respondents (RQ2), we focus on the studies for which VI was the primary mode for a period of time – NCDS, BCS70, and ELSA (for sample size reasons and variables available). We identified three variables common to the three studies: sex, highest qualification/education, economic activity (for definition and categorisation of the latter see Table B2). We compare the share of in-person and video interviews achieved with females, the employed and those with a degree-level education in Figure 3.

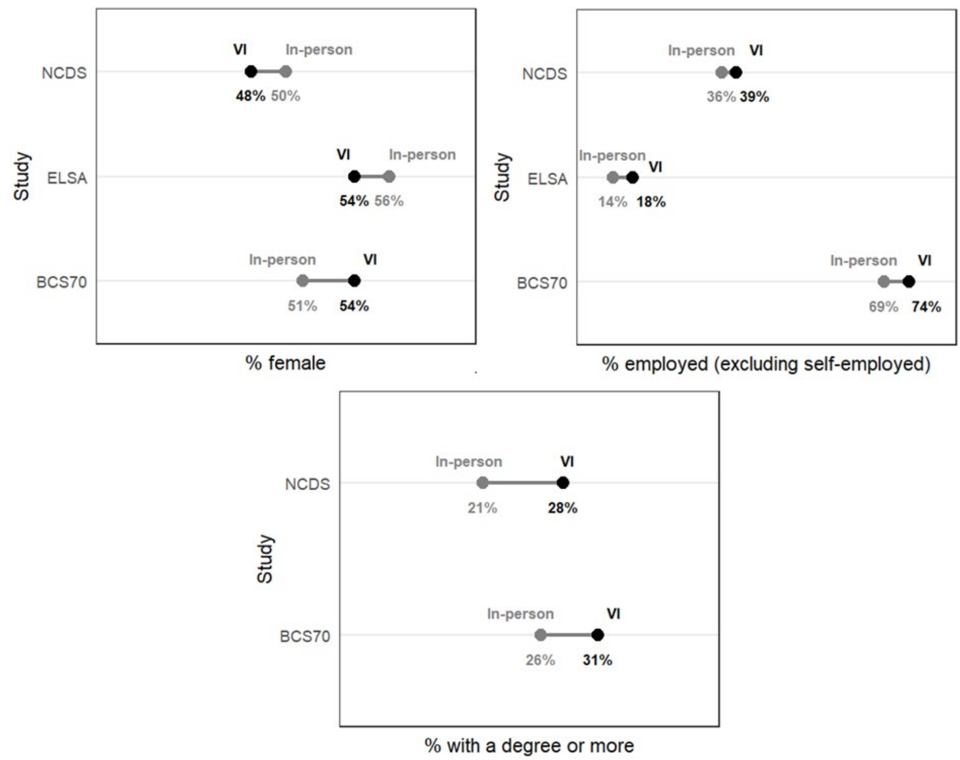


Figure 3 Socio-demographic differences between respondents who were interviewed in person and via VI in three studies – NCDS, ELSA, BCS70 (based on: sex, economic activity (employed), highest qualification/education).

Sex. The analysis revealed minor, statistically insignificant differences in the proportion of women and men who participated via VI and in person in the NCDS and in ELSA. However, women accounted for 54% of VI participants and 51% of in-person respondents in the BCS70, a difference which is statistically significant (ϕ = .026, p = .040).

Economic activity. In BCS70 and ELSA, VI respondents were more likely to be employed (excluding self-employed) than in-person respondents. In BCS70, 74% of VI respondents were employed compared to 69% of in-person respondents (ϕ = .055, p < .001). In ELSA, 18% of VI respondents were employed versus 14% in-person (ϕ = .043, p = .004). The mode difference in employment was not significant for NCDS. In ELSA, retirees were less likely to participate via VI than in person (VI: 70%, in-person: 75%, ϕ = .047, p = .001), consistent with the effect of age below. The same effects were found in NCDS and BCS70 but the difference was small and not significant. Results for economic activity are provided in Table B2 in Appendix B.

Highest qualification. In both NCDS and BCS70[7], VI participants were more likely to have a degree than in-person participants, with the gap being between 5% (BCS70, ϕ = .064, p < .001) and 7% (NCDS, ϕ = .082, p < .001).

Age. The distribution by age is relevant for ELSA only, since the other studies are birth cohort studies in which all participants are the same age. ELSA participants are aged between 50 and 100. When grouped into 5-year age bands, all bands up to age 74 were over-represented amongst VI respondents (compared to in-person respondents) and the reverse was true amongst all bands over 75. Overall, significantly more respondents aged 50-74 participated via VI than in-person (VI: 72%, in-person: 58%, ϕ = .118, p < .001). (For more information, see Table B1 in Appendix B.)

Based on these findings, we conclude that respondents holding a university degree and younger participants are more likely to opt for video interviews instead of in-person interviews when VI is the primary mode. Comparing the Phi values, as measures of the magnitude of associations, we can conclude that the most notable differences between VI and in-person modes were observed in age profile (ELSA), followed by highest qualification. Differences by economic activity groups, and especially differences by sex, were less pronounced.

Analysis of Response to Consent and Cognitive Assessments Via VI (RQ3)

Consent rates. In the analysed surveys, respondents were asked for consent to link their survey data with (a) health records held by the UK[8] (NHS) and economic records held by (b) the Department for Work and Pensions (DWP) and (c) Her Majesty's Revenue and Customs data. Our third research question asks whether this kind of task could be conducted in VI as well as in person.

In Figure 4, we show the effect of mode on data linkage consent rates in the three surveys for which VI was the primary mode for a period, with average rates presented. The consent rates differ substantially between the studies but no differences in consent rates are identified between VI and in-person interviewing in NCDS (ϕ = .019, p = .554), BCS70 (ϕ = .039, p = .105) or ELSA (ϕ = .001, p = .920).

---

7 We did not receive data for ELSA on highest qualification.

8 In ELSA, respondents were asked for consent to link their survey data to the following NHS data sources: Mortality, Hospital Episode Statistics and Primary Care data. They were also asked to agree to wear an activity monitor.

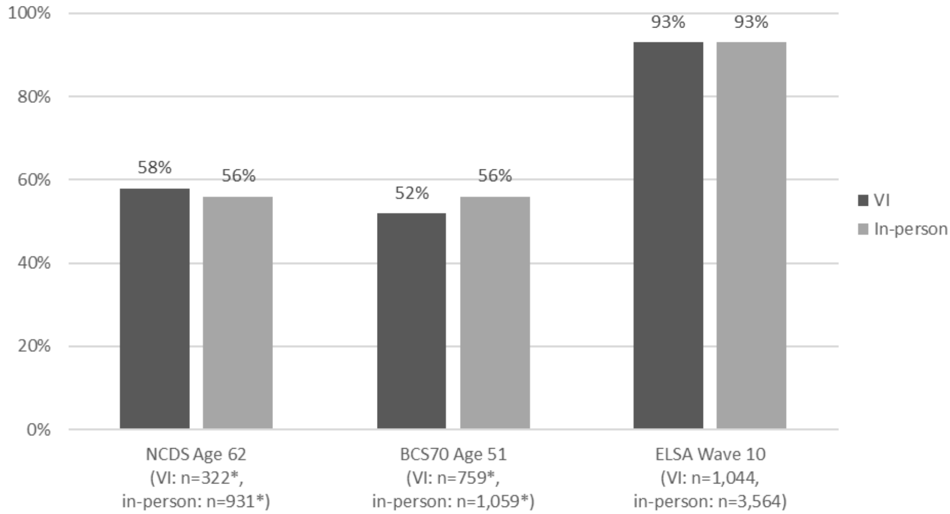


Figure 4 Average data linkage consent rates in NCDS, BCS70 and ELSA by survey mode

Notes: * In NCDS Age 62 and BCS70 Age 51 Surveys, only respondents who did not give consent in a previous wave were asked for consent to data linkage. The numbers of participants asked to give consent differed slightly between different data sources for linkage.

Cognitive assessments. RQ3 also asked about how well cognitive tests can be administered in VI. To address this, we examine completion rates for a series of cognitive tests in the three surveys. The following tests were included in all studies: (a) immediate word list recall test, (b) delayed word list recall test, (c) animal naming and (d) letter cancellation. The first three tasks were scored by the interviewer during the interview, so we combine their completion rate into a single measure – completion of "cognitive tasks". The letter cancellation task, on the other hand, was self-completed on a paper sheet during the interview and had to be returned to the study team for scoring. In VI interviews, respondents were required to return this by post, whereas in in-person interviews, the sheet was collected by the interviewer who was responsible for posting the sheet back to the study team for scoring.

The results, presented in Figure 5, compare completion and return rates for the cognitive tests between VI and in-person participants in NCDS, BCS70 and ELSA[9]. For the "cognitive tasks" scored by the interviewer during the interview, the completion rates were very high in all three surveys and with no significant differences between modes (with all $p > .05$). However, the return rates for the letter cancellation task, which required VI participants to post the paper sheet, were significantly lower in VI than in-person, in both NCDS (VI: 79%, in-person:

9 No letter cancellation test was included in ELSA.

88%, ϕ = .116, p < .001) and BCS70 (VI: 77%, in-person: 88%, ϕ = .146, p < .001). This no doubt reflects the additional burden associated with this extra task.

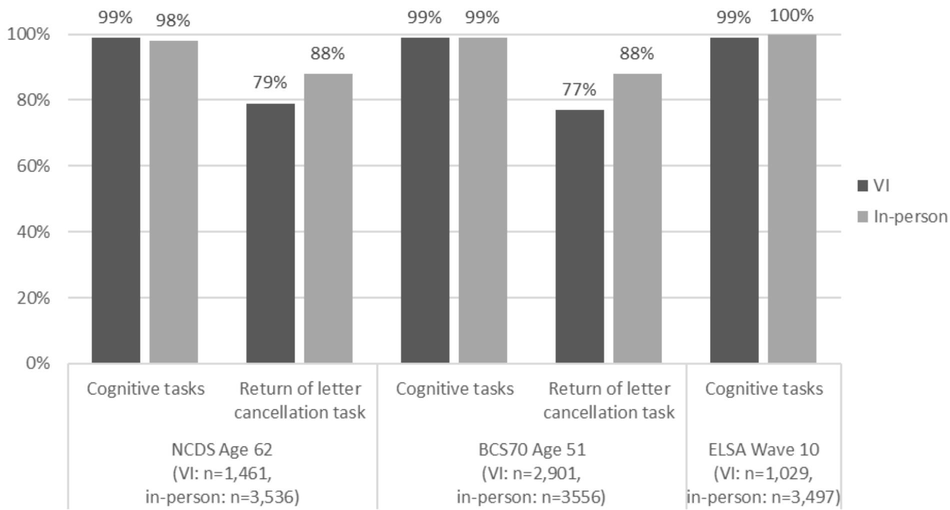


Figure 5 Return rates of cognitive test documents by mode.

Based on these findings, we conclude that respondents participating via VI are no less likely to complete questionnaire sections that are more complex, such as cognitive assessments than those participating in person (unless a paper document must be returned).

Discussion and Conclusions

This article reports evidence on conducting interviews via live, two-way video from the major UK social surveys that have implemented VI and presents key findings on response rates, sample composition, rates of consent to link administrative data and completion of cognitive assessments. We explore VI, either as a primary survey mode or as a supplement to traditional survey modes, such as in-person interviewing, telephone, and web. The main findings and conclusions are:

1. **VI designs:** The UK studies reviewed used VI in different ways and with different objectives. But they all used VI in a mixed mode design; in no case was VI used as the sole mode of data collection for the full duration of fieldwork. Some studies (i.e., NCDS, BCS70 and ELSA) used VI as the only (or first/primary) survey mode for a period when in-person data collection was not pos-

sible, i.e., in the initial stages of the pandemic. In those circumstances, VI acted as a substitute for in-person interviewing and allowed these studies to proceed with data collection and, based on the literature (Endres et al., 2023), it seems likely that this substitution did not affect measurement equivalence. However, these studies all reverted to in-person interviewing when pandemic restrictions were lifted, with VI continuing to be offered as an alternative mode. Some studies used VI to reach participants in areas where no in-person interviewers were available (e.g., Northern Ireland in ESS) and thus to increase geographical coverage. Some studies used VI as a response maximisation approach in mixed-mode designs; it was either given as an option to attract those who were reluctant to be interviewed in person (e.g., in ELSA) or not suggested until respondents refused to participate via the primary mode(s) (e.g., in Next Steps).

2. **VI response rates (uptake) (RQ1):** The first challenge is that response rates in both cross-sectional and longitudinal studies analysed here are generally lower for VI than other modes, consistent with the previous literature (Conrad et al., 2023; Guggenheim & Howell, 2021; Phillips et al., 2023). VI uptake was highest in studies which offered VI as the sole or primary mode (at least for a period of time, e.g., in certain parts of fieldwork). In NCDS and BCS70 VI response rates of 48% and 46% respectively were achieved using a video-only approach although in ELSA the VI response rate was lower at 25%. In all cases, overall response rates were boosted when non-respondents were later invited to take part in person.
   However, in NCDS, BCS70 and ELSA fieldwork was later conducted in person first (with VI as a back-up mode) and overall response rates were significantly higher in NCDS and ELSA, despite low VI response rates. The uptake of VI appears to decrease with age, as the participants in the ELSA Wave 10 cohort were, on average, almost 20 years older than those in the BCS70 cohort and about 8 years older than those in the NCDS cohort and their VI uptake was substantially lower (25%). This pattern can also be seen within ELSA: 50-74 year-olds were more likely to be interviewed by VI than in person but this reversed above age 74. The three studies, which had periods of fieldwork where VI was the sole or primary mode of data collection, were all long-established cohort studies. The lower response rates during VI-only part of fieldwork may in part be attributed to the fact that panel members were very familiar with in-person interviewing through repeated participation across multiple waves and so were reluctant to participate via VI, a phenomenon previously discussed for studies conducted in online panels (Conrad et al., 2023; Phillips et al., 2023).
   Another factor is that for some of the cohort studies here, participants were substantially older than the general population, and so may have been especially unfamiliar with and averse to using new technology.

VI uptake in the studies where VI was offered as one of a range of possible modes was generally low (ESS UK, Next Steps, Cots2020, HSE). From this evidence, it may be concluded that in mixed-mode studies, respondents may be more likely to choose an alternative mode to which they are more accustomed over VI. Uptake will also relate to what is most convenient in the moment, and if an interviewer is present at the doorstep, respondents may well take up that opportunity.

3. **Composition of VI samples (representation) (RQ2):** Another challenge is that the observed lower response rates in VI could lead to an increase in representation bias. Our results suggest that those who are more educated, employed, and to a lesser extent, women may generally be more likely to participate via VI, whereas long-term sick/disabled individuals and older people might be less likely. These findings are based on a comparison of the final composition of the in-person-administered and VI-administered subsamples, while considering that respondents could have been invited to take part in VI-first but responded in person (and vice versa). We conclude that in a longitudinal study context, switching from in-person interviewing to VI for the whole or a larger part of the sample could not only result in an increased loss of sample units but also in differential attrition, meaning that the proportion of already underrepresented and/or smaller subpopulations (e.g., those with chronic conditions) would decrease. This could lead to a limitation in studying certain groups due to an increased loss of statistical power over time associated with the newly introduced survey mode, VI. Another factor to consider is that people with certain characteristics, such as those with technical difficulties and inabilities, do not take part in VI to start with.

4. **Obtaining of consent and cognitive assessments (RQ3):** Promisingly, the analysis of data linkage consent rate and completion rates for cognitive assessments suggested that VI can be used effectively for collecting these kinds of measures. The evidence showed that completion rates/consent rates are generally comparable with those achieved in person. The only notable exception concerns the letter cancellation task which is self-completed on paper during the VI interview and is returned by post after the interview; completion and return rates for this task were about 10%-points higher via in-person interviewing than VI.

5. **Limitations:** While we investigated differences between VI and other modes, with a focus on comparing VI to in-person interviewing, none of the included surveys used experimental designs. Thus, we could not fully disentangle the effects of mode from the effects of the composition of samples that were self-selected or nonrandomly invited to take part via a particular mode. Therefore, most of our findings on the differences between VI and in-person

interviewing warrant further investigation in studies designed to test these differences.

Another important question that should be addressed in a future study is how VI performs if it is the only mode in a survey. Ideally, such a study would be informed by analyses of the cost savings from reduced travel on the one hand and higher levels of differential nonresponse on the other. In addition, the focus of the current article is on errors of representation not errors of measurement including those due to mode effects or interviewer effects, as we did not have access to the necessary data for such analyses. However, there is a small but growing literature on measurement in VI (see Asensio et al., 2025; Conrad et al., 2023; Endres et al., 2023; Kocar et al., 2025) but this is beyond the scope of the current paper. Lastly, VI has the potential to generate substantial cost-savings. In this paper, however, we are not able to address of cost-savings because, again, we did not have access to the necessary data.

Overall Feasibility of VI as a Mode for Conducting Large-Scale Social Surveys

Our study has demonstrated that VI has the potential to continue playing a role in large-scale social surveys, including in the UK. Whilst we investigated UK surveys here, we believe that (some of) the findings are generalisable to survey settings in other contexts. Based on the evidence presented in this study, we agree with the authors of similar studies (e.g., Conrad et al., 2023; Endres et al., 2023) that VI has the necessary characteristics to serve as an alternative to the in-person mode, especially with methodological optimisations and further technological developments. Our key generalisable conclusions are:

1. **Longitudinal studies have the potential to deploy VI most effectively.** This is particularly true for those that have typically collected data in person. Such studies have two notable advantages: (a) a previous relationship with participants, which fosters engagement and commitment, making them more likely to trial a new mode, and (b) access to contact details, such as emails and phone numbers, which can be used to make remote contact. In cross-sectional studies that require an in-person visit to arrange a video interview (e.g., ESS UK), VI uptake is expected to be low.

2. **VI appears to be more suitable for mixed-mode surveys.** At present, VI is unlikely to be suitable as a single-mode due to recruitment challenges, nonresponse, and indicators for potential representation bias identified in our study and others (e.g., Conrad et al., 2023; Guggenheim & Howell, 2021). In mixed-mode surveys, other modes can mitigate certain negative aspects of VI, while VI offers benefits such as response maximization, improved inclu-

sivity and facilitation of participation of certain sub-groups, and recruitment in areas where in-person interviewers are unavailable.

3. **VI can be cost- and time-efficient.** We agree with previous studies (e.g., Conrad et al., 2023) that highlight VI's potential for substantial cost savings compared to in-person interviews due to reduced travel time and expenses. Reduced travel also makes fieldwork more time-efficient, and so we recommend initially pushing sample members to VI before offering other modes in mixed-mode surveys for greater overall efficiency. However, an initial investment in VI technology and training is required, and cost savings can only be fully realised, if contact can be made remotely rather than through in-person visits.

4. **VI can be better suited for specific purposes than many other modes.** Besides cost- and time-efficiency compared to in-person interviewing, VI offers the most measurement equivalence with in-person interviewing compared to modes such as telephone and web (see Endres et al., 2023). It also appears to be suitable for including specific survey tasks or questions (e.g., cognitive assessments, consent), potentially more so than web-surveys into which it can be challenging to integrate complex tasks (Al Baghal, 2019; Emery et al., 2023; Ofstedal et al., 2021).

5. **Interviewer training and the resulting effects have notable implications.** In the analysed studies, VI was conducted by interviewers, who were previously used to conducting in-person interviews. We conclude that the challenges and cost implications of VI interviewer training should not be underestimated, consistent with the recommendations by Schober et al. (2020). Interviewer effects have not (yet) been explored for VI in these UK surveys.

6. **The future of VI looks promising.** VI has the potential to serve as a data collection tool for an increasingly broad range of surveys. This will become even more feasible as video-calling becomes more ubiquitous in our work and personal lives, especially for certain demographic profiles, and with future technological innovations. Nevertheless, evidence from this study suggests that VI is already a suitable alternative to more established survey modes in particular contexts.

Funding

This research was supported by UKRI-ESRC strategic research grant ES/X014150/1 for “Survey data collection methods collaboration: securing the future of social surveys”, known as Survey Futures. Survey Futures is directed by Professor Peter Lynn, University of Essex, and is a collaboration of twelve organisations, benefitting from additional support from the Office for National Statistics and the ESRC National Centre for Research Methods. Further information can be found at www.surveyfutures.net. The research reported here forms part of Research Strand 3 of Survey Futures, led by Matt Brown, University College London, which focuses on video interviewing.

References

Al Baghal, T. (2019). The effect of online and mixed-mode measurement of cognitive ability. *Social Science Computer Review,* 37(1), 89-103. https://doi.org/10.1177/0894439317746328

Al Baghal, T., Sloan, L., Jessop, C., Williams, M. L., & Burnap, P. (2020). Linking twitter and survey data: The impact of survey mode and demographics on consent rates across three UK studies. *Social Science Computer Review, 38*(5), 517-532. https://doi.org/10.1177/0894439319828011

Anderson, A. H. (2008). Video-mediated interactions and surveys. In F. G. Conrad & M. F. Schober (Eds.), *Envisioning the survey interview of the future* (pp. 95–118). Wiley.

Asensio M., Brown M., Silverwood R., Hanson T. & Durrant G. (2025) Assessing the impact of video interviewing on measurement quality: Evidence from an experimental study on mode effects, Survey Futures Working Paper no. 8. Colchester, UK: University of Essex. Available at https://surveyfutures.net/wp-content/uploads/2025/11/working-paper-08-video-interviews.pdf

Centre for Longitudinal Studies. (n.d.-a). *1958 National Child Development Study*. Retrieved March 27, 2024, from https://cls.ucl.ac.uk/cls-studies/1958-national-child-development-study/

Centre for Longitudinal Studies. (n.d.-b). *1970 British Cohort Study*. Retrieved March 27, 2024, from https://cls.ucl.ac.uk/cls-studies/1970-british-cohort-study/

Centre for Longitudinal Studies. (n.d.-c). *What is Next Steps?.* Retrieved March 27, 2024, from https://nextstepsstudy.org.uk/home/about/

Centre for Longitudinal Studies. (n.d.-d). *Children of the 2020s study*. Retrieved March 27, 2024, from https://cls.ucl.ac.uk/cls-studies/children-of-the-2020s-study/

Cernat, A. (2015). The impact of mixing modes on reliability in longitudinal studies. *Sociological Methods & Research, 44*(3), 427-457. https://doi.org/10.1177/004912411455380

Conrad, F. G., Schober, M. F., Hupp, A. L., West, B. T., Larsen, K. M., Ong, A. R., & Wang, T. (2023). Video in survey interviews: Effects on data quality and respondent experience. *Methods, Data, Analyses, 17*(2), 135-170. https://doi.org/10.12758/mda.2022.13

Conrad, F. G., Hupp, A. L., Hanson, T., & Durrant, G. B. (2026). Editorial introduction: Video survey interviewing special issue. *Methods, Data, Analyses, 20*(1), 6–20.

Deakin, H., & Wakefield, K. (2014). Skype interviewing: Reflections of two PhD researchers. *Qualitative research, 14*(5), 603-616. https://doi.org/10.1177/1468794113488126

Dillman, D. A., & Christian, L. M. (2005). Survey mode as a source of instability in responses across surveys. *Field Methods, 17*(1), 30-52. https://doi.org/10.1177/1525822X04269550

Dillman, D. A., Phelps, G., Tortora, R., Swift, K., Kohrell, J., Berck, J., & Messer, B. L. (2009). Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet. *Social Science Research, 38*(1), 1-18. https://doi.org/10.1016/j.ssresearch.2008.03.007

Dulaney, R., Kelley, J., Arrue, J., & Edwards, B. (2023, July 17-21). Who is Choosing Video Interviewing? A Look into Respondent Characteristics of Video Adopters [Conference presentation]. *10th Conference of the European Survey Research Association.* Milan, Italy.

Emery, T., Cabaco, S., Fadel, L., Lugtig, P., Toepoel, V., Schumann, A., Lück, D., & Bujard, M. (2023). Breakoffs in an hour-long, online survey. *Survey Practice, 16*(1). https://doi.org/10.29115/SP-2023-0008.

Endres, K., Hillygus, D. S., DeBell, M., & Iyengar, S. (2023). A randomized experiment evaluating survey mode effects for video interviewing. *Political Science Research and Methods, 11*(1), 144-159. https://doi.org/10.1017/psrm.2022.30

English Longitudinal Study of Ageing. (n.d.). *Study documentation.* Retrieved March 27, 2024, from https://www.elsa-project.ac.uk/study-documentation.

European Social Survey European Research Infrastructure (ESS ERIC) (2023) ESS10 – integrated file, edition 3.2 [Data set]. Sikt – Norwegian Agency for Shared Services in Education and Research. https://doi.org/10.21338/ess10e03_2

European Social Survey. (n.d.). *United Kingdom.* Retrieved March 27, 2024, from https://www.europeansocialsurvey.org/about/national-pages/united-kingdom/english

Fomby, P., Sastry, N., Johnson, H., Crane, S., & Connors-Burge, V. (2026). Video interviewing for standardized achievement assessments with children. *Methods, Data, Analyses, 20*(1), 87–107. https://doi.org/10.25521/mda.657

Gooch, A. (2015). Measurements of cognitive skill by survey mode: Marginal differences and scaling similarities. *Research & Politics, 2*(3). https://doi.org/10.1177/2053168015590681

Groves, R. M., Fowler, F. J., Jr., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2011). *Survey methodology* (2nd ed.). Wiley.

Guggenheim, L., & Howell, D. (2021, January 27). Live Video Interviewing in the 2020 ANES Time Series Study [Seminar presentation]. *Joint Program in Survey Methodology Seminar Series.* College Park, Maryland, United States.

Hanson, T., Briceno-Rosas, R., & Mitchell, J. (2023, July 17-21). Video interviewing on the European Social Survey [Conference presentation]. 10th Conference of the European Survey Research Association. Milan, Italy.

Hanson, T., Briceno-Rosas, R., Ainsaar, M., & Kartau, H. K. (2026). Live online video interviewing as a complementary mode to in-person interviews: Evidence from the European Social Survey. *Methods, Data, Analyses, 2*0(1), 179–208. https://doi.org/10.25521/mda.663

Höhne, J. K. (2024, January 10). How to incorporate AI interviewers in contemporary web surveys [Seminar presentation]. *Going online: using video interviewing in surveys research.* Organised by the *Royal Statistical Society,* London, Online.

Hupp, A. L., Conrad, F. G., Larsen, K. M., Schober, M. F., West, B. T., & Harrison, M. (2026). Dynamics of participating in video survey interviews. *Methods, Data, Analyses, 20*(1), 21–45. https://doi.org/10.25521/mda.635

Irani, E. (2019). The use of videoconferencing for qualitative interviewing: Opportunities, challenges, and considerations. *Clinical Nursing Research, 28*(1), 3-8. https://doi.org/10.1177/105477381880317

Jäckle, A., Beninger, K., Burton, J., & Couper, M. P. (2021). Understanding data linkage consent in longitudinal surveys. In P. Lynn (Ed.), *Advances in longitudinal survey methodology*. https://doi.org/10.1002/9781119376965.ch6

Jäckle, A., Burton, J., Couper, M. P., Crossley, T. F., & Walzenbach, S. (2022). How and why does the mode of data collection affect consent to data linkage?. *Survey Research Methods, 16*(3), 387–408. https://doi.org/10.18148/srm/2022.v16i3.7933.

Jäckle, A., Burton, J., Couper, M. P., Crossley, T. F., & Walzenbach, S. (2023). Survey consent to administrative data linkage: five experiments on wording and format. *Journal of Survey Statistics and Methodology, 12*(5), 1174–1199. https://doi.org/10.1093/jssam/smad019

Jeannis, M., Terry, T., Heman-Ackah, R., & Price, M. (2013). Video Interviewing: An exploration of the feasibility as a mode of survey application. *Survey Practice, 6*(1). https://doi.org/10.29115/SP-2013-0001

Kocar, S., Matt, B., Sanchez, C., Ploubidis, G.B. and Durrant, G. (2025) Measurement Mode Effects in Live Video Interviewing: Evidence from a Mixed-Mode Quasi-Experimental Survey, Survey Futures Working Paper, Colchester, UK: University of Essex. Forthcoming. Shortly available from https://surveyfutures.net/working-papers/.

Kunz, T., Daikeler, J., & Ackermann-Piek, D. (2024). Interviewer-observed paradata in mixed-mode and innovative data collection. *International Journal of Market Research, 66*(1), 14-26. https://doi.org/10.1177/14707853231184742

Martin, L., & Fradier, Y. (2023, July 17-21). ESS + CAVIsio [Conference presentation]. *10th Conference of the European Survey Research Association.* Milan, Italy.

National Health Service. (n.d.). *Health Survey for England*. Retrieved March 27, 2024, from https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england

Office for National Statistics. (2021, April 6). *Internet users, UK: 2020*. ONS. https://www.ons.gov.uk/businessindustryandtrade/itandinternetindustry/bulletins/internetusers/2020.

Ofstedal, M. B., McClain, C. A., & Couper, M. P. (2021). Measuring cognition in a multi-mode context. In P. Lynn (Ed.), Advances in longitudinal survey methodology (pp. 250–271). https://doi.org/10.1002/9781119376965.ch11

Olson, K., Smyth, J. D., Horwitz, R., Keeter, S., Lesser, V., Marken, S., Mathiowetz, N. A., McCarthy, J. S., O'Brien, E., Opsomer, J. D., Steiger, D., Sterrett, D., Su, J., Suzer-Gurtekin, Z. T., Turakhia, C., & Wagner, J. (2021). Transitions from telephone surveys to self-administered and mixed-mode surveys: AAPOR task force report. *Journal of Survey Statistics and Methodology, 9*(3), 381-411. https://doi.org/10.1093/jssam/smz062

Peycheva, D. N., Sakshaug, J. W., & Calderwood, L. (2021). Occupation coding during the interview in a web-first sequential mixed-mode survey. *Journal of Official Statistics, 37*(4), 981-1007. https://doi.org/10.2478/jos-2021-0042

Phillips, B., Neiger, D., Slamowicz, S., Lester, G., Luddon, S., Farrell, E., Gerlach, K., & Carmo, P. (2023, July 17-21). Video-Assisted Live Interviewing in Comparison to Other Survey Methods in Australia [Conference presentation]. *10th Conference of the European Survey Research Association.* Milan, Italy.

Sakshaug, J. W., Couper, M. P., Ofstedal, M. B., & Weir, D. R. (2012). Linking survey and administrative records: Mechanisms of consent. *Sociological Methods & Research, 41*(4), 535-569. https://doi.org/10.1177/0049124112460381

Sakshaug, J. W., Hülle, S., Schmucker, A., & Liebig, S. (2017). Exploring the effects of interviewer-and self-administered survey modes on record linkage consent rates and bias. *Survey Research Methods, 11(*2), 171-188. https://doi.org/10.18148/srm/2017.v11i2.7158

Schober, M. F. (2018). The future of face-to-face interviewing. *Quality Assurance in Education*, *26*(2), 290-302. https://doi.org/10.1108/QAE-06-2017-0033

Schober, M. F., Conrad, F. G., Hupp, A. L., Larsen, K. M., Ong, A. R., & West, B. T. (2020). Design considerations for live video survey interviews. *Survey Practice, 13*(1). https://doi.org/10.29115/SP-2020-0014

Schober, M. F., Okon, S., Conrad, F. G., Hupp, A. L., Ong, A. R., & Larsen, K. M. (2023). Predictors of willingness to participate in survey interviews conducted by live video. *Technology, Mind, and Behavior, 4*(2), 215–229. https://doi.org/10.1037/tmb0000100

Selvam, R., Hu, R., Musselman, R., Raiche, I., & Moloo, H. (2020). Video-based interviewing in medicine: protocol for a scoping review. *Systematic Reviews*, *11*, 94. https://doi.org/10.1186/s13643-022-01959-8

Silber, H., Ell, T., Repke, L., & Cernat, A. (2024, January 10). Collecting egocentric social network data in live video-based interviews [Seminar presentation]. *Going online: using video interviewing in surveys research*, Online.

Sun, H., Conrad, F. G., & Kreuter, F. (2021). The relationship between interviewer-respondent rapport and data quality. *Journal of Survey Statistics and Methodology, 9*(3), 429-448. https://doi.org/10.1093/jssam/smz043

Thornby, M., Calderwood, L., Kotecha, M., Beninger, K., & Gaia, A. (2018). Collecting Multiple Data Linkage Consents in a Mixed-mode Survey: Evidence from a large-scale longitudinal study in the UK. *Survey methods: Insights from the field*, 1-14. https://doi.org/10.13094/SMIF-2018-00004

Thorolfsson, A., Hjaltason, A. B., Jonsdottir, G. A., & Gudmundsson, H. (2023, July 17-21). Mode effects of video interviewing as a proxy of face-to-face interviewing in R10 of ESS in Iceland [Conference presentation]. *10th Conference of the European Survey Research Association.* Milan, Italy.

Tsakos, G., Bernabé, E., O'Brien, K., Sheiham, A., & de Oliveira, C. (2008). Comparison of the self-administered and interviewer-administered modes of the child-OIDP. *Health and Quality of Life Outcomes*, *6*(1), 1-8. https://doi.org/10.1186/1477-7525-6-40

Van Zeeland, I., Van den Broeck, W., Boonen, M., & Tintel, S. (2021). Effects of digital mediation and familiarity in online video interviews between peers. *Methodological Innovations, 14*(3). https://doi.org/10.1177/20597991211060743

West, B. T., Ong, A. R., Conrad, F. G., Schober, M. F., Larsen, K. M., & Hupp, A. L. (2022). Interviewer effects in live video and prerecorded video interviewing. *Journal of Survey Statistics and Methodology, 10*(2), 317-336. https://doi.org/10.1093/jssam/smab040

Zager Kocjan, G., Lavtar, D., & Sočan, G. (2023). The effects of survey mode on self-reported psychological functioning: Measurement invariance and latent mean comparison across face-to-face and web modes. *Behavior Research Methods, 55*(3), 1226-1243. https://doi.org/10.3758/s13428-022-01867-8

Appendix A

Mixed Mode Data Collection

Table A1 Mixed-mode data collection in NCDS Age 62 Survey, BCS70 Age 51 Survey and ELSA Wave 10

Study	Stage[a]	Data collection period	Assigned to VI	Assigned to in-person
NCDS Age 62 Survey	1. Video-only pilot	Spring 2021	*N* = 311	*N* = 0
	2. Video-first main stage waves	October 2021 – summer 2022	*N* = 1,777[b]	*N* = 0
	3. Predominantly in-person-first	Spring 2022 – November 2023	*N* = 1,002[c]	*N* = 4,559
BCS70 Age 51 Survey	1. Video-only pilot	Autumn 2020	*N* = 60	*N* = 0
	2. Video-first soft launch	Summer 2021	*N* = 1,067	*N* = 0
	3. Video-first main stage waves	October 2021 – early 2022	*N* = 5,269[b]	*N* = 0
	4. Predominantly in-person-first	Spring 2022 – November 2023	*N* = 894[c]	*N* = 4,661
ELSA Wave 10	1. Video-only pilot	May-June 2021	*N* = 317	*N* = 0
	2. Video-first main stage waves	October 2021 – early 2022	*N* = 2,881	*N* = 1,115
	3. Predominantly in-person-first	February 2022 – March 2023	*N* = 3,523	*N* = 426

Notes:
[a] the evidence presented in this article is based on all NCDS stages, BCS70 stages 3–4 and ELSA stages 2-3.
[b] This excludes cases who were originally assigned to in-person interviewing before COVID and then assigned to VI later. It includes 192 (NCDS) and 419 (BCS70) cases that were deemed ineligible for VI due to the absence of a telephone number or email address and were assigned to an in-person interviewer later.
[c] They were due to be assigned to an in-person interviewer first but had to be assigned to VI-first as there was no interviewer available.

Additional Information about the Specifics of VI Fieldwork (VI-First Mode of Invitation)

In addition to the stages described in Table A1, the NCDS fieldwork operation in these periods also included cases that had been assigned to in-person interviewers before the pandemic struck but where an interview had not been achieved by the time fieldwork was paused.

Further, there were cases that were initially earmarked for in-person visits following the restart of this approach, but that were switched to VI due to a lack of available field interviewer resource in specific geographical locations, which was also the case for BCS70 and ELSA. This reflected a cross-sector fieldwork capacity problem at this time following the loss of interviewers from agencies' panels and a challenging labour market for recruitment. A strength of the VI approach is that geographical restrictions do not apply.

The analysis in this study focuses on comparing the VI-first approach with the in-person approach. To maintain comparability and provide a full understanding of non-response to VI, the VI-first groups in NCDS and BCS waves include a group of cases that were part of the sample earmarked for allocation to VI that could not be worked in that mode due to an absence of telephone numbers and email addresses.

Additional Information about the HSE Pilot

The sample frame for the HSE pilot contained no names or other contact information for selected households. In the first stage in January 2021, those who opted in and provided contact information were contacted by interviewers who gave the option of a VI or a telephone interview. Of the initial sample of households ($N = 828$), 24% opted in and interviews were achieved by one or other mode in 76% of these households. Of the individuals interviewed ($N = 237$), 13% were conducted via video with the remainder by telephone. With the February and March samples, those opting in were asked to indicate whether they preferred a telephone or a video interview when opting in via the portal. Of the opt-in sample in these months ($N = 561$) a smaller proportion opted for the video route with this approach (4%, $N = 21$ interviews).

Appendix B

Descriptive Statistics

Table B1 Age distribution of ELSA participants by survey mode.

Age group	VI (N = 1,055)	In-person (N = 3,609)
50–54	3.1%	1.7%
55–59	13.4%	10.8%
60–64	8.5%	7.4%
66–69	22.7%	17.5%
70–74	24.0%	20.6%
75–79	16.2%	19.0%
80–84	8.4%	11.9%
85–89	2.8%	8.4%
90+	0.7%	2.6%
Missing	0.3%	0.1%
Total	100.0%	100.0%

Table B2 Economic activity distribution by survey mode in NCDS, BCS and ELSA (Economic activity is a measure of whether or not a person is an active participant in the labour market, with "economic active" being employed or self-employed, and "economic inactive" being unemployed, i.e. currently without employment, permanently sick or disabled, looking after home or family, or retired.).

Economic activity	NCDS		BCS70		ELSA	
	VI (N = 1,453)	In-person (N = 3,487)	VI (N = 2,794)	In-person (N = 2,964)	VI (N = 1,055)	In-person (N = 3,609)
Employed	38.9%	36.4%	73.6%	68.6%	18.1%	14.4%
Self-employed	15.3%	12.3%	16.7%	16.1%	5.7%	4.7%
Unemployed	1.2%	1.8%	0.9%	1.6%	0.7%	0.6%
Permanently sick or disabled	4.5%	8.4%	2.2%	5.4%	1.4%	2.5%
Looking after home or family	7.4%	6.9%	4.4%	4.8%	2.8%	1.8%
Retired	30.5%	31.9%	1.0%	1.1%	69.8%	74.8%
Other	2.1%	2.3%	1.2%	2.5%	1.5%	1.2%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

methods, data, analyses | Vol. 20(1), 2026, pp. 269-301 DOI: 10.25521/mda.674

Investigating Live Video Interviewing in the Life in Australia™ Panel Survey: An Experimental Study

Carina Cornesse[1], Benjamin Phillips[2], Dina Neiger[2], Emma Farrell[3] & Philip Carmo[3]

[1] *GESIS – Leibniz Institute for the Social Sciences*
[2] *The Social Research Centre (SRC)*
[3] *Australian Bureau of Statistics*

Abstract

Video calls have become common in many people's lives, especially since the COVID-19 pandemic. Interviewing respondents via video calls may be a useful addition to the survey mode toolbox as it may increase response rates as well as data quality for at least some population subgroups. To assess the potential of live video interviews (VI) in Australia, a study was implemented, examining VI in a probability-based online panel in Australia in 2022, using an experimental design. The focus of the assessment presented in this paper is on (a) factors that influence VI uptake (as measured by general willingness, consent, appointment-making and response to VI), and (b) factors that influence the data quality of the resulting survey responses (as measured by satisficing and social desirability indicators), comparing the VI response group to a group using the panel's regular online self-completion design. We find that panel members who are highly educated, frequently use the internet, and have high prior panel wave completion rates are the most likely to try VI. Older panel members are more willing to try VI than younger ones, but do not actually participate at higher rates. Regarding data quality of responses, VI respondents produce more socially desirable responses compared to the online self-completion respondents while no differences in terms of satisficing behavior are detected.

Keywords: video interviewing, data quality, panel surveys, social desirability, satisficing

© The Author(s) 2026. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 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 the past, face-to-face interviewing was considered the gold-standard among survey modes and was widely used as the primary mode for data collection in most major survey programs. However, over the years, response rates have steadily decreased in this mode, while its costs have risen (de Leeuw et al., 2018; Smith, 1995; Williams & Brick, 2018). In today's digital era, there is a growing preference for more flexible survey modes which align with our modern lives (Schober, 2018; Smyth et al., 2014). This shift is particularly evident among certain socio-demographic subgroups, such as younger adults and highly educated people (Haan, 2015). Due to growing cost restrictions and difficulties with reaching people at home, many major survey programs are increasingly incorporating online surveys as a self-completion option for some or all their participants (Burton et al., 2020; Cernat & Revilla, 2021; Maslovskaya & Lugtig, 2022).

Online self-completion surveys promise reduced costs, faster data collection, and low social desirability bias (Bianchi et al., 2017; Kreuter et al., 2008). However, they may also come with notably higher nonresponse rates and therefore likely higher nonresponse bias, particularly among older individuals or those with lower educational attainment (e.g., Cornesse & Schaurer, 2021), ignoring the offline population altogether (Bach et al., 2024), and data quality issues, such as break-offs and speeding (e.g., Heerwegh & Loosveldt, 2008; Revilla & Ochoa, 2015; Zhang & Conrad, 2014). This is especially true in comparison to face-to-face interviews, where respondents can see, hear, and interact with interviewers (Al Baghal & Lynn, 2015; Daikeler et al., 2020; Peytchev, 2009). Live video interviewing, where respondents and interviewers are connected via video conferencing software, may offer an adequate compromise between the different types of survey modes: convenient and more cost-efficient than face-to-face interviewing, but with personal real-time interviewer interaction and assistance (Weller, 2017) and potentially higher data quality (comparable to face-to-face). In the following, we will refer to live video interviewing as just "video interviewing" (abbreviated VI)[1]. At present systematic assessments of VI are still scarce, in particular in the context of panel studies. Also, most studies use observational and not experimental data that systematically assess the implementation of VI. We address these gaps here.

The study presented here examines VI using an experimental design in a probability-based online panel in Australia. We investigate (a) how factors including characteristics of respondents, access to the internet, and previous participation history in an online panel impact uptake of VI (as measured by general

1 VI can also include the use of (pre-)recorded videos (e.g., Fuchs, 2009; Haan et al., 2017; Höhne et al., 2024), but this is outside the focus of this study.

Direct correspondence to
Carina Cornesse, GESIS – Leibniz Institute for the Social Sciences, B6 4-5, 68159 Mannheim, Germany
E-mail: Carina.Cornesse@gesis.org

willingness, consent, appointment-making and response to VI). Furthermore, we investigate (b) factors that influence the data quality of survey responses (as measured by satisficing and social desirability indicators), comparing the VI response group to a group using the panel's regular online self-completion design. Considerations and design choices around the practical implementation of VI are also briefly presented.

The remainder of the paper is structured as follows: The Background section presents relevant context on VI, followed by a section outlining our expectations and hypotheses regarding effects of factors and characteristics that may influence uptake of VI (here in an online panel setting). We then present our data and analytical strategy to assess VI and the details on the VI implementation in the Data and Methods section. Subsequently, we discuss the main results from our modelling in the Results section and conclude with a summary of findings and discussion of limitations and future research avenues.

Background

In this section, we first discuss the potential that VI may have as a new survey mode based on the evidence of video call use in the general population and then the potential benefits that VI may offer.

Trends in Uptake of Video Calling

VI may be particularly attractive as a new survey mode because an increasing number of individuals use video calls regularly and confidently in their professional and social lives. For example, in the United States, 81% of adults utilized video calls during the pandemic, with 12% using them multiple times a day (McClain et al., 2021). In Australia, official statistics indicate a significant rise in the use of app-based video calls during and after the pandemic: While in 2019, only 41% of adults had engaged in app-based video calling, this figure increased to 64% in 2023 (Australian Communications and Media Authority [ACMA], 2023). The potential for even more people to do video calls in the futures is high, since 93% of Australian households already have a home internet connection, 99% of the Australian population generally use the internet, and 92% use the internet multiple times a day (ACMA, 2023; ACMA, 2024).

As people have become more accustomed to video calls, they may also increasingly find it convenient and even enjoyable to participate in surveys through this mode (Schober et al., 2023). However, willingness and enjoyment of VI participation will likely differ across population subgroups depending on their internet access, digital skills, and use behavior. Official statistics on video call use from various countries convey that people who live in urban areas, are younger, female, and more highly educated are more likely to have (regular) experiences with video calls than people who live in more remote areas, are older, male, and less highly educated (ACMA, 2023; Eurostat, 2022; Thomas et al., 2023). In Australia, 74% of 18– to 24–year-olds had used app-based video calls within 6 months in 2023, but only 31% of people aged 75 and older. Furthermore, a higher share of women (69%) than men (58%) used video calls (ACMA, 2023). While official statistics on the interaction with education are unavailable for Australia, data from other countries suggests that such a correlation exists, with more highly educated people using video calls more often than people with low educational attainment (e.g., Germany: International Standard Classification of Education (ISCED) 1 and 2 (primary and lower secondary education): 72% video call users vs. ISCED 5 to 8 (tertiary education): 83%, see Destatis, 2024).

Benefits and Uptake of VI

VI offers several benefits. Compared to online self-completion surveys, for example, recent studies indicate lower cognitive response burden and higher data quality in VI (e.g., Conrad et al., 2023; Endres et al., 2023). Moreover, undesirable interviewer effects appear to be less pronounced in video than in offline in-person interviewing (West et al., 2022). VI may thus strike an optimal balance in terms of interviewer social presence: sufficiently close to provide assistance but not overly intrusive into the respondent’s private space (Schober, 2018). Additionally, the ability for both respondents and interviewers to share screens, upload documents, or exchange text messages in the chat enhances the flexibility of this survey mode. However, the extent to which VI can be a successful alternative to other modes depends on its implementation, including the choice of video call platform, interview scheduling tools, interviewer hardware, and setup (camera, background, attire, screen configuration), and interviewer training (e.g., Engward et al., 2022; Jeannis et al., 2013; Oates et al., 2022; see Schober et al., 2020 for a comprehensive overview and recommendations). In a panel survey context, the success of VI may also depend on the survey modes respondents are used to and whether they perceive VI as an improvement over the status quo.

Several studies have shown that younger and more highly educated individuals are more likely to be video interviewed than others (Martin & Fradier, 2023; Thorolfsson et al., 2023; Dulaney et al. 2023). Evidence regarding gender bias is inconclusive. Dulaney et al. (2023) and Kocar et al. (2025) report no gender differences in video participation, whereas Thorolfsson and colleagues (2023) observe that women are more likely to participate via video than men. In contrast, Fradier and Martin (2023) find higher VI uptake among men than women. Schober and colleagues (2023) suggest that socio-demographic characteristics may be completely unrelated to survey respondents’ willingness to participate in VI.

Expectations of VI Uptake and Data Quality Effects

Our expectations regarding which subgroups are more likely to take up VI and which circumstances may lead to a higher data quality in our study are informed by considerations relating to the concept of digital inequality. The concept evolved as an alternative to theories on the so-called digital divide and claims that nowadays, digital inclusion should not be viewed as a binary variable anymore, where people either do or do not use the internet. Instead, it postulates a continuum of digital inclusion, where most people use the internet, but they systematically differ in how easy it is for them to do so as well as their use preferences. Digital inequality consists of the infrastructure component of digital access and the individual-level components of digital skills and use behavior (see Helsper, 2021 for a detailed theory of digital inequality). It can generally be expected that people who benefit from good digital access (i.e., high speed internet connections and high quality devices) and who use the internet confidently for a variety of tasks (e.g., taking online surveys but also shopping, banking, social media; e.g., Jäckle et al., 2024) will be more willing and able to try VI than people in a less digitally privileged situation. Based on the concept of digital inequality, as well as the empirical evidence on uptake of video calling and VI (see Background section), we develop seven hypotheses around the characteristics of those that take up VI. We expect panel respondents used to online self-completion surveys to be more likely to try VI if …

… they live in urban rather than remote areas (H1), because of the better high-speed internet access.

… they are younger rather than older (H2) and/or more highly educated rather than not (H3), because they may have more confidence in their digital skills.

… they are female rather than male (H4), because video interviews fit well with women's higher video call use as already evidenced in official video call statistics.

… they are more frequent internet users (H5) and use the internet for other less traditional communication tasks such as social media (H6), because they should be more confident in their digital skills.

… they have high prior panel survey wave completion rates, indicating a high willingness to take part in a survey and suggesting low barriers to doing so (H7).

These hypotheses are all in line with the concept of digital inequality as well as official statistics on video call use (ACMA, 2023). In addition, H2 and H3 are also in line with some studies on VI (see Martin and Fradier, 2023; Thorolfsson et al., 2023; Dulaney et al. 2023) but not others (Schober et al. 2023). Furthermore, while H4 is in line with research by Martin and Fradier (2023) and Thorolfsson and colleagues (2023), it is not supported by Dulaney and colleagues (2023) or Schober and colleagues (2023). Regarding H7, no study to our knowledge has yet examined the association between online panel members' participation history and their likelihood to engage in video interviews. Completing online panel survey waves is a digital task like many others and requires digital access and skills. We therefore also expect this to be positively associated with uptake of VI. We expect that the seven hypotheses apply to the different phases of uptake which include general willingness to participate in VI, consent to actually do so, making an appointment for VI, and completion of the actual VI.

Apart from uptake to VI, it is also important to assess whether data quality of the resulting responses differs between this survey mode and the regular Life in Australia™ protocol, which requires online self-completion. It is reasonable to expect that common findings on comparing interviewer-administered to self-administered survey modes regarding cognitive burden, data quality, and interviewer effects apply to our setting (e.g., West & Blom 2017; Kreuter et al. 2008). This is especially true since the scarce existing literature on VI data quality outlined in the sections above suggests that such differences between modes exist. We, therefore, expect that the presence of the interviewer in the VIs will reduce satisficing behavior (H8) but increase socially desirable responding (H9) as compared to online self-completion.

Data and Methods

Life in Australia™ and the Australian Comparative Study of Survey Methods (ACSSM)

The study we report on in this manuscript was conducted in the Life in Australia™ panel, which is Australia's only probability-based online panel (Kaczmirek et al., 2019). Life in Australia™ was initially recruited in 2016 using telephone interviewing and dual-frame random digit dialing (RDD; n=3,322), topped up in 2018 using mobile-only RDD (n=267), expanded in 2019 using address-based sampling (ABS; n=1,810), topped up in 2020 using a combination of ABS, interactive voice response calls to mobile RDD sample, and SMS push-to-web to mobile RDD sample (n=612), topped up again in 2021 using SMS push-to-web to mobile RDD sample (n=510) in a trial of retention initiatives, and expanded in 2021 using ABS (n=3,715). Panel members unwilling or unable to participate in the panel online are interviewed via telephone. Following Callegaro and DiSogra's (2008) definitions, the weighted average Recruitment Rate for the study sample was 11.0% and the Profile Rate was 93.6%.

All panel members receive incentives of 10 AUD for completion of surveys of typically 20 minutes or less. The survey content consists of social and health research. Examples include surveys of Australian opinion on international affairs, social cohesion, disability attitudes, and topics in women's and sexual health. Surveys were released monthly at the time of the ACSSM, although not all panel members are invited to complete surveys each month and some are invited to complete multiple surveys in a given month. At the time of sample selection for the ACSSM in 2022, there were 7,265 active panel members. Over the period from October 2021 to November 2022, panel members were invited to an average of 12 monthly waves and completed an average of 9 surveys.

The experiment we report on in this paper is part of the ACSSM (Pennay et al., 2023; Phillips et al. 2023), which had the goal to evaluate and improve contemporary and emerging practices for general population surveys building on earlier work conducted in Life in Australia™ and beyond (Cornesse et al., 2020; Dutwin & Lavrakas, 2016; Lavrakas et al., 2022; Pennay et al., 2018; Pew Research Center, 2019). For the full experimental design of the ACSSM see its technical report composed by Pennay et al. (2023). In the ACSSM, surveys with questionnaires as similar as possible were fielded in different designs (e.g., different modes and sample selection procedures). Our paper focuses on the comparison of VI versus online self-completion implemented experimentally in Life in Australia™.

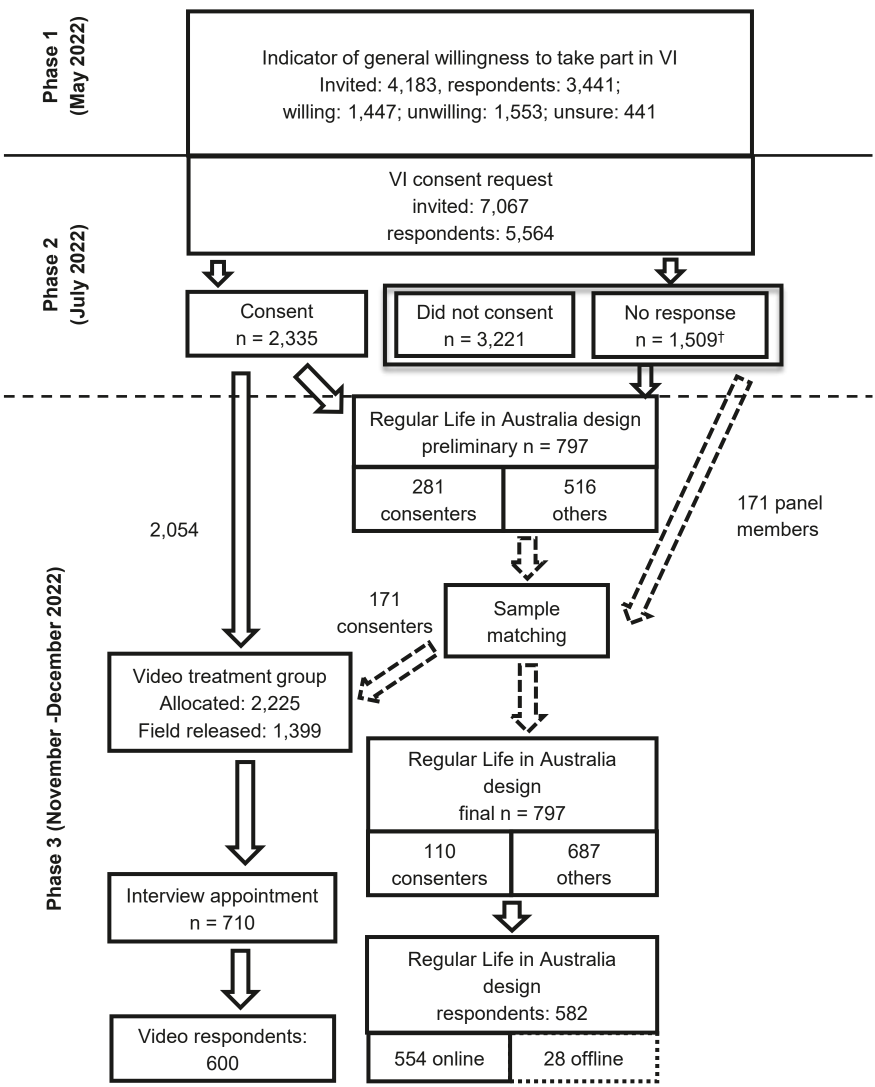
VI Study Setup and Sample Allocation

Our VI study consists of three phases, with the first one exploring general willingness to take part in VI, phase 2 asking about consent to take part in VI, and phase 3 VI implementation (treatment) and parallel data collection phases. Figure 1 schematically depicts key outcomes and sample allocation at each phase. At phase 1, we explored the general potential of VI in our panel by asking for general willingness to participate in the new mode. At phase 2, we asked a subset of panel participants to consent to participate in a VI study, which we then conducted at phase 3. In the following, we describe each study phase in more detail. Key variables of interest that are later used as dependent variables in our models are indicated as DV.

Study Phase 1

To understand the potential of offering VI (study phase 1), we implemented a question on general willingness to be video interviewed as part of the panel's May 2022 data collection wave, which was introduced for a random subsample of panel participants. Offliners (i.e., panel participants unwilling or unable to participate in the panel online) were excluded from this question, the actual VI treatment later that year, and hence from our analyses. Of the 4,183 invited panel members, 3,441 responded to this VI willingness indicator. In total, 1,447

(42%) were in principle willing to participate, 1,553 (45%) were unwilling, and 441 (13%) were unsure.



Notes: † Includes 3 partial respondents who did not reach the consent question.

Figure 1 Schematic depiction of the Life in Australia™ VI study in 2022 (numbers refer to observed sample cases at each phase

Study Phase 2

The results from the willingness indicator variable were sufficiently encouraging to subsequently trial VI in Life in Australia™. To be able to do so, all online respondents of the July 2022 wave were asked for VI consent (study phase 2). Of the 7,067 online Life in Australia™ members, 1,506 (21%) did not complete the July 2022 survey and 3 (< 1%) were partial interviews that did not reach the consent question. Of the remaining respondents, 2,335 (42%) provided consent, 3,221 (58%) did not consent.

Study Phase 3

Study phase 3 started with dividing the panel sample into two groups to be surveyed between November and December 2022. The first is a comparison group, which consists of a random draw of 797 Life in Australia™ active panel members stratified by age, education, and gender. This group was supposed to be surveyed using the regular Life in Australia™ study design, which consists predominantly of online self-completion questionnaires while non-internet users are interviewed via telephone.

By chance, however, a relatively high number, 281 people (35%), in the comparison group were VI consenters. Since it was unclear how many cases would have to be released into field to be able to meet the target number of 600 VIs, keeping 281 consenters from the VI group was deemed too risky. Therefore, the decision was taken to substitute a certain number of these cases from the randomly drawn comparison group. Through sample matching based on a VI consent propensity model, 171 (61% of consenters in the comparison group) were substituted by other active panel members who had not consented (see Table 1). This procedure maximized the pool of consenters that could be invited to VI but also ensured that both groups still contained VI consenters for comparative analyses.

To ensure that not all VI consenters would be removed from the comparison group, 600 cases were held out of the sample replacement but were still used in the propensity modeling.

Propensity modeling variables in the sample matching procedure included location (state × capital city/rest of state), area-level socio-economic status quintile, remoteness, gender, citizenship status, using a language other than English at home, Indigenous status, country of birth groupings, number of adults in the household, household type, household tenure, marital status, employment status, income, education, television viewing, digital affinity, Big Five personality indicators, self-assessed health, life satisfaction, alcohol and tobacco consumption, internal migration, concession card status, and telephony status. Characteristics that were highly correlated were excluded from the propensity model to avoid multicollinearity. There were a total of 480 combinations of propensity classes and selection strata (40 classes × 12 strata). Matched panel members were defined as being in the same propensity class and selection stratum combination as the selected panel members. Each panel member who consented to VI and was included in the regular Life in Australia™ design was then replaced with one of the matched panel members who had not consented to VI. When there were multiple matched panel members available, one was selected at random. Robustness analyses comparing substitute cases with their substitutes show no differences with the exception that the cases brought in from the remaining panel sample into the comparison group to replace 171 VI consenters had significantly lower prior survey completion rates than VI consenters (see the additional model Add. M1 in Appendix B).

Table 1 Selection into comparison group before and after sample matching case substitution

	Selected	Not selected	Total
Before substitution:			
VI consenters	281	2,054	2,335
Remaining panel members[a]	516	4,414	4,930
Total	797	6,468	7,265
After substitution:			
VI consenters	110	2,225	2,335
Remaining panel members[a]	687	4,243	4,930
Total	797	6,468	7,265

Notes: [a]Remaining panel members include non-consenters, nonrespondents to the VI request, and offliners not asked for their consent

In the end, 797 panel members were invited to the survey in their accustomed mode design, with 110 of them (14%) being VI consenters. The invitation was sent via email, followed by three reminders and a mop-up call the week before fieldwork ended. In total, 582 people completed the survey, with 554 of them participating online and 28 via telephone. The latter are excluded from our analyses to focus on the comparison between online self-completion versus VI.

The second group in this study is the VI treatment group. It consists of all panel members who consented to VI in study phase 2 except for the 110 cases randomly drawn into the regular Life in Australia™ design comparison group and not replaced in the sample matching procedure. These 2,225 cases in total were released into field in batches of approximately 200 cases until the target of 600 completed video interviews was reached. In each batch, invitations to make an appointment for the video interview were sent via email, followed by three reminders and a mop-up call the week before fieldwork ended. Only 1,399 cases (63%) had to be released into field to achieve the target number of video interviews, meaning that our expectations about how many panel members would participate in the VI were exceeded. Of those cases released into field, 128 had initially been drawn into the comparison group but then re-assigned to the VI group as a result of the sample matching procedure. Of all cases released into field, 710 people made a VI appointment and 600 completed the interview via video.

VI Implementation

We describe briefly the VI implementation, guided by prior work conducted by Schober et al. (2020) and Hanson (2021). First, we selected our treatment sample as described above and sent out invitations and reminders asking participants to schedule an appointment. To support panel members for whom VI was likely to be a new concept, a microsite was created on the Social Research Centre (SRC) website to explain what was being asked of them (see Supplementary Document 1 for all VI communication).

Appointment-Setting

Scheduling an interview appointment was necessary because maintaining a group of interviewers available to complete interviews without appointments (i.e., on-demand) was deemed infeasible in terms of costs and logistics due to lack of prior experience with VI in Life in Australia™. Study participants set and, as necessary, canceled and rescheduled appointments using the commercial booking website OnceHub. The OnceHub site look and feel was customized for Life in Australia™. Confirmation emails (and optionally SMS) were sent for appointments, cancelations and reschedules. Reminders were sent 24 hours, 1 hour and 10 minutes before the scheduled time of the appointment.

Conducting VIs

Interviews were conducted on MS Teams. A typical interviewing set-up consisted of a desktop computer with keyboard, mouse, and two screens. On the first screen, the interviewer would see the interview software (UNICOM Intelligence computer-assisted interviewing interface) with the survey questionnaire as well as the respondent video window on MS Teams. On the second screen, the interviewer would see the showcards, which they shared with the respondent via MS Teams. Interviewers were instructed to have the survey questionnaire on the screen where the camera was facing, so that the interviewer was looking in the respondent's direction when questions were being asked. A neutral, standardized background was used via the MS Teams' background function.

The interviewers were selected from the SRC established interviewing staff, who usually conducted telephone interviews for Life in Australia™. Altogether, six interviewers were trained. VI interviewer training included an overview of VI, how appointments were made, how to access appointments, computer set-up and use of software (calendar, showcards, CATI software), being present on Teams two minutes before the scheduled interview start, interviewer etiquette (relaxed and friendly rapport-building introduction including greetings and thanks, neutral and unbiased manner during interview, relaxing one's face and gently smiling, maintaining eye contact with the camera, good posture), identifying signs of distress, and interviewer welfare (see the Supplementary Document 2 for our video interviewer manual).

Analysis Strategy

In our analyses, we investigate the characteristics of panel respondents who are generally willing to participate in video interviews, consent to it for our VI experiment, make an appointment, and actually participate. Moreover, we explore differences in data quality between respondents in the VI group and respondents in the online self-completion mode control group.

Dependent Variables

In total, we estimate 16 regression models (M1 to M16; see Table 2). The first four focus on the group of panel members invited to VI. Key dependent variables are general VI willingness (M1) and actual consent (M2), making a video interview appointment (given consent and inclusion in the VI group; M3), and completing VI (given having an appointment; M4). The remaining models (M5 to M16) focus on data quality of responses in the VI group versus the comparison group. We calculate three separate models to examine satisficing (M5 to M7) and nine models to examine social desirability indicators (M8 to M16). In the data quality models, the dependent variables are the responses to survey questions supposedly attracting satisficing behavior (number of selected health conditions, rounded numeric punch, reported payments) or social desirability (support for multiculturalism, life satisfaction, general health, psychological distress, generalized trust, smoking, alcohol consumption, exercise frequency, unpaid care; for a full overview of the dependent variables and models see Table 2).

Table 2 Models and variable types across study phases (DV = dependent variable, study phases: 1 = May, 2 = July, 3 = November to December)

Model	Regression model type	Study phase	DV type	DV (categories)
M1	Multinomial logistic	1	Nominal	General willingness to use VI (unwilling, unsure, willing)
M2	Binomial logistic	2	Binary	Consent to participate in VI study (no consent, consent)
M3	Binomial logistic	3	Binary	Made VI appointment (no, yes)
M4	Binomial logistic	3	Binary	Completed VI (no, yes)
M5	Poisson	3	Count	Number of selected health conditions (0–7)
M6	Binomial logistic	3	Binary	Rounded punches (no, yes)
M7	Poisson	3	Count	Number of reported payments (0–5)
M8	Ordinal logistic	3	Ordinal	Support for multiculturalism (Strongly disagree, Disagree, Neither agree nor disagree, Agree, Strongly agree)
M9	Ordinary least squares	3	Continuous	Square of life satisfaction (0–100)
M10	Poisson	3	Count	General health (Poor, Fair, Good, Very good, Excellent)
M11	Ordinary least squares	3	Continuous	Psychological distress[2]
M12	Ordinal logistic	3	Ordinal	Generalized trust (Strongly disagree, Disagree, Neither agree nor disagree, Agree, Strongly agree)
M13	Binomial logistic	3	Binary	Smoking behavior (no, yes)
M14	Ordinal logistic	3	Ordinal	Alcohol consumption frequency (Do not drink alcohol, Less often than 1 day a month, About 1 day a month, 2–3 days a month, 1–2 days a week, 3–4 days a week, 5–6 days a week, Every day)
M15	Ordinary least squares	3	Ordinal	Exercise frequency (Not at all, Less than once a week, 1 to 2 times a week, 3 times a week, More than 3 times a week (but not every day), Every day)
M16	Binomial logistic	3	Binary	Provides unpaid care (no, yes)

We analyze our video interview study data using multivariate regression analyses. To analyze general VI willingness (3-category nominal variable; study phase 1), we apply a multinomial logistic regression model. For the analyses of the VI experiment (study phases 2 and 3) analyzing the different stages of taking part in VI and our data quality indicators, we use - depending on the dependent vari-

[2] Kessler scores 6–30 (Kessler et al., 2010)

able type - binomial logistic regression models, Poisson models, ordinal logistic models, and Ordinary Least Square models.

Independent Explanatory Variables

Independent variables that may explain effects on uptake of VI were selected based on our hypotheses for direction of effects outlined in the section above. The variables include: degree of area remoteness (H1), age (H2), education (H3), gender (H4), frequency of internet use (H5), amount of daily social media use (H6), completion rate across the surveys in the four months prior to the video interview study (H7), and the survey mode (video vs. online self-completion; H8 and H9). As control variables, we also include panel recruitment year, state of residence as well as country of birth and indigenous status (see Table A1 in Appendix A for a full overview of the independent variables and rationale for selection).

Robustness Analyses for Group Selection Effects

As described above, in the VI group, only panel members who had previously consented to VI could be invited while the regular survey comparison group contained VI consenters, non-consenters, and panel members who did not answer the consent question (e.g., because they skipped the phase 2 survey).This difference between the groups is of no consequence for M1 to M4, which focus exclusively on the group invited for video interviews. However, this potential selection effect may impact the data quality comparisons between the VI respondents and the web self-completion respondents in M5 to M16. We, therefore, examined to what extent the invited and responding samples across these two groups differed (see Table B1). Results show that panel members invited to (see additional model Add. M2) as well as participating in the video interviews (see additional model Add. M3) were significantly more likely to have higher education, higher internet use frequency, and higher survey completion rates than the online self-completion comparison group invitees and respondents. Panel members invited to the video interviews were also more likely to be in the oldest age group (75 years and older; reference category: 18 to 24 years) and use social media less than 30 minutes a day (reference category: not at all). Women were less likely to respond to the video interview than the online self-completion survey. All differences are accounted for through inclusion of the relevant characteristics in the set of analyses variables applied throughout all models.

Results

First, we model uptake of VI (models M1-M4) and focus on understanding the influence of the key variables of interest (relating to hypotheses H1 to H7) across all three study phases We then examine the data quality indicators (satisficing [H8] and socially desirable responding [H9]) of the actual responses on key variables of interest, comparing the VI group with the comparison group.

Uptake of VI: Willingness, Consent, Appointment-Making, and Survey Completion

To summarize the earlier overview described above, we asked an initial question on general willingness to participate in VI in May 2022, where 42% of panel members agreed in principle to participate, 13% were unsure, and 45% were unwilling. Actual VI consent was asked in July 2022, again with 42% consenting. Only a subset of panel members who consented were selected for the VI arm of the trial. Of the 2,225 panel members assigned to the VI group, 1,399 were released into field in batches until the target number 600 VIs was reached. Among the 1,399 panel members who were invited, attrition occurred at appointment setting (51% invited did not set an appointment) and appointment keeping (8% invited set an appointment but did not complete an interview), with 43% of panel members invited being interviewed. By contrast, 73% of panel members invited to the regular arm completed the survey in their usual mode of interview.

Higher VI participation was expected in major cities as compared to regional and remote areas (H1). No support for this hypothesis was found for general willingness to participate (M1) or at any of the three study phases (M2–M4).

Table 3 Analysis of VI uptake: estimated regression outcomes for models M1-M4, for key explanatory variables (as they relate to our hypotheses)

Variable	M1: General willingness		M2: Consent	M3: Appointment	M4: Completed
Category	Yes	Maybe	Yes	Yes	Yes
Reference category	No	No	No	No	No
Estimate	*RR (SE)*	*RR (SE)*	*OR (SE)*	*OR (SE)*	*OR (SE)*
Remoteness (reference = Major Cities) (H1)					
Inner Regional	0.94 (0.10)	1.09 (0.16)	0.98 (0.08)	1.13 (0.17)	1.34 (0.41)
Outer Regional	0.86 (0.14)	0.87 (0.20)	0.79 (0.10)	0.80 (0.20)	0.78 (0.35)
(Very) Remote	0.82 (0.35)	0.47 (0.36)	0.82 (0.27)	0.58 (0.39)	0.35 (0.40)
Age group (reference = 18–24 years) (H2)					
25–34 years	1.52 (0.34)	1.61 (0.61)	1.75** (0.30)	1.11 (0.46)	2.07 (1.66)
35–44 years	1.46 (0.32)	2.10* (0.77)	1.54* (0.26)	1.16 (0.48)	1.14 (0.88)
45–54 years	1.72* (0.39)	2.42* (0.89)	1.58** (0.27)	1.30 (0.54)	3.33 (2.70)
55–64 years	2.14** (0.48)	1.95 (0.72)	2.06*** (0.35)	1.38 (0.57)	2.36 (1.86)
65–74 years	2.12** (0.49)	2.24* (0.84)	2.17*** (0.38)	1.65 (0.69)	2.77 (2.18)
75 or more years	1.79* (0.45)	2.67* (1.05)	2.37*** (0.45)	1.13 (0.50)	3.26 (2.78)
Highest educational attainment (reference = less than Year 12) (H3)					
Year 12	1.22 (0.22)	0.67 (0.15)	1.27 (0.16)	0.74 (0.21)	1.45 (0.79)
Advanced certificate	1.99*** (0.30)	0.92 (0.16)	1.71*** (0.19)	0.72 (0.18)	1.13 (0.49)
Bachelor‘s degree	2.36*** (0.37)	0.69 (0.13)	1.86*** (0.21)	0.91 (0.23)	1.51 (0.69)
Graduate degree	2.53*** (0.39)	0.61* (0.12)	2.41** (0.27)	0.90 (0.22)	1.58 (0.70)
Unable to establish	1.65 (0.49)	0.51 (0.20)	1.26* (0.25)	0.95 (0.45)	0.92 (0.73)
Gender (reference = male) (H4)					
Female	0.71*** (0.06)	0.99 (0.12)	0.80*** (0.05)	0.82 (0.10)	0.74 (0.18)
Non-binary	0.89 (0.48)	1.27 (1.05)	1.01 (0.45)	0.27 (0.32)	–

Table 1 (continued)

Variable	M1: General willingness		M2: Consent	M3: Appointment	M4: Completed
Category	Yes	Maybe	Yes	Yes	Yes
Reference category	No	No	No	No	No
Estimate	*RR (SE)*	*RR (SE)*	*OR (SE)*	*OR (SE)*	*OR (SE)*
Frequency of internet use (reference = less than once a day) (H5)					
About once a day	1.36 (0.38)	0.76 (0.23)	1.25 (0.25)	1.82 (0.79)	2.61 (1.91)
Several times a day	1.96** (0.47)	0.89 (0.23)	1.64** (0.27)	2.24* (0.84)	4.12 (2.63)
Almost constantly	2.46*** (0.61)	0.81 (0.22)	1.90*** (0.33)	2.35* (0.91)	8.15** (5.53)
Amount of daily social media use (reference = Not at all) (H6)					
< 30 minutes	1.14 (0.13)	1.23 (0.19)	1.42*** (0.12)	1.03 (0.17)	1.04 (0.33)
30 to 59 minutes	1.29* (0.16)	0.96 (0.18)	1.48*** (0.14)	0.72 (0.14)	0.98 (0.36)
1 to < 2 hours	1.12 (0.15)	1.05 (0.20)	1.31** (0.13)	0.66* (0.14)	1.45 (0.65)
2 to 4 hours	1.40* (0.20)	1.03 (0.22)	1.43** (0.16)	1.02 (0.22)	1.32 (0.57)
Over 4 hours	0.97 (0.17)	1.02 (0.26)	1.40* (0.18)	1.01 (0.29)	0.40 (0.19)
Completion rate × 100 (H7)	1.01** (0.00)	1.01** (0.01)	1.01** (0.00)	1.03*** (0.00)	1.03** (0.01)
N	3,420		5,523	1,388	710
X^2	270.27		233.87	127.10	69.23
D.F.	76		38	33	37
P	< .001		< .001	< .001	< .001
McFadden's R^2	0.04		0.03	0.07	0.11
Log-likelihood	−3,232.67		−3,641.71	−898.17	−278.22

Notes: Relative risk ratios (*RR*), odds ratios (*OR*), and standard errors (*SE*); controls and intercept not shown.*$p < .05$; **$p < .01$; ***$p < .001$.

Higher participation from younger adults was expected (H2). For general willingness to participate (M1), contrary to expectations, older adults were more willing to participate in VI than younger adults (highest odds ratios for those aged 55 to 74 years). Moreover, older panel members, particularly those 75 years and older, are more likely to be undecided than unwilling to participate. Similarly, H2 was rejected for consent to participate in VI (M2), where older adults were also more likely to consent to a VI than younger adults. Among those who consented, however, older adults were not significantly more likely to make an appointment for (M3) or complete the VI (M4) than younger adults[3].

Higher participation in VI for more highly educated adults was expected (H3). Results were consistent with H3 for general willingness (M1) and actual consent (M2) to participate. However, no differences were found for scheduling an appointment (M3) or completing the survey (M4).

Women were expected to be more likely to participate in VI than men (H4). Contrary to H4, women were generally less willing to try VI (M1) and to consent to it (M2) than men. There was no statistically significant gender difference, however, regarding appointment making (M3) or survey completion (M4).

More frequent internet users were expected to be more likely to participate in VI than those who used the internet less frequently (H5). Consistent with H5, frequent internet users were more likely to generally be willing to try VI (M1), consent to participate in VI (M2), schedule an appointment (M3), and complete the survey (M4) than less frequent internet users.

More frequent users of social media were expected to be more likely to participate than less frequent users (H6). In partial alignment with H6, panel members reporting 30 to 59 minutes per day of social media use and those reporting 2 to 4 hours are more likely to be generally willing to participate in video interviews than those who did not use social media (M1). Moreover, panel members who used social media were more likely to consent to VI than those who did not (M2). However, users with higher levels of social media use were not more likely to make an appointment (M3) or complete VI (M4).

Panel members with high prior wave completion rates were expected to be more likely to participate in VI than panel members with lower completion rates (H7). In line with H7, panel members with higher completion rates were more likely to be willing to participate in VI (M1), consent to VI (M2), make an appointment (M3), and complete the VI (M4) than panel members with lower prior survey wave completion rates.

Regarding our control variables, we find that panel members born overseas are less likely to be willing to agree and to consent to VI than panel members

[3] A potential post hoc explanation for these findings is that older panel members may be drawn to VI due to a desire for interviewer assistance and a more conversational survey environment. However, they may lack confidence in their digital skills or equipment perceived as necessary to master the task.

born in Australia (see Appendix B). Moreover, panel members recruited in 2018 are more likely to be willing to try video interviews or be undecided about it than panel members recruited in 2016. We find no significant association with state of residence (except a lower likelihood of Queensland panel members to complete the video interviews) or Indigenous status.

The explanatory power of all results reported in this section is small with McFadden's McFadden's R2 statistics between 0.03 and 0.11.

Data Quality of Survey Responses: Satisficing and Social Desirability

This section reports the results for the data quality indicators on satisficing and social desirability in survey responses, comparing VI and online response modes.

Satisficing (H8)

We expected that satisficing would be lower for VI mode (H8). Our regression models reveal no support for H8 (see Table 4 for results and Table 2 for details on the satisficing variables). There were no significant differences between the modes regarding the number of selected health conditions and rounding numeric punches. In M7, the coefficient was significant, albeit only at the $p < .05$ level, but in the opposite direction to H8 (i.e., VI had fewer reported payments, from which we infer more satisficing).

Table 4 Regression model coefficients comparing satisficing indicators in the video and online self-completion group (key independent variable: survey mode, controlling for sample composition)

Dependent variable of interest (survey response variable)	Model	Coefficient for VI mode (reference: online mode): *b (SE)*	*N*
Number of selected health conditions	M5	0.07 (0.06)	1,176
Rounded punches	M6	0.13 (0.16)	1,168
Number of reported payments	M7	−0.27* (0.11)	1,176

Notes: Model type: Poisson (M5, M7) or logit (M6). Controls (not shown; all dummy variables except completion rate): year of recruitment to panel, completion rate in four prior waves, state, remoteness, gender, age group, place of birth (Australia, overseas), Indigenous status (Indigenous, not Indigenous), and frequency of internet use, frequency of social media use.
*$p < .05$; **$p < .01$; ***$p < .001$.

Social Desirability (H9)

We find evidence of higher socially desirable responding (SDR) for VI (see results in table 5) supporting our hypothesis H9. Table 5 shows the coefficient for mode of interview (VI, where online is the reference), controlling for sample composition. The results are consistent with the hypothesized (H9) higher SDR for VI in seven of nine variables tested (i.e., significant at the $p < .01$ level in the hypothesized direction of effect): higher support for multiculturalism, higher life satisfaction, better general health, higher generalized trust, higher exercise frequency, higher reported provision of unpaid care, and lower psychological distress after controlling for sample composition differences across the two groups (see Table 2 for details). The only results inconsistent with higher SDR for VI are for substance use: smoking and alcohol consumption were not significantly different between VI and online mode at the $p < .05$ level.

Table 5 Regression model coefficients comparing social desirability indicators in the VI and online self-completion sample (key independent variable: survey mode, controlling for sample composition)

Dependent variable of interest (hypothesized direction of SDR)	Coefficient of VI mode (reference: online mode): b (*SE*)	*N*
M8: Support for multiculturalism (+)	1.07*** (0.12)	1,174
M9: Square of life satisfaction (+)	5.25*** (1.30)	1,176
M10: General health (+)	0.39** (0.11)	1,176
M11: Psychological distress (−)	−0.89*** (0.23)	1,175
M12: Generalized trust (+)	0.39** (0.12)	1,176
M13: Smoking behavior (−)	−0.25 (0.26)	1,174
M14: Alcohol consumption frequency (−)	0.06 (0.11)	1,174
M15: Exercise frequency (+)	0.41*** (0.11)	1,176
M16: Provides unpaid care (+)	0.50** (0.17)	1,143

Notes: Model type: ordinal logit (M8, M12, M14, M15), OLS (M9, M11), Poisson (M10), binomial logit (M13, M16). Controls (not shown; all dummy variables except completion rate): year of recruitment to panel, completion rate in four prior waves, state, remoteness, gender, age group, place of birth (Australia, overseas), Indigenous status (Indigenous, not Indigenous), and frequency of internet use, frequency of social media use. **p* < .05; **p < .01; ****p* < .001.

Summary and Conclusion

We assessed the feasibility of conducting live VI in Australia using the probability-based online panel Life in Australia™. We investigated the factors predicting uptake of VI and data quality of resulting survey responses. We found that VI attracted frequent internet users with a history of high panel survey completion rates. When compared to a subsample surveyed in the regular Life in Australia™ online self-completion mode, we found that VI seems to lead to more socially desirable responses while not helping to reduce satisficing behavior. Some of these results were expected while others were not (see Table 6 for an overview of which of our hypotheses were supported or had to be rejected).

Table 6 Summary of influencing factors on uptake of VI and the data quality of survey responses: our hypotheses and to what extent they are supported by the findings

Hypothesis	Supported?
Respondents used to online self-completion surveys will be more likely to take up video interviews if …	
… they live in urban rather than remote areas (H1).	No
… they are younger (H2)	No
… they are more highly educated (H3)	Partially
… they are female (H4).	No
… they are more frequent internet users (H5).	Yes
… they use social media (H6).	Partially
… they have a history of high panel survey completion rates (H7).	Yes
The presence of the interviewer in the video interviews will reduce satisficing behavior (H8).	No
The presence of the interviewer in the video interviews will increase social desirability (H9).	Yes

There may be multiple explanations for our findings. For example, Life in Australia™ panel members are used to and seem to be comfortable with online self-completion (or telephone interviews without appointments) and may, therefore, not see any reason to switch to other modes. In addition, having to schedule an appointment rather than participating 'spontaneously' may not have appealed to the respondents. Our results thus indicate that in the context of an already established online panel, it may have very limited benefits to offer respondents VI as a new survey mode option as respondents already routinely participate in a more cost-efficient and flexible survey mode. Results may be different in studies on cross-sectional surveys or panel recruitment, warranting further research.

When interpreting these findings, it is important to consider that results may differ in research settings with longer and more complex questionnaires, different topics, or interviewers with prior experience in face-to-face interviewing (rather than telephone interviewing). Furthermore, one could speculate that the finding that older panel members expressed higher willingness and consent for video interviews than younger panel members but then did not participate in the video interviews at higher rates suggests that some people may want interviewer assistance but face digital skill or equipment barriers preventing them from seizing the opportunity. A crucial barrier may have been the necessity to make an appointment before they could complete the survey.

We expect our results to be generalizable to other panels with online self-completion as the only or dominant data collection mode, although more research from other panels is needed to confirm this. However, we expect limited generalizability to cross-sectional surveys, where respondents do not have a survey mode history. We do not expect our results to be generalizable to panel surveys where respondents have an interviewer-administered survey mode history, especially face-to-face surveys. It is possible that respondents from those surveys may perceive video interviews as a more flexible and convenient mode option compared to what they are used to. Moreover, fieldwork costs for the survey agency can be expected to be lower in VI than face-to-face interviewing. Furthermore, our analyses are limited by the fact that in the data quality analyses we compare panel members who consented to VI to panel members who mostly did not do so, either due to consent refusal, nonresponse, or ineligibility.

Since our study did not indicate much added value of offering VI in an established online panel, future research should focus on other survey design settings in which VI may have a greater impact, for example panel recruitment or cross-sectional surveys. Future research with fully randomized experimental design allocation may show to what extent our findings hold and are causally attributable to the VI mode. Future studies should also consider finding ways to avoid the necessity to make an appointment to the VI and instead offer on-demand VI. In addition, it may be useful to collect interview paradata through the interviewers, for example on the extent to which technical problems occurred during an interview and whether the respondents seemed comfortable with the interview situation as such aspects may influence the quality of the provided data. All in all, more research into various aspects of VI and its optimal implementation is necessary if this mode is to establish itself as part of the survey mode toolbox.

Funding

We gratefully acknowledge the funding of this study provided by the Australian Bureau of Statistics and The Social Research Centre (SRC).

References

Al Baghal, T., & Lynn, P. (2015). Using Motivational Statements in Web-Instrument Design to Reduce Item-Missing Rates in a Mixed-Mode Context. *Public Opinion Quarterly, 79*(2), 568–579. https://doi.org/10.1093/poq/nfv023

Australian Bureau of Statistics. (2012). *Information paper: Use of the Kessler Psychological Distress Scale in ABS health surveys, Australia, 2007-08* (cat. no. 4817.0.55.001). https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/4817.0.55.001Chapter192007-08

Australian Bureau of Statistics. (2021a). *Remoteness Areas: Australian Statistical Geography Standard (ASGS) edition 3*. https://www.abs.gov.au/statistics/standards/australian-statistical-geography-standard-asgs-edition-3/jul2021-jun2026/remoteness-structure/remoteness-areas

Australian Bureau of Statistics. (2021b). *Standard for sex, gender, variations of sex characteristics and sexual orientation variables*. https://www.abs.gov.au/statistics/standards/standard-sex-gender-variations-sex-characteristics-and-sexual-orientation-variables/latest-release

Australian Communications and Media Authority. (2023). *How we communicate* (Communications and media in Australia). https://www.acma.gov.au/publications/2023-12/report/communications-and-media-australia-how-we-communicate

Australian Communications and Media Authority. (2024). *How we use the internet. Executive summary and key findings* (Communications and media in Australia). https://www.acma.gov.au/sites/default/files/2024-12/How%20we%20use%20the%20internet%20-%20Executive%20summary%20and%20key%20findings.pdf

Australian Government Department of Education. (2019). *International Standard Classification of Education 2011 (ISCED 2011) to Australian Standard Classification of Education (ASCED) Concordance*. https://www.tcsisupport.gov.au/sites/default/files/2019-12/isced-asced-concordance.pdf

Bach, R. L., Cornesse, C., & Daikeler, J. (2024). Equipping the Offline Population with Internet Access in an Online Panel: Does It Make a Difference? *Journal of Survey Statistics and Methodology, 12*(1), 80–93. https://doi.org/10.1093/jssam/smad003

Bianchi, A., Biffignandi, S., & Lynn, P. (2017). Web-Face-to-Face Mixed-Mode Design in a Longitudinal Survey: Effects on Participation Rates, Sample Composition, and Costs. *Journal of Official Statistics, 33*(2), 385–408. https://doi.org/10.1515/jos-2017-0019

Burton, J., Lynn, P., & Benzeval, M. (2020). How Understanding Society: The UK Household Longitudinal Study adapted to the COVID-19 pandemic. *Survey Research Methods, 14*(2), 235–239. https://doi.org/10.18148/srm/2020.v14i2.7746

Callegaro, M., & DiSogra, C. (2008). Computing response metrics for online panels. *Public Opinion Quarterly, 72*(5), 1008–1032. https://doi.org/10.1093/poq/nfn065

Cernat, A., & Revilla, M. (2021). Moving from Face-to-Face to a Web Panel: Impacts on Measurement Quality. *Journal of Survey Statistics and Methodology, 9*(4), 745–763. https://doi.org/10.1093/jssam/smaa007

Conrad, F. G., Schober, M. F., Hupp, A. L., West, B. T., Larsen, K. M., Ong, A. R., & Wang, T. (2023). Video in Survey Interviews: Effects on Data Quality and Respondent Experience. *Methods, Data, Analyses, 17*(2), 135–170. https://doi.org/10.12758/mda.2022.13

Cornesse, C., Blom, A. G., Dutwin, D., Krosnick, J. A., de Leeuw, E. D., Legleye, S., Pasek, J., Pennay, D., Phillips, B., Sakshaug, J. W., Struminskaya, B., & Wenz, A. (2020). A review of conceptual approaches and empirical evidence on probability and nonprobability sample survey research. *Journal of Survey Statistics and Methodology, 8*(1), 4–36. https://doi.org/10.1093/jssam/smz041

Cornesse, C., & Schaurer, I. (2021). The Long-Term Impact of Different Offline Population Inclusion Strategies in Probability-Based Online Panels: Evidence From the German Internet Panel and the GESIS Panel. *Social Science Computer Review, 39*(4), 687–704. https://doi.org/10.1177/0894439320984131

Daikeler, J., Bošnjak, M., & Lozar Manfreda, K. (2020). Web Versus Other Survey Modes: An Updated and Extended Meta-Analysis Comparing Response Rates. *Journal of Survey Statistics and Methodology, 8*(3), 513–539. https://doi.org/10.1093/jssam/smz008

de Leeuw, E., Hox, J., & Luiten, A. (2018). International Nonresponse Trends across Countries and Years: An analysis of 36 years of Labour Force Survey data. *Survey Methods: Insights from the Field,* 1–11. https://doi.org/10.13094/SMIF-2018-00008

Destatis (2024). *Informations- und Kommunikationstechnologien privater Haushalte (Mikrozensus-Unterstichprobe) – Endergebnis 2024* https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Einkommen-Konsum-Lebensbedingungen/IT-Nutzung/Publikationen/Downloads-IT-Nutzung/statistischer-bericht-ikt-privater-haushalte-2150400247005.xlsx?__blob=publicationFile&v=5.

Dulaney, R., Kelley, J., Arrue, J., & Edwards, B. (2023, July 17-21). *Who is Choosing Video Interviewing? A Look into Respondent Characteristics of Video Adopters* [Conference presentation]. European Survey Research Association Conference 2023, Milan, Italy.

Dutwin, D., & Lavrakas, P. (2016). Trends in telephone outcomes, 2008-2015. *Survey Practice, 9*(3). https://doi.org/10.29115/SP-2016-0017

Endres, K., Hillygus, D. S., DeBell, M., & Iyengar, S. (2023). A randomized experiment evaluating survey mode effects for video interviewing. *Political Science Research and Methods, 11*(1), 144–159. https://doi.org/10.1017/psrm.2022.30

Engward, H., Goldspink, S., Iancu, M., Kersey, T., & Wood, A. (2022). Togetherness in Separation: Practical Considerations for Doing Remote Qualitative Interviews Ethically. *International Journal of Qualitative Methods, 21,* 16094069211073212. https://doi.org/10.1177/16094069211073212

Eurostat. (2022, August 22). *Increase in high-speed internet coverage in 2021.* https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20220822-1

Fuchs, M. (2009). Gender-of-Interviewer Effects in a Video-Enhanced Web Survey: Results from a Randomized Field Experiment. *Social Psychology, 40*(1), 37–42. https://doi.org/10.1027/1864-9335.40.1.37

Haan, M. (2015). *Mode matters: Effects of survey modes on participation and answering behavior* [Doctoral dissertation]. University of Groningen. https://research.rug.nl/en/publications/mode-matters-effects-of-survey-modes-on-participation-and-answeri/

Haan, M., Ongena, Y. P., Vannieuwenhuyze, J. T. A., & de Glopper, K. (2017). Response Behavior in a Video-Web Survey: A Mode Comparison Study. *Journal of Survey Statistics and Methodology, 5*(1), 48–69. https://doi.org/10.1093/jssam/smw023

Hanson, T. (2021, May 11). *The European Social Survey during COVID-19: using video interviews and other innovations* [conference paper]. American Association for Public Opinion Research Annual Conference, online.

Heerwegh, D., & Loosveldt, G. (2008). Face-to-Face versus Web Surveying in a High-Internet-Coverage Population: Differences in Response Quality. *Public Opinion Quarterly, 72*(5), 836–846. https://doi.org/10.1093/poq/nfn045

Helsper, E. (2021). *The Digital Disconnect: The Social Causes and Consequences of Digital Inequalities.* SAGE Publications. https://doi.org/10.4135/9781526492982

Höhne, J. K., Ziller, C., & Lenzner, T. (2024). Investigating respondents' willingness to participate in video-based web surveys. *International Journal of Market Research, 66*(1), 3–13. https://doi.org/10.1177/14707853231198788

Jäckle, A., Cornesse, C., Wenz, A., & Couper, M. P. (2024). Measuring Expenditure with a Mobile App: Do Probability-Based and Nonprobability Panels Differ? *Journal of Survey Statistics and Methodology, 12*(5), 1224–1253. https://doi.org/10.1093/jssam/smae026

Jeannis, M., Terry, T., Heman-Ackah, R., & Price, M. (2013). Video Interviewing: An exploration of the feasibility as a mode of survey application. *Survey Practice, 6*(1). https://doi.org/10.29115/SP-2013-0001

Kaczmirek, L., Phillips, B., Pennay, D., Lavrakas, P. J., & Neiger, D. (2019). *Building a probability-based online panel: Life in Australia™* (CSRM and SRC Methods Paper No. 2/2019). Social Research Centre of the Australian National University. https://polis.cass.anu.edu.au/research/publications/building-probability-based-online-panel-life-australia

Kessler, R. C., Green, J. G., Gruber, M. J., Sampson, N. A., Bromet, E., Cuitan, M., Furukawa, T. A., Gureje, O., Hinkov, H., Hu, C.-Y, Lara, C., Lee, S., Mneimneh, Z., Myer, L., Oakley-Browne, M., Posada-Villa, J., Sagar, R., Viana, M. C., & Zaslavsky, A.M. (2010). Screening for serious mental illness in the general population with the K6 screening scale: Results from the WHO World Mental Health (WMH) survey initiative. *International Journal of Methods in Psychiatric Research, 19*(S1), 4–22. https://doi.org/10.1002/mpr.310

Kocar, S., Brown, M., Sanchez, C., Ploubidis, G. B., & Durrant, G (2025) Measurement Mode Effects in Live Video Interviewing: Evidence from a Mixed-Mode Quasi-Experimental Survey. Survey Futures Working Paper no. 11. Colchester, UK: University of Essex. Available at https://surveyfutures.net/working-papers/

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

Lavrakas, P. J., Pennay, D., Neiger, D., & Phillips, B. (2022). Comparing probability-based surveys and nonprobability online panel surveys in Australia: a total survey error perspective. *Survey Research Methods, 16*(2), 241–266. http://dx.doi.org/10.18148/srm/2022.v16i2.7907

Martin, L., & Fradier, Y. (2023, July 17-21). *ESS + CAVIsio* [Conference presentation]. European Survey Research Association Conference 2023, Milan, Italy.

Maslovskaya, O., & Lugtig, P. (2022). Representativeness in Six Waves of Cross-National Online Survey (CRONOS) Panel. *Journal of the Royal Statistical Society Series A: Statistics in Society, 185*(3), 851–871. https://doi.org/10.1111/rssa.12801

McClain, C., Vogels, E. A., Perrin, A., Sechopoulos, S., & Rainie, L. (2021). *The Internet and Technology During the COVID-19 Pandemic: April 2021 Survey*. Pew Research Center. https://policycommons.net/artifacts/1811100/the-internet-and-the-pandemic/2547007/

Oates, M., Crichton, K., Cranor, L., Budwig, S., Weston, E. J. L., Bernagozzi, B. M., & Pagaduan, J. (2022). Audio, video, chat, email, or survey: How much does online interview mode matter? *PLOS ONE, 17*(2), e0263876. https://doi.org/10.1371/journal.pone.0263876

Pennay, D. W., Neiger, D., Lavrakas, P. J., & Borg, K. (2018). *The Online Panels Benchmarking Study: a total survey error comparison of findings from probability-based surveys and non-probability online panel surveys in Australia* (CSRM and SRC Methods Paper No. 2/2018). Social Research Centre of the Australian National University. https://polis.cass.anu.edu.au/research/publications/online-panels-benchmarking-study-total-survey-error-comparison-findings

Pennay, D. Phillips, B., Neiger, D., Ward, A., Slamowicz, S., & Lethborg, A. (2023). *Results from the 2022 Australian Comparative Study of Survey Methods.* Social Research Centre of the Australian National University. https://srcentre.com.au/wp-content/uploads/2023/12/ACSSM-Analytical-Report-20231005.pdf

Pew Research Center. (2019). *Response rates in telephone surveys have resumed their decline*. https://www.pewresearch.org/fact-tank/2019/02/27/response-rates-in-telephone-surveys-have-resumed-their-decline/

Peytchev, A. (2009). Survey Breakoff. *Public Opinion Quarterly, 73*(1), 74–97. https://doi.org/10.1093/poq/nfp014

Phillips, B., Neiger, D., Ward, A., Slamowicz, S., & Lethborg, A. (2023). *Australian Comparative Study of Survey Methods: Technical Report.* Social Research Centre of the Australian National University. https://srcentre.com.au/wp-content/uploads/2024/01/ACSSM-Technical-Report-with-Appendices.pdf

Revilla, M., & Ochoa, C. (2015). What are the Links in a Web Survey Among Response Time, Quality, and Auto-Evaluation of the Efforts Done? *Social Science Computer Review, 33*(1), 97–114. https://doi.org/10.1177/0894439314531214

Schober, M. F. (2018). The future of face-to-face interviewing. *Quality Assurance in Education, 26*(2), 290–302. https://doi.org/10.1108/QAE-06-2017-0033

Schober, M. F., Conrad, F. G., Hupp, A. L., Larsen, K. M., Ong, A. R., & West, B. T. (2020). Design considerations for live video survey interviews. *Survey Practice, 13*(1). https://doi.org/10.29115/SP-2020-0014

Schober, M. F., Okon, S., Conrad, F. G., Hupp, A. L., Ong, A. R., & Larsen, K. M. (2023). Predictors of Willingness to Participate in Survey Interviews Conducted by Live Video. *Technology, Mind, and Behavior, 4*(2), 215–229. https://doi.org/10.1037/tmb0000100

Smith, T. W. (1995). Trends in non-response rates. *International Journal of Public Opinion Research, 7*(2), 157–171. https://doi.org/10.1093/ijpor/7.2.157

Smyth, J. D., Olson, K., & Millar, M. M. (2014). Identifying predictors of survey mode preference. *Social Science Research, 48,* 135–144. https://doi.org/10.1016/j.ssresearch.2014.06.002

Thomas, J., McCosker, A., Parkinson, S., Hegarty, K., Featherstone, D., Kennedy, J., Holcombe-James, I., Ormond-Parker, L., & Ganley, L. (2023). *Measuring Australia's digital divide: The Australian digital inclusion index 2023*. ARC Centre of Excellence for Automated Decision-Making and Society, RMIT University, Swinburne University of Technology, & Telstra. https://doi.org/10.25916/528S-NY91

Thorolfsson, A., Hjaltason, A. B., Jonsdottir, G. A., & Gudmundsson, H. (2023, July 17-21). *Mode effects of video interviewing as a proxy of face to face interviewing in R10 of ESS in Iceland* [Conference presentation]. European Survey Research Association Conference 2023, Milan, Italy.

Weller, S. (2017). Using internet video calls in qualitative (longitudinal) interviews: Some implications for rapport. *International Journal of Social Research Methodology, 20*(6), 613–625. https://doi.org/10.1080/13645579.2016.1269505

West, B. T., & Blom, A. G. (2017). Explaining Interviewer Effects: A Research Synthesis. *Journal of Survey Statistics and Methodology, 5*(2), 175–211. https://doi.org/10.1093/jssam/smw024

West, B. T., Ong, A. R., Conrad, F. G., Schober, M. F., Larsen, K. M., & Hupp, A. L. (2022). Interviewer Effects in Live Video and Prerecorded Video Interviewing. *Journal of Survey Statistics and Methodology, 10*(2), 317–336. https://doi.org/10.1093/jssam/smab040

Williams, D., & Brick, J. M. (2018). Trends in US face-to-face household survey nonresponse and level of effort. *Journal of Survey Statistics and Methodology, 6*(2), 186–211. https://doi.org/10.1093/jssam/smx019

Zhang, C., & Conrad, F. (2014). Speeding in Web Surveys: The tendency to answer very fast and its association with straightlining. *Survey Research Methods, 8*(2), 127-135. https://doi.org/10.18148/srm/2014.v8i2.5453

Appendix A. Models and Variables

Table A1 Independent variables of interest with rationale for inclusion and control variables

Hypothesis	Variable	Description	Categories / range
H1	Remoteness	Remoteness Index of Australia plus area of residence (Australian Bureau of Statistics, 2021b), Remote and Very Remote Australia were collapsed due to small cell sizes	Major Cities of Australia, Inner Regional Australia, Outer Regional Australia, Remote or Very Remote Australia
H2	Age group	Respondent age in ranges. Older adults are generally overrepresented in Life in Australia™	18-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75+
H3	Highest educational attainment	Educational attainment (Australian Government Department of Education, 2019), cells were collapsed due to small sample size. Highly educated adults are generally overrepresented in Life in Australia™	Less than Year 12, Year 12, Certificate III/IV or Diploma or Advanced Diploma, Bachelor's Degree, Graduate Certificate or Graduate Diploma or Postgraduate Degree
H4	Gender	Respondent gender (Australian Bureau of Statistics, 2021a), non-binary and self-described were collapsed due to small cell sizes. Women are overrepresented in Life in Australia™	man or male, woman or female, non-binary/self-described
H5	Frequency of internet use	Self-reported frequency of using the internet	Less than once a day, about once a day, several times a day, almost constantly
H6	Amount of daily social media use	Self-reported frequency of social media use	Not at all, less than 30 mins, 30 to 59 mins, 1 to less than 2 hours, 2 to 4 hours, over 4 hours
H7	Completion rate	Proportion of waves completed between August and November 2022 out of waves invited to, multiplied by 100	0 (i.e., no completed waves) to 100 (i.e. all waves completed)
H8 and H9	survey mode	operationalizes the experimental and control group to examine mode differences in responses	Online self-completion, video interview

Table A1 (continued)

Hypothesis	Variable	Description	Categories / range
Controls	Recruitment year	Year of recruitment into panel, control for differential attrition due to time in panel and method of recruitment	2016 to 2021
	State of residence	State/territory of residence, control for small degree of over- and under-representation on Life in Australia™ by state	NSW, VIC, QLD, SA, WA, TAS, NT, ACT
	Born overseas	Australian analogue of race/ethnicity together with indigenous status	born in Australia, born overseas
	Indigenous status	Australian analogue of race/ethnicity, in combination with nativity	Not Indigenous, Aboriginal or Torres Strait Islander

Appendix B. Full Sample Composition Analyses Including Sensitivity Analyses

Table B1 Regression outcomes in multivariate analyses across VI selection steps including control variables and results for additional robustness analysis models

Variable	M1: General willingness		M2: Consent	M3: Appointment	M4: Completed	Add. M1: Substituted	Add. M2: Invited	Add. M3: Respondents
Category	Yes	Maybe	Yes	Yes	Yes	Substitute	Video	Video
Reference	No	No	No	No	No	Substituted	Online	Online
Estimate	*RR (SE)*	*RR (SE)*	*OR (SE)*	*OR (SE)*	*OR (SE)*	*OR (SE)*	*OR (SE)*	*OR (SE)*
Remoteness (reference = Major Cities)								
Inner Regional	0.94 (0.10)	1.09 (0.16)	0.98 (0.08)	1.13 (0.17)	1.34 (0.41)	1.30 (0.42)	0.99 (0.13)	1.05 (0.17)
Outer Regional	0.86 (0.14)	0.87 (0.20)	0.79 (0.10)	0.80 (0.20)	0.78 (0.35)	2.03 (0.94)	0.71 (0.14)	0.65 (0.18)
(Very) Remote	0.824 (0.35)	0.47 (0.36)	0.82 (0.27)	0.58 (0.39)	0.35 (0.40)	0.55 (0.74)	0.70 (0.39)	0.99 (0.94)
Age group (reference = 18–24 years)								
25–34 years	1.52 (0.34)	1.61 (0.61)	1.75** (0.30)	1.11 (0.46)	2.07 (1.66)	0.62 (0.59)	1.55 (0.46)	1.69 (0.77)
35–44 years	1.46 (0.32)	2.10* (0.77)	1.54* (0.26)	1.16 (0.48)	1.14 (0.88)	0.67 (0.65)	1.38 (0.45)	1.53 (0.62)
45–54 years	1.72* (0.39)	2.42* (0.89)	1.58** (0.27)	1.30 (0.54)	3.33 (2.70)	1.05 (1.01)	1.36 (0.40)	1.50 (0.67)
55–64 years	2.14** (0.48)	1.95 (0.72)	2.06*** (0.35)	1.38 (0.57)	2.36 (1.86)	0.52 (0.51)	1.69 (0.50)	1.92 (0.86)
65–74 years	2.12** (0.49)	2.24* (0.84)	2.17*** (0.38)	1.65 (0.69)	2.77 (2.18)	0.73 (0.72)	1.66 (0.50)	1.99 (0.90)
75 or more	1.79* (0.45)	2.67* (1.05)	2.37*** (0.45)	1.13 (0.50)	3.26 (2.78)	0.56 (0.58)	1.98* (0.65)	2.15 (1.04)
Highest educational attainment (reference = less than Year 12)								
Year 12	1.22 (0.22)	0.67 (0.15)	1.27 (0.16)	0.74 (0.21)	1.45 (0.79)	0.51 (0.29)	1.53 (0.35)	1.61 (0.48)
Advanced certificate	1.99*** (0.30)	0.92 (0.16)	1.71*** (0.19)	0.72 (0.18)	1.13 (0.49)	0.58 (0.27)	1.60* (0.30)	1.49 (0.37)
Bachelor's degree	2.36*** (0.37)	0.69 (0.13)	1.86*** (0.21)	0.91 (0.23)	1.51 (0.69)	0.66 (0.30)	1.66** (0.46)	1.84* (0.32)

Table B1 (continued)

Variable	M1: General willingness		M2: Consent	M3: Appointment	M4: Completed	Add. M1: Substituted	Add. M2: Invited	Add. M3: Respondents
Category	Yes	Maybe	Yes	Yes	Yes	Substitute	Video	Video
Reference	No	No	No	No	No	Substituted	Online	Online
Estimate	*RR (SE)*	*RR (SE)*	*OR (SE)*	*OR (SE)*	*OR (SE)*	*OR (SE)*	*OR (SE)*	*OR (SE)*
Graduate degree	2.53*** (0.39)	0.61* (0.12)	2.41** (0.27)	0.90 (0.22)	1.58 (0.70)	0.59 (0.28)	2.03*** (0.56)	2.28** (0.39)
Unknown	1.65 (0.46)	0.51 (0.20)	1.26* (0.25)	0.95 (0.45)	0.92 (0.73)	0.96 (0.86)	0.97 (0.33)	1.04 (0.48)
Gender (reference = male)								
Female	0.71*** (0.06)	0.99 (0.12)	0.79*** (0.05)	0.82 (0.10)	0.74 (0.18)	1.07 (0.27)	0.95 (0.10)	0.77* (0.10)
Non-binary	0.89 (0.48)	1.27 (1.05)	1.01 (0.45)	0.27 (0.32)	–	–	1.60 (1.37)	0.49 (0.70)
Frequency of internet use per day (reference = less than once a day)								
About once	1.36 (0.38)	0.76 (0.23)	1.25 (0.25)	1.82 (0.79)	2.61 (1.91)	1.89 (1.43)	2.34** (0.71)	5.02** (2.49)
Several times a day	1.96** (0.47)	0.89 (0.23)	1.64** (0.27)	2.24* (0.84)	4.12 (2.63)	0.89 (0.56)	2.42*** (0.59)	5.66*** (2.45)
Almost always	2.46*** (0.61)	0.81 (0.22)	1.90*** (0.33)	2.35* (0.91)	8.15** (5.53)	0.74 (0.48)	2.58*** (0.67)	6.62*** (2.95)
Amount of daily social media use (reference = Not at all)								
< 30 minutes	1.14 (0.13)	1.23 (0.19)	1.42*** (0.12)	1.03 (0.17)	1.04 (0.33)	1.22 (0.43)	1.33* (0.19)	1.35 (0.24)
30 to 59 minutes	1.29* (0.16)	0.96 (0.18)	1.48*** (0.14)	0.72 (0.14)	0.98 (0.36)	1.25 (0.52)	1.23 (0.20)	1.05 (0.21)
1 to less than 2 hours	1.12 (0.15)	1.05 (0.20)	1.31** (0.14)	0.66* (0.20)	1.45 (0.65)	1.35 (0.54)	1.06 (0.18)	0.90 (0.20)
2 to 4 hours	1.40* (0.20)	1.03 (0.22)	1.43** (0.16)	1.02 (0.22)	1.32 (0.57)	1.66 (0.74)	1.25 (0.23)	1.28 (0.31)
Over 4 hours	0.97 (0.17)	1.02 (0.26)	1.40* (0.18)	1.01 (0.29)	0.40 (0.19)	1.34 (0.74)	1.05 (0.24)	0.84 (0.26)

Table B1 (continued)

Variable	M1: General willingness		M2: Consent	M3: Appointment	M4: Completed	Add. M1: Substituted	Add. M2: Invited	Add. M3: Respondents
Category	Yes	Maybe	Yes	Yes	Yes	Substitute	Video	Video
Reference	No	No	No	No	No	Substituted	Online	Online
Estimate	*RR (SE)*	*RR (SE)*	*OR (SE)*	*OR (SE)*	*OR (SE)*	*OR (SE)*	*OR (SE)*	*OR (SE)*
Completion rate								
× 100	1.01** (0.00)	1.01** (0.01)	1.01** (0.01)	1.03*** (0.00)	1.03** (0.01)	0.97*** (0.01)	1.02*** (0.00)	1.03*** (0.01)
Recruitment year (reference = 2016)								
2018	1.83* (0.47)	2.14* (0.77)	1.18 (0.25)	0.53 (0.23)	–	0.12 (0.16)	1.54 (0.62)	1.07 (0.53)
2019	0.85 (0.09)	1.07 (0.16)	0.96 (0.08)	0.85 (0.15)	–	0.64 (0.23)	0.97 (0.14)	0.88 (0.17)
2020	1.24 (0.24)	0.90 (0.28)	0.10 (0.12)	0.89 (0.22)	–	1.20 (0.60)	0.93 (0.19)	0.82 (0.22)
2021	0.94 (0.09)	1.04 (0.16)	1.08 (0.08)	0.77 (0.11)	–	0.73 (0.21)	1.10 (0.13)	1.03 (0.16)
State (reference = New South Wales)								
Victoria	0.89 (0.09)	0.96 (0.14)	0.89 (0.07)	1.34 (0.20)	0.84 (0.26)	1.08 (0.34)	0.97 (0.12)	1.10 (0.18)
Queensland	0.96 (0.11)	1.25 (0.20)	0.86 (0.07)	1.05 (0.17)	0.35** (0.11)	1.03 (0.36)	1.00 (0.14)	0.88 (0.17)
South Australia	1.09 (0.16)	1.37 (0.28)	1.10 (0.12)	1.04 (0.24)	0.61 (0.27)	1.07 (0.57)	1.10 (0.22)	1.15 (0.28)
Western Australia	0.85 (0.12)	1.02 (0.21)	1.06 (0.11)	1.19 (0.25)	0.62 (0.25)	1.09 (0.49)	1.05 (0.19)	1.04 (0.24)
Tasmania	0.94 (0.25)	1.00 (0.39)	0.94 (0.19)	0.99 (0.41)	0.91 (0.80)	1.28 (0.91)	0.77 (0.26)	0.64 (0.28)
Northern Territory	1.18 (0.55)	0.92 (0.75)	1.15 (0.43)	1.01 (0.66)	0.29 (0.32)	1.63 (2.58)	2.38 (1.60)	3.43 (3.76)
Australian Capital	1.19 (0.29)	1.54 (0.55)	1.06 (0.20)	0.74 (0.26)	2.27 (2.45)	1.26 (0.82)	0.89 (0.27)	0.83 (0.33)
Born overseas (reference = born in Australia)								
Born overseas	0.72*** (0.06)	1.06 (0.13)	0.76*** (0.05)	0.90 (0.12)	0.94 (0.24)	1.07 (0.30)	0.84 (0.09)	0.86 (0.12)

Table B1 (continued)

Variable	M1: General willingness		M2: Consent	M3: Appointment	M4: Completed	Add. M1: Substituted	Add. M2: Invited	Add. M3: Respondents
Category	Yes	Maybe	Yes	Yes	Yes	Substitute	Video	Video
Reference	No	No	No	No	No	Substituted	Online	Online
Estimate	*RR (SE)*	*RR (SE)*	*OR (SE)*	*OR (SE)*	*OR (SE)*	*OR (SE)*	*OR (SE)*	*OR (SE)*
Indigenous status (reference = non-Indigenous)								
Indigenous	1.32 (0.43)	1.40 (0.64)	1.17 (0.28)	0.80 (0.36)	0.35 (0.26)	1.16 (1.29)	2.42 (1.20)	2.82 (2.00)
Intercept	0.08*** (0.04)	0.06*** (0.03)	0.09*** (0.03)	0.04*** (0.03)	0.05* (0.06)	20.50 (26.86)	0.04*** (0.02)	0.00*** (0.00)
N	3,420		5,523	1,388	710	342	2,179	1,154
X^2	270.27		233.87	127.11	69.23	60.78	139.10	139.10
D.F.	76		38	38	33	37	38	38
p	< .001		< .001	< .001	< .001	.008	< .001	< .001
McFadden R^2	0.04		0.03	0.07	0.11	0.12	0.10	0.09
Log-likelihood	−3,232.67		−3,641.71	−898.17	−278.22	−224.24	−1,280.10	−745.45

Notes: Odds ratios (OR), relative risk ratios (RR), and standard errors (SE).
*$p < .05$; **$p < .01$; ***$p < .001$.

methods, data, analyses | Vol. 20(1), 2026, pp. 302-316 DOI: 10.25521/mda.581

How Does Video Interviewing Affect the Interviewers' and Respondents' Paralinguistic Behaviors? A First Exploration

Hanyu Sun[1] & Frederick G. Conrad[2]

[1] *Westat*
[2] *University of Michigan*

Abstract

Video interviews have been gradually adopted by survey organizations as an alternative to in-person interviews as a mode of data collection. Recent studies have shown that video and in-person interviews elicited similar levels of respondents' rapport with interviewers and similar quality data. However, little is known about whether the presence and prevalence of interviewer and respondent paralinguistic behaviors (e.g., disfluencies such as "*uh*" and "*um*" or laughter) vary between the two modes, and when they do, how might this affect survey outcomes? To address these questions, we coded the presence of six paralinguistic behaviors in 710 question-answer (Q-A) sequences in 15 in-person interviews and 12 video interviews conducted by professional survey interviewers in a laboratory experiment. Most of the paralinguistic behaviors occurred equally often in the two interviewing modes except laughter which was significantly more prevalent in video than in-person interviews. We attributed the increased laughter in video interviews as a nervous response to greater communication difficulties in that mode. Nonetheless, this did not differentially impact the prevalence of respondents' adequate responses, indicating (indirectly) that data quality was equivalent in the two modes. These findings bolster the emerging narrative that when interviewed via video, respondents' experience and their answers are very similar to when they are interviewed in person.

Keywords: live video interviews, in-person interviews, paralinguistic behaviors

© The Author(s) 2025. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Video interviewing has been increasingly adopted by survey organizations as the underlying technology has become more mature and widely available, and because the pandemic drove organizations to find alternative ways to safely collect data requiring face-to-face interaction. In both in-person and video interviews, interviewers and respondents can see and hear each other in real time, enabling interviewers to rely on visual information about respondents, such as their facial expressions, and to provide visual materials to respondents, such as showcards or standardized test materials. It is plausible that reading respondents' facial expressions and presenting visual material to them is generally no more difficult and sometimes easier via video than in person. For instance, a recent study showed that interviewers exhibited similar question reading behaviors in both in-person and video interviews and respondents provided a higher percentage of responses that matched one of the provided response categories in video than in-person interviews (Kokoska et al., 2024). Moreover, respondents in video interviews and their counterparts in in-person interviews reported similar levels of rapport with interviewers, disclosed similar amounts of sensitive information (Sun et al., 2020), exhibited similar levels of item nonresponse and straightlining, provided answers of similar length to open-ended questions, and had similar interview experiences (Endres et al., 2022). However, little is known about how communication between interviewers and respondents might be affected by video mediation and how this might impact the respondent's answers to a variety of questions.

Paralinguistic Behaviors in Spontaneous Speech

An indication of how well two conversational partners, including a survey interviewer and a respondent, are communicating can sometimes be found in their paralinguistic utterances. These are generally (but not always) non-lexical parts of speech that can serve as cues about the speaker's current state of mind including their understanding of the speaker and their affective state or mood (e.g., Brown, 1977; Hancock, 2004). Speakers produce *disfluencies*—"um" and "uh", hesitations or pauses, and repairs—to indicate planning activities or difficulty in word finding (Clarck & Fox Tree, 2002; Conrad et al., 2013). In spontaneous speech, speakers use "um" and "uh" to signal the start of expected minor or major delays (Clark & Fox Tree, 2002). Speakers can use these utterances to suggest they are searching for the right word, formulating an accurate description, preparing to repair what they just said, and deciding what to say next (Clark & Fox Tree, 2002). *Pauses* are periods of silences in speech and may signal that the speaker is

Direct correspondence to
Hanyu Sun, Westat, Rockville, MD, USA
E-mail: hanyusun@westat.com

organizing their thoughts or experiencing uncertainty (Fox Tree, 2002). Pauses are used for holding the floor or maintaining the turn (Maclay & Osgood, 1959). In replying to what the speaker has said, the addressee may start their reply with "um" or a pause, possibly indicating difficulty formulating a reply or that they are not comfortable with the topics under discussion (Fox Tree, 2010). *Repairs*—spontaneous corrections to an immediately prior utterance—occur when the speaker realizes they have made an error (Clark, 1994), for instance, "2 to 3 times a week... I mean frequently". In addition, speakers repair their speech to reduce ambiguity (Levelt, 1983). In everyday conversation, the speaker and the listener finely coordinate turn-taking (Sacks et al., 1974). *Overlapping* speech (i.e., where two speakers talk simultaneously) is often considered a breakdown in turn-taking. A conversation with large amount of overlapping speech often signals troubled communication (e.g., Leighton et al., 1971). Overall, the presence of disfluencies, pauses, repairs, and overlapping speech suggests that the speaker is experiencing communication difficulties to some extent.

Laughter alternates with speech to regulate the flow of the conversation (Jefferson, 1979). Not all laughter is an expression of amusement; laughter with high vowel resonances could reflect nervousness and submissiveness, while laughter with low vowel resonances could reflect coarseness and dominance (Kohler, 2008). Laughter by a first speaker may serve as an invitation to the addressee to laugh, resulting in shared laughter (Jefferson, 1979). Shared laughter has been linked to ending a topic, while declining to engage in reciprocal laughter rejects the proposed topic termination, leading to continued conversation on the topic (Holt, 2010). But laughter is not always shared. There are circumstances where it is inappropriate for the listener to accept the laughter invitation, for instance, when the speaker is talking about their troubles, the "troubles-teller" often produces an utterance and then laughs, but the "troubles-recipient" typically does not join in with the laughter (Jefferson, 1984). Laughter and smile can be considered as two extremes in the same continuum (Fried et al., 1988). The listener response smile seems to be equivalent to verbal acknowledgement like "mm-hmm" and "good", and is often accompanied by head nodding (Ekman, 1985). *Backchannels* (e.g., "uh huh" and "mmm") are thought to signal that the listener is attending to the speaker (Brunner 1979; Clark & Schaeffer, 1989) and possibly agrees with the speaker's message, consenting to the speaker's continuing to speak (Conrad et al., 2013).

Paralinguistic Behaviors in Standardized Interviews

Standardized interviews are different from everyday conversations. In standardized interviews, the interviewer directs the flow, and the interviewer-respondent interaction are expected to follow the paradigmatic question-answer sequence in which the interviewer asks a question and then the respondent selects a response option from those provided by the interviewer (Schaeffer and Maynard, 1996). For respondents, their sole task is to answer the interviewer's question. Interviewers, however, must keep the respondent engaged and ensure the collection of high-quality response data. To explore interaction in telephone and in-person survey interviews, researchers have studied the paralinguistic utterances of interviewers and respondents, especially disfluencies, hesitation, and pauses (e.g., Conrad et al., 2008; Conrad & Schober, 2021; Garbarski et al., 2016; Schober & Bloom, 2004; Schober et al., 2012). The presence of such behaviors can be used to flag respondent difficulty and, thus, increased likelihood of lower quality responses (e.g., Schaeffer & Dykema, 2011; Schober et al., 2012; Min et al., 2020). For instance, previous research has shown that respondents' disfluent answers—turns that include "um" and "uh", pauses, or repairs and restarts—are more likely to be unreliable than their fluent answers (Schober et al., 2012). Interviewers' disfluencies can affect survey outcomes by encouraging or discouraging sample member's decision to participate in a telephone interview (e.g., Conrad et al., 2013; Min et al., 2020). An interviewer who speaks without disfluencies may sound robotic to the respondent, giving the impression that they are only focused on completing the interviewing task rather than engaging with the respondent in a meaningful way. To the extent that this is detected by respondents, it can potentially reduce willingness to participate and if they do participate, their conscientiousness leading to reduced response quality. In addition, how the interviewer reacts to the respondent's "laughter invitation" may also affect the course of specific question-answer sequences (dialogue) and the overall interviewing experience. For instance, a telephone interviewer might use "smile voice" or quasi-laughter to create a sense of rapport with the respondent in the context of standardized interviewing (e.g., Lavin & Maynard, 2001). In addition, the more time interviewers spend smiling, the more likely the respondents were to provide an adequate response (Welles et al., 2022). (Adequate response is an indirect measure of data quality. A response is adequate if it matches one of the response categories provided in the survey question [Welles et al., 2022]).

Paralinguistic Behaviors in In-Person and Video Interviews

Video interviews may differ from in-person interviews in subtle ways due to the mediated nature of interaction. Because the interviewer and the respondent are in separate physical spaces, it is possible that they are less attuned to subtle social cues such as body language and nonverbal behaviors. At the same time, the increased sense of distance and privacy may make some respondents feel more comfortable sharing sensitive information. To the best of our knowledge, no research has been conducted comparing the prevalence of paralinguistic behaviors in in-person and video interviews. This paper explores that topic. In this study, we used the term “video interview” to refer to real-time, synchronized, two-way communication, mediated by video, distinguishing it from prerecorded interviews where a video file of the interviewer reading the question is embedded in a web survey. Specifically, the current study asks the following research questions:

RQ1: *Does the prevalence of paralinguistic behaviors vary between in-person and video interviews?*

RQ2: *Does the prevalence of paralinguistic behaviors vary between in person and video interviews differently for interviewers versus respondents?*

RQ3: *Does data quality differ between modes in a corresponding way?*

Data and Method

To address these research questions, we examined the presence of paralinguistic behaviors in 15 in-person interviews and 12 video interviews in a laboratory experiment carried out in 2014. Five, professional female interviewers at a university research center administered standardized interviews in both modes (see Sun et al., 2020 for more information); respondents (14 females and 13 males, mean age of 46 [*SD* = 12], 24 of 27 held bachelor’s degrees) were full-time staff employees of the University of Michigan, recruited via email and on-campus flyers, and randomly assigned to one of the two modes at the time of recruitment. Upon completion, they received a $15 token of thanks for their participation. The questions were on various topics including dietary behaviors, mental health, and sexual behaviors, adapted from existing national surveys such as the National Survey of Family Growth. The average interview length was about 35 minutes. The in-person interviews were conducted in an office, where the interviewer sat across a table from the respondent. The interviewer read the questions aloud from a laptop screen, and the respondent replied orally. The video interviews were conducted via Adobe Connect (www.adobe.com/products/adobeconnect.html). Both the interviewer and the respondent connected to the video call using the desktop computer. For the video respondents, most of their window was filled with the interviewer’s video. The respondent’s self-view video thumbnail appeared in the upper right corner. For the interviewer, most of their window was filled with the respondent’s video. Interviewer’s self-view video thumbnail appeared in the right upper corner. The interviewer shared their screen with the respondent to display showcards as needed.

The dialog between interviewers and respondents was audio-recorded in both in-person and video interviews. To transform the dialog into data that are amenable to quantitative analysis, we developed a coding scheme and then coded (in Sequence Viewer[1] [http://www.sequenceviewer.nl/]) the presence (Yes/No) of six paralinguistic behaviors for the interviewers and six for the respondents within each Q-A sequence (see Table 1); the behaviors included disfluencies, pauses, repairs, overlapping speech, laughter, and backchannels. In this study, disfluencies include fillers (e.g., "uh" and "um"). Laughter was coded as either interviewer-initiated or respondent-initiated during a Q-A sequence. Regarding overlapping speech, if the interviewer interrupted or cut off the respondent, we coded it as interviewer-initiated overlapping speech; if the respondent interrupted or cut off the interviewer's speech, we coded it as respondent-initiated overlapping speech. We coded the presence of interviewer and respondent paralinguistic behaviors in each of 710 audio-recorded, question-answer (Q-A) sequences (i.e., for each question, we coded each conversational turn between the start of the question's delivery [by the interviewer] and the start of the next question's delivery). Two coders independently classified the interviewers' and respondents' behaviors; Cohen's kappa = .75 was "substantial" (McHugh, 2012).

Table 1 Definition of the paralinguistic behaviors

Paralinguistic behaviors	Definition
Disfluencies	Fillers (e.g., "um" and "uh")
Pauses	Periods of silence during a speaking turn
Repairs	Spontaneous corrections to an immediately prior utterance
Overlapping speech	Two speakers talk simultaneously
Laughter	Interviewer-initiated or respondent-initiated laughter during a Q-A sequence
Backchannels	Utterances such as, "uh huh" and "mmm", to signal that the listener is attending to the speaker

Once the interactions were coded, we compared the presence of paralinguistic behaviors in the two modes. We first examined the presence of each paralinguistic behaviors by mode overall and separately for interviewer and respondent. We also computed dichotomous variables to measure communication difficulties overall and separately for interviewers and respondents. If a Q-A sequence included any of the paralinguistic behaviors—disfluencies, repairs, pauses, or overlapping speech—it is considered to exhibit communication difficulties. In the analysis, we accounted for the fact that interviewer and respondent utterances are clustered in an interview and used Wald chi-square tests for independence to compare the percentage of the presence of each paralinguistic behavior between the in-person and video interviews. We used PROC SURVEYFREQ in SAS 9.4 for the analysis and specified interviews (i.e., case id) as clusters.

[1] We used Sequence Viewer 6.0d (http://www.sequenceviewer.nl/).

Because of how they were recruited, the participants comprised a convenience sample and thus the results are unweighted. We also adjusted for multiple comparisons using the Benjamini-Hochberg method (1995).

In addition, we fitted a crossed random effects logistic regression model to explore the effects of paralinguistic behaviors on the respondent's adequate response (Yes or No). A response is adequate if it matches one of the response categories provided in the survey question. As described earlier, the data includes 710 audio-recorded, question-answer (Q-A) sequences in total, combined across 15 in-person interviews and 12 video interviews conducted by 5 interviewers. The mean number of turns for each Q-A sequence for in-person and video interviews was 3.2 ($SD = 1.8$) and 3.6 ($SD = 2.6$), respectively. We initially fitted a two-level cross-classified random effects regression model with adequate response cross-classified by respondents and questions, nested within interviewers. However, the variance associated with interviewer was estimated to be zero in the unconditional model and therefore we simplified the statistical approach by fitting crossed random effects logistic regression models. The general model specification is as below:

$$\begin{aligned} logit\,(\Pr(Y_{ij} = 1)) \\ &= ß_0 + ß_1 \times Mode + ß_2 \times Interviewer\ Paralinguistics \\ &+ ß_3 \times Respondent\ Paralinguistics + u_i + v_j + \varepsilon_{ij} \\ &u_i \sim N(0, \sigma_i^2),\ v_j \sim N(0, \sigma_j^2),\ \varepsilon_{ij} \sim N(0, \sigma^2) \end{aligned}$$

where $Y_{ij} = 1$ indicates adequate response for question i by respondent j, u_i represents the random effects associated with question i, v_j represents the random effects associated with respondent j, and ε_{ij} represents the residuals. The models are estimated using the default residual pseudo-likelihood (RSPL) in SAS 9.4 PROC MIXED with random intercepts for questions and respondents.

Results

RQ1 explores whether the prevalence of paralinguistic behaviors vary between in-person and video interviews. To address RQ1, we first examined the overall presence of paralinguistic behaviors by mode. As shown in Table 2, the overall prevalence of paralinguistic behaviors was at or below 30% in both in-person and video interviews. For instance, disfluencies occurred in 30.3% of speaking turns in in-person interviews and in 22.8% in video interviews, while laughter occurred in only of 2.8% of speaking turns in in-person and 13.8% in video interviews. Disfluencies and backchannels occurred significantly less frequently in video than in-person interviews while laughter and overlapping speech occurred significantly more often in video than in in-person interviews.

Table 2 The overall presence of paralinguistic behaviors by mode (by either the interviewer or the respondent)

Paralinguistic behaviors	In-person (%)		Video (%)		Wald chi-square test statistic[a]
	%	n[b]	%	n[b]	
Disfluencies	30.3	121	22.8	71	$F(1,26) = 1.68, p = .02$
Pauses	11.5	46	8.7	27	$F(1,26) = 0.50, p = .84$
Repairs	4.8	19	10.9	34	$F(1,26) = 9.79, p = .21$
Overlapping speech	10.5	42	15.8	49	$F(1,26) = 2.11, p = .01$
Laughter	2.8	11	13.8	43	$F(1,26) = 12.47, p = .01$
Backchannels	19.3	77	18.3	57	$F(1,26) = 0.03, p = .01$

Notes: Based on a total of 710 question-answer sequences irrespective of mode. The question-answer sequences were clustered by interview.
[a] Adjusted for multiple comparisons using the Benjamini-Hochberg method. The Wald Chi-square test statistics were from PROC SURVEYFREQ, specified interviews (i.e., case id) as clusters
[b] This column displays the number of times a paralinguistic behavior occurred.

RQ2 explores whether the prevalence of paralinguistic behaviors vary between in-person and video interviews differently for interviewers versus respondents. To address RQ2, we examined the presence of paralinguistic behaviors separately for interviewers and respondents. As shown in Figure 1a and 1b, there were significantly higher percentages of interviewer laughter ($F(1,26) = 12.93$, $p = .01$) and respondent laughter ($F(1,26) = 9.47$, $p = .03$) in video than in-person interviews. However, there were no significant differences between the two modes for other paralinguistic behaviors. As stated earlier, if the Q-A sequence included any of the paralinguistic behaviors—disfluencies, pauses, repairs, or overlapping speech—it is considered to exhibit communication difficulties. As shown in Figure 1a and 1b, interviewers experienced marginally more communication difficulties in video than in-person interviews ($F(1,26) = 3.30$, $p = .08$). Respondents experienced communication difficulties to the same extent in video and in-person interviews ($F(1,26) = 0.78$, $p = .38$).

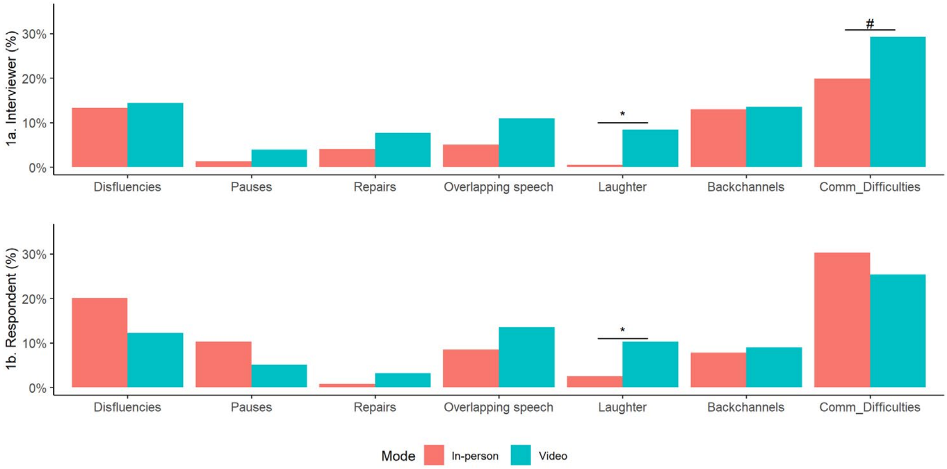


Figure 1 The presence of paralinguistic behavior by mode, separately for interviewers and respondents. Wald chi-square test statistics, specified interviews as clusters. Adjusted for multiple comparisons using the Benjamini-Hochberg method. * $p < .05$, # $p < .10$.

As noted earlier, laughter can serve different purposes in speech—it might be an expression of nervousness, or it could be someone laughing off their own troubles during a conversation. To better understand what drives the differences, we compared the prevalence of interviewer and respondent laughter when there were communication difficulties (either from the interviewer or the respondent) and when there were no difficulties for in-person and video interviews.

Figure 2 presents the prevalence of laughter by any communication difficulties in the Q-A sequence by mode. As shown in Figure 2a, no laugher was observed in the absence of communication difficulties in in-person interviews. It was only when there were communication difficulties that there was interviewer laughter and respondent laughter in in-person interviews. The percentage of respondent laughter was higher than the percentage of interviewer laughter in in-person interviews. As shown in Figure 2b, the percentage of interviewer laughter and the percentage of respondent laughter were higher when there were communication difficulties as compared to when there were no difficulties in video interviews. The prevalence of laughter was much lower in in-person than video interviews. For instance, as shown in Figure 2a, 6.1% of the Q-A sequences contain respondent laughter where there were no communication difficulties in video interviews as compared to 0% of respondent laughter when there were no communication difficulties in in-person interviews (a condition that was never actually observed as all sequences in this mode contained communication difficulty either from the interviewer or the respondent). The percentage of respondent laughter increased to 16.8% when there were communication difficulties in video interviews as shown in Figure 2b. The findings are consistent when separately examining the interviewer and respondent communication difficulties (see Figures A1 and A2 in the Appendix).

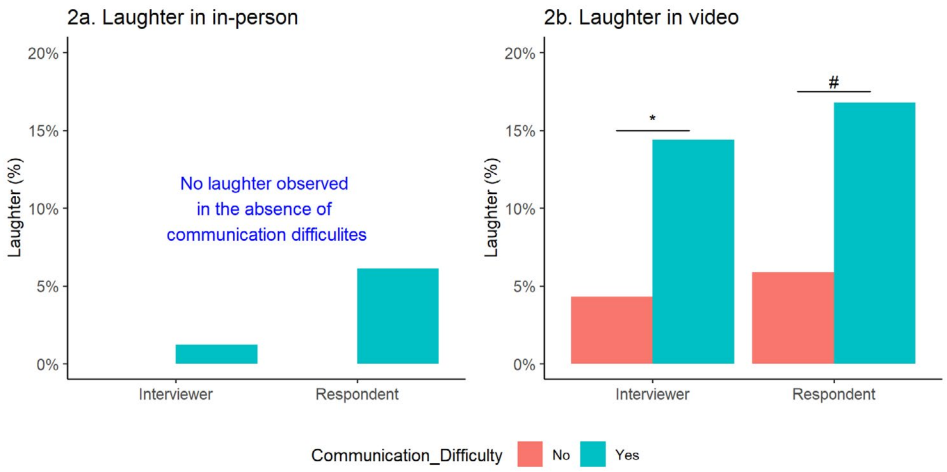


Figure 2 The prevalence of laughter by communication difficulties (by either the interviewer or the respondent) by mode. Adjusted for multiple comparisons using the Benjamini-Hochberg method. Chi-square tests cannot be computed if at least one table cell has 0 frequency. * $p < .05$, # $p < .10$.

RQ3 explores whether the presence of paralinguistic behaviors affect data quality by mode. To address RQ3, we fitted a crossed random effects logistic regression model to predict adequate response—an indirect measure of data quality—by the paralinguistic behaviors (present or absent) and mode (in-person or video). As Shown in Table 3, we found that respondents were less likely to provide an adequate response when interviewers repaired their speech (e.g., "How long has it been since you last used any prescription transit uh tranquilizer..."), when respondents were disfluent, and when respondents interrupted the interviewer, after controlling for the other paralinguistic behaviors and mode. But there were no significant two-way interactions between any of the paralinguistic behaviors and mode on the prevalence of adequate responses (results not shown). As a result, although we found significant differences between laughter in in-person and video interviews, this did not differentially impact the prevalence of respondents' adequate responses (i.e., was not related to [our measure of] data quality).

Table 3 Crossed random effects logistic regressions predicting adequate response by the presence of paralinguistic behaviors and mode

Parameters	*b (SE)*	
Intercept	3.58***	(0.36)
Video interview (ref: In-person)	−0.09	(0.39)
Interviewer paralinguistic behaviors		
Disfluencies (ref: Not present)	0.72	(0.54)
Repairs (ref: Not present)	−1.10#	(0.59)
Pauses (ref: Not present)	1.14	(1.14)
Overlapping speech (ref: Not present)	0.74	(0.65)
Laughter (ref: Not present)	0.19	(0.87)
Backchannels (ref: Not present)	−0.21	(0.43)
Respondent paralinguistic behaviors		
Disfluencies (ref: Not present)	−1.60***	(0.38)
Repairs (ref: Not present)	−1.26	(0.87)
Pauses (ref: Not present)	0.06	(0.53)
Overlapping speech (ref: Not present)	−1.47**	(0.48)
Laughter (ref: Not present)	−0.37	(0.59)
Backchannels (ref: Not present)	−0.29	(0.53)
$\sigma^2_{int:question}$	0.89	(0.43)
$\sigma^2_{int:respondent}$	0.22	(0.27)
n	710	

Notes: ref = reference. *** $p < .001$, ** $p < .01$, * $p < .05$, # $p < .10$.

Discussion and Conclusion

In this study, we compared the presence of interviewer and respondent paralinguistic behaviors between in-person and video interviews. One paralinguistic behavior differed between modes: Interviewers and respondents laughed significantly more in video than in in-person interviews. Although we did not observe any mode differences for the other paralinguistic behaviors, the presence of laughter was correlated with the presence of communication difficulties. One interpretation of this finding is that laughter was functioning as a "social lubricant" to keep the conversation flowing when it was disrupted by these behaviors.

The study was conducted in 2014, a time when video as a mode of data collection was new to interviewers and respondents. While it is possible that technical issues such as bandwidth limitation, lags in transmission, and poor audio/video quality may have led to communication difficulties, we are not aware of any technical problems having occurred during the interviews. It is plausible that video requires participants to split their attention between planning what to say and monitoring activities across multiple views (i.e., the interviewer needed to monitor the respondent's video and survey instrument, the respondent may have monitored their self-view video thumbnail), therefore increasing the prevalence of paralinguistic behaviors that indicate communication difficulties. The one mode effect on paralinguistic behavior (i.e., laugher) observed in the current study was small but persistent when the data were collected in 2014, and video was relatively new. This effect of mode on paralinguistic behavior is likely to be even smaller now that video is common in everyday communication. Future research may explore this by conducting similar studies to test mode differences between in-person and video interviews, since video technologies are now widespread and familiar to much of the public. In a more contemporary test, the researcher would be wise to recruit a larger sample of participants from a more general population.

In addition, we found that the presence of interviewer repairs, respondent disfluencies, and respondent-initiated overlapping speech affected adequate response. However, there were no significant two-way interactions between any of the paralinguistic behaviors and mode on adequate response. The very similar patterns of interaction in video and in-person interviews are consistent with several findings in the literature indicating very similar respondent experiences and data quality in video and in-person interviews. As previously indicated, Sun et al. (2020) reported similar levels of respondents' rated rapport with the interviewer and Endres et al. (2022) reported similar quality data in the two modes, in particular equivalent length of open responses, straightlining, item nonresponse, and social desirability bias.

The study has some limitations. When the study was conducted (2014), video was still a relatively new method of data collection for both interviewers and respondents. The communication difficulties caused by technical issues have likely been greatly reduced today due to improved Internet connectivity (at least in urban areas) and widespread familiarity with video conferencing platforms. It was difficult to assess the impact of technical issues on communication difficulties arising from the mode difference between in-person and video interviews because we only had audio recordings of the interviews. Technical issues were a known aspect of video interviews in 2014. However, recent studies have found that respondent experiences and data quality are similar between video and in-person interviews (e.g., Endres et al., 2022). This has led us to believe that part of the laughter we observed in 2014 may have been due to technical issues more than mode differences. Future research should tease apart technical issues from communication difficulties in testing mode effects on paralinguistic behaviors. For instance, developing a coding scheme to identify technical issues and then comparing Q-A sequences with and without these difficulties could help answer how paralinguistic behaviors vary by mode. To the extent that communication difficulties appear when there are no technical difficulties, this might suggest benefit in training interviewers on how to best multitask while maintaining respondent engagement in video interviews.

Nevertheless, these findings strongly argue for using video to conduct survey interviews when the measurement goals of the research require in-person data collection. Although we observed slightly different interaction patterns in video than in-person interviews, the adequacy of the answers in the two modes was indistinguishable.

References

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B: Statistical Methodology, 57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Brown, G. (1977). *Listening to spoken English* (2nd ed.). Routledge. https://doi.org/10.4324/9781315538518

Brunner, L. J. (1979). Smiles can be back channels. *Journal of Personality and Social Psychology, 37*(5), 728–734. https://doi.org/10.1037/0022-3514.37.5.728

Clark, H. H. (1994). Managing problems in speaking. *Speech Communication, 15*(3–4), 243–250. https://doi.org/10.1016/0167-6393(94)90075-2

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition, 84*(1), 73–111. https://doi.org/10.1016/S0010-0277(02)00017-3

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science, 13*(2), 259–294. https://doi.org/10.1207/s15516709cog1302_7

Conrad, F. G., Broome, J. S., Benkí, J. R., Kreuter, F., Groves, R. M., Vannette, D., & McClain, C. (2013). Interviewer speech and the success of survey invitations. *Journal of the Royal Statistical Society Series A: Statistics in Society, 176*(1), 191–210. https://doi.org/10.1111/j.1467-985X.2012.01064.x

Conrad, F. G., & Schober, M. F. (2021). Clarifying question meaning in standardized interviews can improve data quality even though wording may change: A review of the evidence. *International Journal of Social Research Methodology, 24*(2), 203–226. https://doi.org/10.1080/13645579.2020.1824627

Conrad, F. G., Schober, M., & Dijkstra, W. (2008). Cues of communication difficulty in telephone interviews. In J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. de Leeuw, L. Japec, P. J. Lavrakas, M. W., Link, & R. L. Sangster (Eds.), *Advances in telephone survey methodology* (pp. 212–230). John Wiley & Sons. https://doi.org/10.1002/9780470173404.ch10

Eckman, P. (1985). *Telling lies: Clues to deceit in the marketplace, politics, and marriage*. Norton.

Endres, K., Hillygus, D. S., DeBell, M., & Iyengar, S. (2022). A randomized experiment evaluating survey mode effects for video interviewing. *Political Science Research and Methods, 11*(1), 144–159. https://doi.org/10.1017/psrm.2022.30

Fried, I., Wilson, C. L., MacDonald, K. A., & Behnke, E. J. (1998). Electric current stimulates laughter. *Nature, 391,* 650. https://doi.org/10.1038/35536

Fox Tree, J. E. (2002). Interpreting pauses and ums at turn exchanges. *Discourse Processes, 34*(1), 37–55. https://doi.org/10.1207/S15326950DP3401_2

Garbarski, D., Schaeffer, N. C., & Dykema, J. (2016). Interviewing practices, conversational practices, and rapport: Responsiveness and engagement in the standardized survey interview. *Sociological Methodology, 46(*1)*,* 1–38. https://doi.org/10.1177/0081175016637890

Hancock, J. T. (2004). Verbal irony use in face-to-face and computer-mediated conversations. *Journal of Language and Social Psychology, 23*(4)*,* 447–463. https://doi.org/10.1177/0261927X04269587

Holt, E. (2010). The last laugh: Shared laughter and topic termination. *Journal of Pragmatics, 42*(6)*,* 1513–1525. https://doi.org/10.1016/j.pragma.2010.01.011

Jefferson, G. (1979). A technique for inviting laughter and its subsequent acceptance/declination. In G. Psathas (Ed.), *Everyday language: Studies in ethnomethodology* (pp. 79–96). Irvington.

Kohler, K. J. (2008). "Speech-smile", "speech-laugh", "laughter" and their sequencing in dialogic interaction. *Phonetica, 65*(1–2)*,* 1–18. https://doi.org/10.1159/000130013

Kokosak A., Sun, H., & Yan, T. (2024, May 15–17). *How do behavior coding results differ by mode in a multimode survey?* [Paper presentation]. 79th Annual Conference of the American Association for Public Opinion Research (AAPOR), Atlanta, GA, United States. https://aapor.confex.com/aapor/2024/meetingapp.cgi/Paper/2649

Lavin, D., & Maynard, D. W. (2001). Standardization vs. rapport: Respondent laughter and interviewer reaction during telephone surveys. *American Sociological Review, 66*(3), 453–479. https://doi.org/10.2307/3088888

Leighton, L. A., Stollak, G. E., & Ferguson, L. R. (1971). Patterns of communication in normal and clinic families. *Journal of Consulting and Clinical Psychology, 36*(2), 252–256. https://doi.org/10.1037/h0030759

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition, 14*(1)*,* 41–104. https://doi.org/10.1016/0010-0277(83)90026-4

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word, 15*(1)*,* 19–44. https://doi.org/10.1080/00437956.1959.11659682

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276–282. https://biochemia-medica.com/en/journal/22/3/10.11613/BM.2012.031

Min, B. H., Schaeffer, N. C., Garbarski, D., & Dykema, J. (2020). The action structure of recruitment calls and its analytic implications: The case of disfluencies. *Journal of Official Statistics, 36*(3), 529-559. https://doi.org/10.2478/jos-2020-0028

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*(4)*,* 696–735. https://doi.org/10.2307/412243

Schober, M. F., & Bloom, J. E. (2004). Discourse cues that respondents have misunderstood survey questions. *Discourse processes, 38*(3)*,* 287–308. https://doi.org/10.1207/s15326950dp3803_1

Schober, M. F., Conrad, F. G., Dijkstra, W., & Ongena, Y. P. (2012). Disfluencies and gaze aversion in unreliable responses to survey questions. *Journal of Official Statistics, 28*(4), 555–582. https://www.proquest.com/scholarly-journals/disfluencies-gaze-aversion-unreliable-responses/docview/1266766372/se-2?accountid=16283

Sun, H., Conrad, F. G., & Kreuter, F. (2020). The relationship between interviewer-respondent rapport and data quality. *Journal of Survey Statistics and Methodology, 9*(3), 429–448. https://doi.org/10.1093/jssam/smz043

Foucault Welles, B., Sun, H., & Miller, P. V. (2022). Nonverbal behavior in face-to-face survey interviews: An analysis of interviewer behavior and adequate responding. *Field Methods, 34*(1)*,* 52–68. https://doi.org/10.1177/1525822X221075305

Appendix

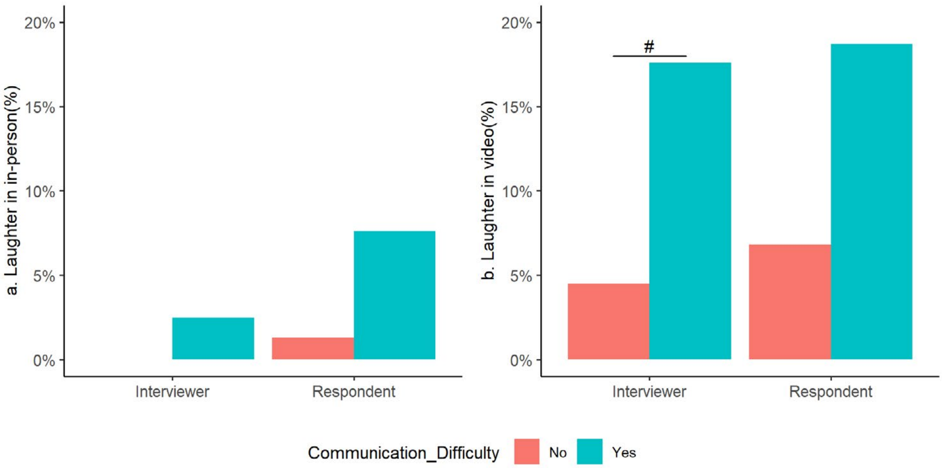


Figure A1 The prevalence of laughter by interviewer communication difficulties by mode. Adjusted for multiple comparisons using the Benjamini-Hochberg method. Chi-square tests cannot be computed if at least one table cell has 0 frequency. # $p < .10$.

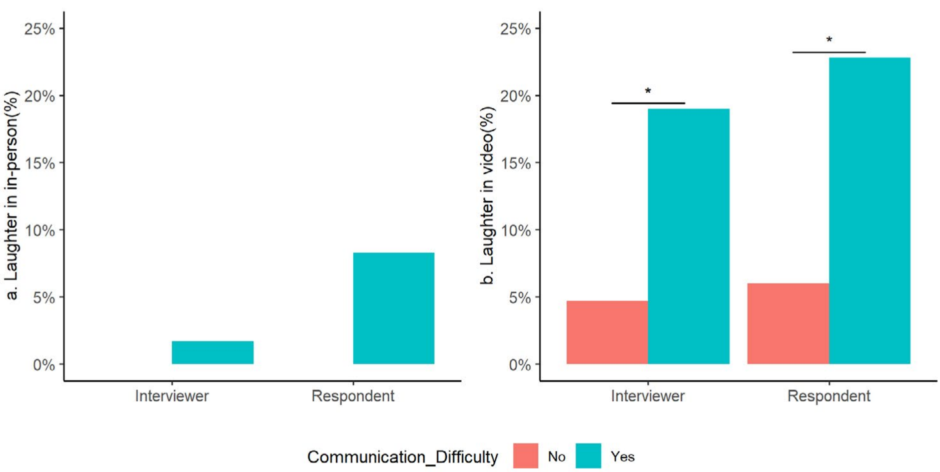


Figure A2 The prevalence of laughter by respondent communication difficulties by mode. Adjusted for multiple comparisons using the Benjamini-Hochberg method. * $p < .05$.

methods, data, analyses | Vol. 20(1), 2026, pp. 317-346 DOI: 10.12758/mda.2025.10

A Survey Mode of the Future? Investigating Respondents' Willingness to Participate in Self-Administered Video-Based Web Surveys

Joshua Claassen[1], Timo Lenzner[2], Jan Karem Höhne[1] & Conrad Ziller[3]

[1] *German Centre for Higher Education Research and Science Studies (DZHW), Leibniz University Hannover*
[2] *GESIS – Leibniz Institute for the Social Sciences*
[3] *University of Duisburg-Essen*

Abstract

Web surveys convince through low costs, save interviewer time, and provide respondents with a high level of flexibility, but often struggle with achieving high response rates and data quality. One way to mitigate these problems might be to conduct video-based web surveys, in which at least one of the two interview components (i.e., questions or answers) is based on video communication technology. However, few studies have investigated respondents' willingness to participate in such web surveys. Therefore, we examine the extent to which respondents are willing to participate in video-based web surveys and the reasons for their hypothetical (non)willingness. This also includes the investigation of characteristics of respondents and video-based web surveys that are associated with (non)willingness. The results of two experimental web surveys conducted in non-probability online panels in Germany reveal that a majority of respondents is willing to participate in web surveys with video-based questions and predefined answer options or open textual answers. Less respondents are willing to participate in web surveys with voice or video-based answers. While respondents mostly mention convenience and time flexibility as reasons for their willingness, the main reasons given for their nonwillingness are excessive effort regarding video-based questions and privacy and data security concerns regarding video-based answers. Video-based web surveys generally appeal to younger, male, and extraverted respondents as well as respondents who rate the survey as more interesting. This article provides researchers and practitioners with new insights into respondents' hypothetical willingness when it comes to video-based web surveys, including empirical-driven survey design recommendations.

Keywords: automatic question reading, pre-recorded interviewers, respondent willingness, self-recording answers, web surveys

© The Author(s) 2025. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

For a long-time, face-to-face interviews were described as the "gold-standard" of data collection in social science and market research because they are characterized by unique strengths, such as allowing interviewers to build rapport with respondents, engage in dynamic social interactions, and motivate respondents to put effort into question answering (Schober, 2018). In the past two decades, and particularly in the wake of the COVID-19 pandemic, web surveys have replaced face-to-face interviews as a predominant data collection method. Web surveys are associated with many advantages. For example, this includes the absence of constraints regarding survey location as well as significant time and cost savings (Callegaro et al., 2015). However, web surveys frequently struggle with achieving high response rates (Daikeler et al., 2020), collecting high-quality data (Silber et al., 2019), and including respondents with literacy difficulties (Bauer et al., 2025). One way to mitigate these problems might be to incorporate elements of face-to-face interviews into web surveys using video communication technology.

Specifically, this can be done by delivering (all) questions of a web survey via pre-recorded videos (Conrad et al., 2023) and/or having respondents answer via self-recorded videos (Höhne et al., 2023). We refer to this method as "video-based web survey" and define it as web surveys in which at least one of the interview components (i.e., questions or answers) is video-based. Importantly, video-based web surveys are self-administered and asynchronous. They are not to be confused with live video interviews, which are based on a synchronous and video-mediated interaction between interviewers and respondents (West et al., 2022). Compared to traditional web surveys and face-to-face interviews, video-based web surveys are promising in at least four respects: First, they resemble human-like conversations, possibly reducing respondent burden and improving overall data quality (Sun et al., 2021; West et al., 2022). Second, video-based questions and answers facilitate the participation for respondents with literacy difficulties (i.e., respondents who have difficulty reading survey questions or providing open textual answers), which can ultimately contribute to more inclusiveness in terms of web survey participation. Third, if the answer component is video-based, respondents' answers allow researchers to analyze rich tonal, facial, and gestural data to learn about data quality and answer behavior (e.g., respondents' interest and engagement; see Höhne, Kern, et al., 2024). Fourth, compared to interviewer-administered surveys (i.e., face-to-face, live video, and telephone interviews), the asynchronous communication associated with video-based web surveys provides respondents with more flexibility, while being cost- and time-efficient (e.g., no in-person contact approach is required).

Direct correspondence to
Joshua Claassen, German Centre for Higher Education Research and Science Studies (DZHW), Leibniz University Hannover, Lange Laube 12, 30159 Hannover, Germany
E-mail: claassen@dzhw.eu

Despite the promising combination of web surveys with elements of face-to-face interviews through video communication technology, there has been little research on video-based web surveys. Importantly, there is limited knowledge about the share of respondents who are willing to participate in video-based web surveys and what conditions drive willingness. In addition, many existing studies have applied observational (or non-experimental) research designs (Höhne et al., 2023; Schober et al., 2023). In the present article, we address this research gap by experimentally investigating respondents' hypothetical willingness to participate in video-based web surveys using survey data from two non-probability online panels in Germany. We also employ follow-up probes asking respondents to elaborate on their (non)willingness and examine characteristics of respondents (e.g., age and education) and video-based web surveys (e.g., video-based answers) that are associated with (non)willingness.

Background and Research Questions

Empirical studies on video-based web surveys are rare and focus on question delivery rather than answer provision. Importantly, video-based questions can be delivered by both pre-recorded (human) interviewers as well as (artificial) virtual interviewers. A few studies have examined the data quality of web surveys with pre-recorded interviewers in comparison to text-only web surveys, face-to-face, live video, and telephone interviews. The results are somewhat mixed. While one study found less information disclosure in a web survey with pre-recorded interviewer videos than in a text-only web survey (Fuchs, 2009), other studies did not find any disclosure differences (Fuchs & Funke, 2007, 2009). Conrad et al. (2023) found less rounding in open numeric answers and more disclosure of sensitive information in a web survey with pre-recorded interviewers compared to a text-only web survey. In addition, in the study by Conrad et al. (2023), about 70% of respondents declared that they have built a connection with the pre-recorded interviewer. Furthermore, Haan et al. (2017) reported that web surveys with pre-recorded interviewers result in less socially desirable answers than face-to-face and telephone interviews. Recency effects, as an indicator of survey-satisficing, were less common in web surveys with pre-recorded interviewers than in telephone interviews. Finally, it was recently shown that if interviewer effects occur, they tend to be larger in live video interviews than in web surveys with pre-recorded interviewers (West et al., 2022; see Table 1 for a systematic outline of studies). These findings point to data quality benefits when it comes to web surveys with pre-recorded interviewers.

With respect to virtual interviewers, Conrad et al. (2015) discovered that those with advanced dialog capabilities, as opposed to virtual interviewers with limited dialog capabilities, elicited more precise responses. Respondents interacted in a more social way, maintained more eye contact, asked more frequently for clarifications, and rated the virtual interviewers as more personal and less distant when they included advanced dialog capabilities. For virtual interviewers with greater facial animation, respondents showed a somewhat higher level of engagement (Conrad et al., 2015). Conrad et al. (2020) reported a decrease in socially desirable responding when the demographic characteristics of the respondent and virtual interviewers matched. When respondents shared the same ethnicity as the virtual interviewers, they were more inclined to report being slightly overweight, which is considered a socially undesirable answer, compared to when respondents did not share the same ethnicity. These findings indicate that video-based web surveys may emulate human-like conversations, possibly improving data quality.

Following the so-called "uses and gratification" theory, we assume that videos, as a specific media type, can serve the gratification of individual needs (Katz et al. 1973; Montag et al. 2021). In line with this assumption, respondents' willingness depends on whether video-based web surveys align with respondents' needs, such as convenience, self-expression, and social interaction. The extent to which these needs are gratified serves as a rationale for explaining individual differences in willingness. For example, the extent to which video-based questions are perceived as convenient might be related to respondents' age. Older respondents may have visual impairments, which impedes question reading. Similarly, the extent to which respondents conceive video-based answers as a positive form of self-expression might be related to their level of extraversion. Since we take an exploratory approach in this article, we refrain from formulating specific research hypotheses. The uses and gratification theory rather serves as a general framework underlining potential differences in respondents' willingness to participate in video-based web surveys.

To the best of our knowledge, only two empirical studies have been published on respondents' willingness to participate in video-based web surveys so far (see Höhne et al., 2023, and Schober et al., 2023). Schober et al. (2023) found that the majority of respondents (63%) is willing to participate in web surveys with video-based questions. This corresponds to the findings reported by Höhne et al. (2023). However, the latter study also reported that only about 20% of respondents are willing to participate in web surveys with video-based answers and that willingness depends on various aspects, such as respondents' belief that technology facilitates communication, their interest in the survey, and their level of extraversion. These studies provide novel insights on the share of respondents who are willing to participate in video-based web surveys, including potential drivers of willingness. However, there is less knowledge on the characteristics of video-based web surveys driving respondents' willingness.

The discrepancy between respondents' comparatively high willingness to participate in web surveys with video-based questions and their comparatively low willingness to participate in web surveys with video-based answers is striking. It suggests that respondents' willingness to participate in video-based web surveys is not only associated with respondent characteristics, such as attitudes on technology and personality traits, but also with the answer component. In particular, the answer component of video-based web surveys can range from selecting predefined answer options to open textual answers to voice and video-based answers. Thus, willingness to participate in video-based web surveys might differ substantially across respondent groups. On the one hand, respondents might be less willing to participate in such web surveys if the answer component is considered as burdensome or inconvenient. This may particularly apply if the answer component is not video-based. For example, open textual answers require respondents to retrieve information from memory without a frame of reference in the form of predefined answer options (Zuell et al., 2015). Respondents need to formulate answers in their own words and manually enter them through the keyboard. This is especially burdensome when completing the web survey on a mobile device with a virtual on-screen keyboard shrinking the viewing space (Höhne et al., 2020). Open textual answers increase respondent burden and might reduce respondents' willingness to participate. On the other hand, respondents might be less willing to participate in video-based web surveys if the answer component invades respondents' privacy (Wenz et al., 2019). For example, some respondents feel uncomfortable to record answers using the built-in microphone of their smartphone (e.g., out of concerns to be overheard by third parties), leading to lower willingness to participate in web surveys with voice answers (Höhne, 2023; Lenzner & Höhne, 2022; Revilla et al., 2018). Web surveys with video-based answers are even more intrusive as they require respondents to not only share their voice but their overall appearance, gestures, facial expressions, and environment (Höhne et al., 2023).

Even though video-based web surveys have the potential to reduce respondent burden and improve data quality, they might lead to low participation rates because respondents feel uncomfortable with respect to video-based answers. Thus, the choice of the answer component is a key design decision in video-based web surveys that should follow empirical evidence. Before employing video-based web surveys, it is essential to gather more knowledge on respondents' willingness to participate in such web surveys and possible variations in the characteristics of video-based web surveys. We fill this research gap by analyzing data from two studies that we conducted in German non-probability online panels and in which we asked respondents for their willingness to participate in video-based web surveys. Specifically, we address the following three research questions (RQs):

> *RQ1*: What is the share of respondents being (non)willing to participate in video-based web surveys with different question and answer components?

RQ2: What are the reasons for respondents' (non)willingness to participate in video-based web surveys?

RQ3: What variables drive respondents' (non)willingness to participate in video-based web surveys?

Method

Study 1

Data Collection

Data for Study 1 was collected in the non-probability SoSci Panel (www.sosci-panel.de), which is a project of the Institute for Communication Science and Media Research at the Ludwig-Maximilian-University Munich (Germany) and the German Society for Journalism and Communication Science (DGPuK). The SoSci Panel does not pursue commercial goals and data can only be collected for academic studies. Researchers are eligible to submit study proposals that undergo a review process evaluating the methodological soundness of the studies. Upon acceptance, the panel staff sends out email invitations to the respondents of the SoSci Panel pool (recruited via an opt-in subscription process) to take part in the web survey. Researchers do not have to pay any fees as web survey data collection is free of charge.

The web survey was fielded from 16th May 2022 to 5th June 2022. A reminder was sent on 25th May 2022. The invitation email included information on the web survey's topic (new communication forms in web surveys), its estimated duration (approx. 20 minutes), and a link to the web survey. Respondents could participate with the device of their choice. The first survey page provided additional details on the web survey (e.g., the research group responsible for the web survey). We also included a statement of confidentiality, stating that the study adheres to EU and national data protection laws and regulations. Respondents took part voluntarily without the provision of incentives.

Sample

The invitation email was sent to 5,676 respondents (out of these emails, 68 could not be successfully delivered). Out of all successfully invited respondents, 1,146 respondents (20%) started the web survey, and 874 respondents (15%) finished it. These 874 respondents had a mean age of 49, and 63% of them were female. In terms of education, 5% had completed lower secondary school or less (low education level), 15% intermediate secondary school (medium education level), and 79% college preparatory secondary school or university-level education (high education level). Overall, 63% of respondents participated with a computer, 3% with a tablet, and 35% with a smartphone.

Experimental Design

Respondents were randomly assigned to one out of three experimental groups. The first group (n = 297) received two closed questions on their willingness to participate in web surveys with video-based questions and answers. The second group (n = 294) received the same questions, but the two willingness questions included an additional motivational statement placed between the question text and answer options, stressing the provision of an additional incentive. The third group (n = 283) again received the same questions, but the two willingness questions included an additional motivational statement, stressing the improvement of respondent experience in surveys. Irrespective of the experimental groups, both willingness questions were followed by an open probing question asking respondents to elaborate on their previous answer.

To evaluate the effectiveness of random assignment, we compared the sample composition between the three experimental groups. We found no statistically significant differences with respect to age, gender, education, and completion device.

Willingness Questions

Respondents received the following two questions (in one out of three versions):

1. *Video-based questions:* In general, are you willing to participate in web surveys in which **interviewers read questions to you via pre-recorded videos**?

   The videos would be played within the web survey, you would not have to download any software. You could start and stop the videos at your convenience.

   Version 1: No additional text (first experimental group).

   Version 2: You would receive an additional incentive for this (second experimental group).

   Version 3: Your willingness would help us to improve the experience of respondents in surveys (third experimental group).

2. *Video-based answers:* In general, are you willing to participate in web surveys in which **you self-record your answers via videos**?

   The video recordings would be gathered within the web survey, you would not have to download any software. You could delete the videos and record them again.

Version 1: No additional text (first experimental group).

Version 2: You would receive an additional incentive for this (second experimental group).

Version 3: Your willingness would help us to improve the experience of respondents in surveys (third experimental group).

The willingness questions were presented on two separate survey pages (single-question presentation) with the following vertically aligned answer options: 1 (*definitely yes*), 2 (*probably yes*), 3 (*probably no*), 4 (*definitely no*), 5 (*don't know how this works*), and 6 (*don't have a device with camera and microphone*; only for the second willingness question). The last two options were visually separated by a space from the remaining options. Both willingness questions were followed by an open probing question asking respondents to elaborate on their previous answer. Appendix A includes screenshots of the willingness and probing questions.

Study 2

Data Collection

Data for Study 2 was collected in the non-probability Bilendi panel (www.bilendi.de) in Germany from 5th December 2023 to 18th December 2023. Bilendi drew a sample from their online access panel based on quotas for age, gender, education, and federal state. In contrast to the SoSci Panel (Study 1), Bilendi is a commercial panel providing respondents incentives for web survey participation. The invitation email included information on the provided incentive and a link to the web survey. Respondents could participate with the device of their choice. On the first survey page, respondents were informed about the topic and procedure of the web survey. In addition, it included a statement of confidentiality, stating that the study adheres to existing data protection laws and regulations. This study was part of a larger study investigating perspectives on current challenges for society and politics. The web survey included questions on political attitudes and attitudes towards technology. In addition, it contained two unrelated experiments. The experiment analyzed in this study was embedded at the end of the web survey.

Sample

The invitation email was sent to 28,455 respondents, out of which 2,356 (8%) started the web survey and 1,993 respondents (7%) finished it. These 1,993 respondents had a mean age of 45, and 50% of them were female. In terms of education, 26% had completed lower secondary school or less (low education level), 28% intermediate secondary school (medium education level), and 46% college preparatory secondary school or university-level education (high education level). Overall, 44% of respondents participated with a computer, 4% with a tablet, and 52% with a smartphone.

Experimental Design

We randomly assigned respondents to one out of four closed question versions (or experimental groups) asking respondents about their willingness to participate in future video-based web surveys. These questions differed with respect to the answer component: The first group ($n = 485$) was asked about the willingness to participate in video-based web surveys with predefined answer options (i.e., selecting answers from a list of predefined options). The second group ($n = 515$) was asked about the willingness to participate in video-based web surveys with open textual answers (i.e., entering answers in an open text field). The third group ($n = 506$) was asked about the willingness to participate in video-based web surveys with voice answers (i.e., self-recording voice answers). The fourth group ($n = 487$) was asked about the willingness to participate in video-based web surveys with video-based answers (i.e., self-recording video answers).

To evaluate the effectiveness of random assignment, we compared the sample composition between the four experimental groups. We found no statistically significant differences with respect to age, gender, education, and completion device.

Willingness Questions

Respondents received one out of the following four question versions:

1. *Predefined answer options:* In general, would you be willing to participate in future web surveys in which interviewers read questions to you via video, and **you select your answers from a list of predefined options**?

   The videos would be played within the web survey. You would not need to download an additional program. The interviewer videos would be recorded so that you can start and pause them yourself.

2. *Open textual answers:* In general, would you be willing to participate in future web surveys in which interviewers read questions to you via video, and **you enter your answers in an open text field**?

   The videos would be played within the web survey. You would not need to download an additional program. The interviewer videos would be recorded so that you can start and pause them yourself.

3. *Voice answers:* In general, would you be willing to participate in future web surveys in which interviewers read questions to you via video, and **you provide your answers via voice recording**?

   The videos would be played within the web survey, the same applies to the voice recordings. You would not need to download an additional program. The interviewer videos would be recorded so that you can start and pause them yourself. Of course, you could delete your voice recordings and record them again.

4. *Video-based answers:* In general, would you be willing to participate in future web surveys in which interviewers read questions to you via video, and **you provide your answers via video recording**?

   The videos would be played within the web survey, the same applies to the video recordings. You would not need to download an additional program. The interviewer videos would be recorded so that you can start and pause them yourself. Of course, you could delete your video recordings and record them again.

The willingness question was presented with the following vertically aligned answer options: 1 (*definitely yes*), 2 (*probably yes*), 3 (*probably no*), 4 (*definitely no*), 5 (*don't know how this works*), and 6 (*don't have a device with a microphone*; only for group 3) or 6 (*don't have a device with camera and microphone*; only for group 4). The last two options were visually separated by a space from the remaining options. Appendix B includes screenshots of the willingness question.

Analytical Strategy

To examine the share of respondents being (non)willing to participate in video-based web surveys (first research question), we remain on a descriptive level and report answer distributions of the willingness questions. In the first step, we draw on data from Study 1 and look at the answer distributions of the two questions on respondents' willingness to participate in web surveys with video-based questions and answers, respectively. In the next step, we draw on data from Study 2 and examine the answer distribution of the question on respondents' willingness to participate in video-based web surveys across the four experimental groups (or answer components).

To investigate the reasons for respondents' (non)willingness (second research question), we draw on the open probing questions asked in Study 1. To do so, respondents' open textual answers were manually coded by a student assistant. The student assistant developed the coding schemes based on the data rather than using preconceived codes. The final coding schemes consisted of 9 (willingness for video-based questions) and 10 (nonwillingness for video-based questions) as well as 9 (willingness for video-based answers) and 10 (nonwillingness for video-based answers) categories, respectively. In addition, the coding schemes contained two non-substantial categories (misunderstanding how video-based questions or answers work as well as noninformative reasons and non-substantial answers). Using these coding schemes, a second student assistant independently coded a randomly selected subset of 10% of the answers to the two open probing questions (video-based questions: n = 77; video-based answers: n = 72) so that we could estimate inter-coder reliability. Inter-coder reliability was deemed "almost perfect" (Landis & Koch, 1977) with Cohen's kappa values of .85 and .88 and agreement rates of 85.7% and 88.9% for the two open probing questions, respectively. Cases in which the two coders disagreed were reviewed by the second author who made a final judgment. The reasons for respondents' (non)willingness are reported in percentages. Importantly, when looking at the reasons for respondents' willingness to participate in web surveys with video-based questions and answers, we only consider respondents who selected "definitely yes" or "probably yes" in the preceding closed question. Conversely, when looking at the reasons for respondents' nonwillingness, we only consider respondents who selected "definitely no" or "probably no" in the preceding closed question.

To identify variables driving respondents' willingness to participate in video-based web surveys (third research question), we run multiple OLS regressions. Based on Study 1, we estimate regression models with respondents' willingness to participate in web surveys with video-based questions and answers as dependent variables, respectively. The two non-substantial answer options are excluded from the regression analyses. We include the experimental groups, which differ in terms of the provision of a motivational statement, as independent dummy variables in the models (provision of an additional incentive: 1 = *yes*; improvement of respondent experience in surveys: 1 = *yes*; control group as reference). Using data from Study 2, we again run regression models with respondents' willingness to participate in video-based web surveys as dependent variable excluding the two non-substantial answer options. As before, we include the experimental groups, which differ in terms of the answer component, as independent dummy variables in the model (open textual answers: 1 = *yes*; voice answers: 1 = *yes*; video-based answers: 1 = *yes*; predefined answer options as reference). In addition, we use the following independent variables that were suggested by prior studies to be associated with respondents' willingness (Höhne, 2023; Höhne et al., 2023; Lenzner & Höhne, 2022; Revilla et al., 2018): belief that technology facilitates communication (7 ascending options), personality traits in terms of openness, conscientiousness, extraversion, agreeableness, and neuroticism (21 ascending options in Study 1 and 14 ascending options in Study 2, respectively[1]), and survey evaluations in terms of interest (7 ascending options), difficulty (7 ascending options), topic sensitivity (7 ascending options), and general evaluation (7 ascending options; Study 2 only). Furthermore, we control for age (continuous), female (1 = *yes*), high education (1 = *yes*) and medium education (1 = *yes*) with low as reference, and smartphone use in the survey (1 = *yes*). English translations of the question wordings of the independent variables are available in Appendix C. Data and analysis code for replication purposes are available via Harvard Dataverse (see https://doi.org/10.7910/DVN/CYNCJJ).

Results

Research Question 1

In a first step, we draw on data from Study 1 to describe the share of respondents being (non)willing to participate in web surveys with video-based questions and answers, respectively. Table 1 presents the results. Across the three experimental groups, about 60% of respondents are definitely or probably willing to participate in web surveys with video-based questions. Respondents' willingness to participate in web surveys with video-based answers is considerably lower. Only about 25% of respondents are definitely or probably willing to participate in such web surveys. According to two chi-squared tests, there are no significant differences between the three experimental groups (i.e., control, provision of an additional incentive, and improvement of respondent experience) regarding willingness to participate in web surveys with video-based questions ($\chi^2(6) = 12.22$, $p = .057$) and video-based answers ($\chi^2(6) = 10.13$, $p = .119$). In addition, about 3% of respondents state that they do not know how the video-based questions or answers work and about 6% of respondents state that they do not have a device with camera and microphone.

Next, we turn to Study 2 examining respondents' willingness to participate in video-based web surveys with different answer components. Table 2 shows the results separately for each of the four experimental groups. In the first group (predefined answer options), 72% of respondents are definitely or probably willing to participate in video-based web surveys. In the second group (open textual answers), willingness is slightly lower (67%). In contrast, only 47% of respondents in the third group (voice answers) and 38% in the fourth group (video-based answers) are definitely or probably willing to participate in video-based

1 In Study 1, we employed the BFI-S inventory by Schupp and Gerlitz (2008), measuring each of the Big Five personality traits with three items and a 7-point response scale. In Study 2, due to time restrictions, we employed the shorter BFI-10 inventory by Rammstedt et al. (2014), measuring each of the Big Five personality traits with two items and a 7-point response scale. For each trait, we summed up answers to the respective items.

web surveys. According to a chi-squared test, the differences between the four experimental groups are statistically significant ($\chi^2(9) = 137.11$, $p < .001$). Across all experimental groups, between 3% and 6% of respondents state that they do not know how video-based web surveys work and about 10% (third and fourth groups) state that they do not have a device with camera and microphone.

Research Question 2

In a second step, we look at the open probing questions in Study 1 and examine respondents' reasons for their (non)willingness to participate in web surveys with video-based questions and answers, respectively. Tables 3 and 4 show the results. With respect to their willingness to participate in web surveys with video-based questions, respondents most frequently indicate willingness without providing a substantial reason (22%; e.g., "I like the idea"), followed by convenience and time flexibility (21%), and mentioning a general openness to new technologies and procedures (12%). Furthermore, 10% of respondents say that this does not represent an additional burden, 9% explain that their willingness depends on the particular survey, situation, and timing, and another 8% indicate that video-based questions are more personal, authentic, and interactive. Less than 5% of respondents mention support for and interest in science as well as already using the technology in daily life. A few respondents (4%) provide other reasons (e.g., "additional incentive"), seem to misunderstand how video-based questions works (6%; e.g., "I can ask questions"), or provide noninformative reasons and non-substantial answers (9%; e.g., "./.").

With respect to their willingness to participate in web surveys with video-based answers, respondents most often indicate their willingness without providing a substantial reason (21%; e.g., "Why not?"). This is followed by convenience and time flexibility (16%), dependence on the particular survey, situation, and timing (13%), and support for and interest in science (13%). In addition, 11% of respondents indicate a general openness to new technologies and procedures and 6% state that they already use the technology in daily life. A few respondents (less than 5%) refer to no additional burden or describe video-based answers as more personal, authentic, and interactive. Furthermore, 6% of respondents mention other reasons (e.g., "Trust in data protection laws"), 3% seem to misunderstand how video-based answers work (e.g., "I can ask the interviewer comprehension questions"), and 7% provide noninformative reasons and non-substantial answers (e.g., "Depends on the weather").

Among respondents who are not willing to participate in web surveys with video-based questions, the most frequently mentioned reason is requiring too much effort (38%), followed by a preference for reading (17%), and concerns about third parties being around (12%). Furthermore, 9% of respondents refer to no added value and 7% mention privacy and data security concerns. Less than

5% of respondents mention technical obstacles, uncomfortableness, and indicate nonwillingness without providing a substantial reason (e.g., “Don’t want to”). Finally, 8% of respondents indicate other reasons (e.g., “Personal reasons”), 19% seem to misunderstand how video-based questions work (e.g., “I don’t want to download unknown software”), and 4% provide noninformative reasons and non-substantial answers (e.g., “No”).

Table 1 Respondents’ willingness to participate in web surveys with video-based questions and answers (Study 1)

Answer options	Control		Provision of an additional incentive		Improvement of respondent experience	
	Video-based questions (%)	Video-based answers (%)	Video-based questions (%)	Video-based answers (%)	Video-based questions (%)	Video-based answers (%)
Definitely no	10	40	11	31	14	42
Probably no	24	29	23	34	29	30
Probably yes	43	15	37	19	34	16
Definitely yes	20	4	26	7	20	5
Don’t know how this works	3	3	3	2	2	3
Don’t have a device with camera and microphone	–	8	–	7	–	4
n	297	297	293	293	283	278

Notes: The last two options were visually separated by a space from the remaining options (see Appendix A for screenshots of the two questions). Due to rounding percentages may not add up to 100%. The control group was also analyzed in Höhne et al. (2023).

With respect to respondents who are not willing to participate in web surveys with video-based answers, more than one third of respondents (41%) state privacy and data security concerns. This is followed by requiring too much effort (21%), uncomfortableness (19%), and nonwillingness without providing a substantial reason (8%; e.g., “Don’t want to”). Furthermore, 8% of respondents mention a lack of ability to control (altering or deleting recordings), 6% refer to no added value, and another 6% indicate concerns about third parties being around. Up to 4% of respondents mention a preference for writing or technical obstacles. Finally, 7% of respondents indicate other reasons (e.g., “Might distort survey results”), 1% seem to misunderstand how video-based answers work (e.g., “Don’t like to participate in a video conference”), and 3% provide noninformative reasons and non-substantial answers (e.g., “Why?”).

Table 2 Respondents' willingness to participate in video-based web surveys with different answer components (Study 2)

Answer options	Predefined answer options (%)	Open textual answers (%)	Voice answers (%)	Video-based answers (%)
Definitely no	9	9	12	21
Probably no	15	17	29	26
Probably yes	35	33	27	21
Definitely yes	37	34	20	17
Don't know how this works	4	6	4	3
Don't have a device with camera and microphone	–	–	9	11
n	485	515	505	487

Notes: The last two options were visually separated by a space from the remaining options (see Appendix B for screenshots of the four question versions). Due to rounding percentages may not add up to 100%.

Table 3 Reasons for willingness to participate in web surveys with video-based questions and answers (Study 1)

Reasons for willingness	Video-based questions (%)	Video-based answers (%)
Willingness (without providing a substantial reason)	22	21
Convenience and time flexibility	21	16
Willingness depends on survey, situation, and timing	9	13
General openness to new technologies and procedures	12	11
Support for and interest in science	4	13
No additional burden	10	3
More personal, authentic, and interactive	8	3
Already using technology in daily life	4	6
Other	4	6
Misunderstanding how video-based questions or answers work	6	3
Noninformative reasons and non-substantial answers	9	7

Notes: $N = 478$ (video-based questions) and $N = 173$ (video-based answers). Respondents were able to mention multiple reasons and thus percentages may not add up to 100%.

Table 4 Reasons for nonwillingness to participate in web surveys with video-based questions and answers (Study 1)

Reasons for nonwillingness	Video-based questions (%)	Video-based answers (%)
Requiring too much effort	38	21
Privacy and data security concerns	7	41
Preference for reading or writing	17	3
Uncomfortableness	1	19
Third parties around	12	6
No added value	9	6
Nonwillingness (without providing a substantial reason)	4	8
Lack of ability to control (altering or deleting recordings)	–	8
Technical obstacles	3	4
Other	8	7
Misunderstanding how video-based questions or answers work	19	1
Noninformative reasons and non-substantial answers	4	3

Notes: $N = 295$ (video-based questions) and $N = 550$ (video-based answers). Respondents were able to mention multiple reasons and thus percentages may not add up to 100%.

Research Question 3

Finally, we investigate which variables drive respondents' non(willingness) to participate in video-based web surveys. To do so, we first draw on data from Study 1 and run OLS regressions on respondents' willingness to participate in web surveys with video-based questions and answers, respectively. For each of the dependent variables we estimate two models. In the first model, we only include the experimental groups as independent dummy variables. In the second model, we add the additional independent variables (see "Analytical strategy"). Table 5 presents the results. Regarding respondents' willingness to participate in web surveys with video-based questions as dependent variable, only the second model is statistically significant and has explanatory power: Model M1 ($F(2,782) = 2.52$, $p = .081$, adjusted $R^2 = .00$) and M2 ($F(16,768) = 5.82$, $p < .001$, adjusted $R^2 = .09$). The experimental groups are not associated with respondents' willingness. Instead, the belief that technology facilitates communication, extraversion, and evaluating the current survey as more interesting and more difficult are positively associated with respondents' willingness to participate in web surveys with video-based questions. In contrast, higher age and being female are negatively associated with respondents' willingness.

Looking at the regression models with respondents' willingness to participate in web surveys with video-based answers as dependent variable, both models are statistically significant. Adding the additional independent variables in the second model increases explanatory power: M1 (F(2,730) = 4.14, p = .016, adjusted R^2 = .01) and M2 (F(16,716) = 7.04, p < .001, adjusted R^2 = .12). In both models, the dummy variable for the second experimental group (statement stressing the provision of an additional incentive) is positively associated with respondents' willingness to participate in web surveys with video-based answers. Similar to video-based questions, the belief that technology facilitates communication, extraversion, and evaluating the current survey as more interesting are positively associated with respondents' willingness. In contrast, using a smartphone to answer the survey is negatively associated with respondents' willingness.

Next, analyzing data from Study 2, we estimate two regression models with respondents' willingness to participate in video-based web surveys as dependent variable and include the experimental groups as independent dummy variables. Table 6 shows the results. Both models are statistically significant. Adding the additional independent variables in the second model increases explanatory power: M1 (F(3,1759) = 38.43, p < .001, adjusted R^2 = .06) and M2 (F(18,1744) = 25.29, p < .001, adjusted-R^2 = .17). In both models, the dummy variables for the third (voice answers) and fourth groups (video-based answers) are negatively associated with respondents' willingness to participate in video-based web surveys. Similarly, higher age and being female are negatively associated with respondents' willingness. In contrast, high education, extraversion, evaluating the survey as more interesting, and evaluating the survey more positively are positively associated with respondents' willingness to participate in video-based web surveys.

Discussion and Conclusion

The aim of this article was to investigate respondents' willingness to participate in video-based web surveys: web surveys in which at least one of the interview components (i.e., questions or answers) is based on video communication technology. For this purpose, we conducted two experimental studies. Similar to previous research (Höhne et al., 2023; Schober et al., 2023), our results show that respondents are more willing to participate in web surveys with video-based questions than to participate in web surveys with voice (Study 2) or video-based answers (Studies 1 and 2). Respondents' reasons for (non)willingness differ between video-based questions and answers (Study 1). Overall, willingness for video-based web surveys is driven by respondent as well as survey character-

Table 5 OLS regressions with willingness to participate in web surveys with video-based questions and answers as dependent variables (Study 1)

Independent variables	Video-based questions				Video-based answers			
	M1		M2		M1		M2	
	b	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*
Motivational statement (reference: control)								
Provision of an additional incentive	.06 (*p* = .458)	.08	.04 (*p* = .625)	.08	.22 (*p* = .007)	.08	.20 (*p* = .008)	.08
Improvement of respondent experience	−.13 (*p* = .118)	.08	−.15 (*p* = .062)	.08	.04 (*p* = .664)	.08	.06 (*p* = .454)	.08
Belief technology facilitates communication			.07 (*p* = .004)	.02			.09 (*p* < .001)	.02
Personality traits								
Openness			.01 (*p* = .342)	.01			−.00 (*p* = .876)	.01
Conscientiousness			−.01 (*p* = .507)	.01			.01 (*p* = .636)	.01
Extraversion			.02 (*p* = .008)	.01			.05 (*p* < .001)	.01
Agreeableness			−.02 (*p* = .142)	.01			−.01 (*p* = .411)	.01
Neuroticism			−.00 (*p* = .664)	.01			−.00 (*p* = .615)	.01
Survey evaluation								
Interest			.14 (*p* < .001)	.03			.09 (*p* < .001)	.02
Difficulty			.07 (*p* = .024)	.03			.02 (*p* = .448)	.03
Topic sensitivity			−.01 (*p* = .582)	.02			−.03 (*p* = .192)	.02

Table 5 (continued)

Independent variables	Video-based questions				Video-based answers			
	M1		M2		M1		M2	
	b	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*
Age			−.01 ($p < .001$)	.00			.00 ($p = .606$)	.00
Female			−.17 ($p = .018$)	.07			−.12 ($p = .096$)	.07
Education (reference: low)								
Medium			−.00 ($p = 1.00$)	.18			.23 ($p = .254$)	.20
High			.05 ($p = .747$)	.17			.06 ($p = .723$)	.18
Smartphone use			−.14 ($p = .060$)	.07			−.16 ($p = .021$)	.07
Constant	2.76 ($p < .001$)	.05	2.10 ($p < .001$)	.43	1.81 ($p < .001$)	.06	.45 ($p = .272$)	.41
N	785		785		733		733	

Notes: M = model, *SE* = robust standard error. Coding of the dependent variables video-based questions and answers: 1 (*definitely no*), 2 (*probably no*), 3 (*probably yes*), and 4 (*definitely yes*). Exclusion of non-substantial answer options and respondents with missing values for any of the independent variables. We also excluded respondents selecting the answer option “divers” when asked for their gender because the share was very low ($n = 15$).

Table 6 OLS regressions with willingness to participate in video-based web surveys as dependent variable (Study 2)

Independent variables	M1		M2	
	b	*SE*	*b*	*SE*
Answer component (reference: predefined answer options)				
Open textual answers	−.06 (p = .355)	.06	−.07 (p = .278)	.06
Voice answers	−.43 (p < .001)	.07	−.46 (p < .001)	.06
Video-based answers	−.63 (p < .001)	.07	−.65 (p < .001)	.06
Belief technology facilitates communication			.03 (p = .107)	.02
Personality traits				
Openness			.01 (p = .094)	.01
Conscientiousness			−.01 (p = .423)	.01
Extraversion			.03 (p = .001)	.01
Agreeableness			.02 (p = .130)	.01
Neuroticism			−.01 (p = .682)	.01
Survey evaluation				
Interest			.16 (p < .001)	.02
Difficulty			−.02 (p = .360)	.02
Topic sensitivity			.00 (p = .911)	.02
General evaluation			.05 (p = .018)	.02
Age			−.01 (p < .001)	.00
Female			−.15 (p = .001)	.05
Education (reference: low)				
Medium			.01 (p = .870)	.06
High			.20 (p = .001)	.06
Smartphone use			.06 (p = .194)	.05
Constant	3.04 (p < .001)	.05	1.66 (p < .001)	.25
N	1,763		1,763	

Notes: M = model, *SE* = robust standard error. Coding of dependent variable: 1 (*definitely no*), 2 (*probably no*), 3 (*probably yes*), and 4 (*definitely yes*). Exclusion of non-substantial answer options and respondents with missing values for any of the independent variables.

istics. In the following, we discuss our empirical findings in light of our three research questions. It is important to note that the results of the two studies can only be compared cautiously because both samples are non-probabilistic and differ in their composition (e.g., in terms of education).

With respect to our first research question on the share of respondents being (non)willing to participate in video-based web surveys we found that many respondents are willing to participate in web surveys with video-based questions. However, when it comes to video-based answers, the picture changes. As shown in Study 1, only about 25% of respondents are definitely or probably willing to participate in web surveys with video-based answers. Interestingly, willingness for video-based answers is substantially higher in Study 2 (almost 40%). While respondents' willingness for voice answers is only slightly higher (about 50%), most respondents are willing to select predefined answer options and enter open textual answers when surveyed through pre-recorded interviewers (about 70%, respectively). On average, less than 10% of respondents indicate that they do not know how the video-based questions and answers work or state that they do not have a device with camera and microphone (Studies 1 and 2). This suggests that only a minority of respondents cannot participate in video-based web surveys because they lack knowledge or equipment. However, it is possible that we underestimate the share of these respondents, as some respondents may have selected "probably no" or "definitely no" due to a lack of knowledge and equipment.

Our second research question investigated the reasons for respondents' (non) willingness to participate in video-based web surveys (Study 1). Overall, we found relatively similar patterns for both video-based questions and answers. The most frequently mentioned reasons for willingness to use this technology were as follows: willing without providing a substantial reason (about 20%) and convenience and time flexibility (about 20%). Particularly, the latter category indicates that video-based communication potentially facilitates web survey completion for some respondents. A substantial minority of respondents also mentioned that their willingness depends on the survey, situation, and timing (about 10%) and that they are generally open to new technologies and procedures (about 10%). The former category suggests that these respondents may change their opinion if the survey setting does not fit their expectations (e.g., presence of third parties). Support for and interest in science was frequently mentioned when it comes to video-based answers (more than 10%), but not when it comes to video-based questions (less than 5%) showing that respondents may have different motivations for both components.

For nonwillingness, we found substantial differences between video-based questions and answers, which indicates that respondents are more divided. While about 40% stated that video-based questions require too much effort, only about 20% stated this reason when it comes to video-based answers. Privacy concerns were much more common for video-based answers (about 40%) than for video-based questions (less than 10%). This similarly applies to the category uncomfortableness when it comes to video-based answers (about 20%), which was indicated by only 1% of respondents for video-based questions. Thus, video-based answers put some respondents in an unpleasant position. Finally, concerns regarding third parties were mentioned both regarding video-based questions (about 10%) and video-based answers (about 5%). This indicates that some respondents may be concerned about disturbing other people.

Video-based answers do not only provide a direct view into respondents' environment (e.g., their home or workplace), but also into their appearance, gestures, and facial expressions. Thus, some respondents may find them intrusive and consider them as a threat to their overall privacy. For example, research on answering web survey questions through pictures has shown that respondents are more willing to provide a picture of what they see (or their location) than a selfie that would allow researchers to infer their in-the-moment feelings (Bosch et al., 2022). In order to pursue such requests in web surveys it appears inevitable to take respondents' privacy concerns seriously and to include measures that potentially alleviate these concerns. Following Keusch and Conrad (2022), this may include the explanation of the importance of providing video-based answers, a reduced utilization of video-based answers (only for few questions instead of the entire web survey), or the provision of additional incentives. As shown in Table 5, promising an additional incentive was positively associated with respondents' willingness to participate in web surveys with video-based answers.

Regarding our third research question on variables driving respondents' (non)willingness to participate in video-based web surveys we found that specific respondent groups seem to be more open than others. For example, respondents believing that technology facilitates communication (Study 1), being extraverted, and having higher interest in the survey (Studies 1 and 2) seem to be more attached to video-based web surveys. We also found that younger, male (Studies 1 and 2), and higher educated (Study 2) respondents show higher willingness. Importantly, the comparison of the findings on education across studies is limited because of the large imbalance in Study 1 (it only includes 5% low educated respondents). Nonetheless, these findings indicate that traditional respondent characteristics, such as age and gender, are not sufficient to infer respondents' willingness for new communication forms in web surveys. For the future of web surveys, it is key to further investigate respondents' motivations for new communication forms, because it may help to recruit respondents that are reluctant to participate in traditional text-only web surveys (Revilla & Höhne, 2020). Considering our results on the reasons for nonwillingness, it would be worthwhile to include direct measures of respondents' privacy concerns in future studies. In addition, future studies should examine whether the use of video communication in other contexts, such as work or personal life, influences willingness to participate in video-based web surveys.

Importantly, these considerations are also related to a more theory-driven investigation of respondents' willingness to participate in video-based web surveys. More specifically, we advocate for the creation of survey participation frameworks that can be empirically tested or from which empirically testable research hypotheses can be inferred. A starting point for the creation of such theoretical frameworks can be the "uses and gratification" theory (Katz et al.

1973). Potentially, this helps to develop more respondent-centered web survey data collection strategies (Wilson & Dickinsson, 2022).

This article has some methodological limitations that provide avenues for future research. First and foremost, we only asked respondents for their hypothetical willingness and did not examine their actual participation in video-based web surveys. Previous research has indicated that hypothetical willingness and successful participation in data collection tasks can differ and may be driven by different variables (Keusch et al., 2024). Future research should investigate whether and to what extent the response rates of video-based web surveys are indeed as high as suggested by respondents' hypothetical willingness. Second, we conducted our studies among respondents of non-probability online panels and thus we cannot infer to the general population. For example, as shown by Höhne (2023) as well as Lenzner and Höhne (2022), willingness among respondents from non-probability and probability-based panels differ. If respondents of probability-based panels are indeed more reluctant to use video-based technology than respondents of non-probability panels, researchers would have to weigh the (possible) benefits of collecting richer and more nuanced data (by employing voice and video-based answers in non-probability panels) against the possibility of inferring from the sample to the general population. Therefore, it is important that future research examines differences in the willingness of respondents from non-probability and probability-based panels as well as cross-sectional web surveys in general. Third, we argued that video-based web surveys may be particularly well suited for respondents with lower literacy levels and with lower competence in the language of the web survey. However, we were not able to include measures of these competencies in the web survey and thus we could not properly distinguish between low educated respondents and those with low literacy when it comes to the survey language (German). We encourage future research to examine this issue more thoroughly by employing methods for determining literacy levels of respondents. For example, this includes lexical decision tasks (Juhasz et al., 2019; Schwarz et al., 2020), sentence reading tests (Bergmann & Wimmer, 2008; Kroh et al., 2016), and literacy self-reports (Snowling et al., 2012). This way, it is possible to better isolate the impact of video-based web surveys on low-literacy groups.

Finally, given the continual growth of web surveys and the significance of video-based communication in our daily lives, it is inevitable that these forms of communication will soon integrate into web surveys. Indeed, several studies are already underway experimenting with video-based web surveys (Conrad et al., 2023; West et al., 2022). This development is further supported by advances in Generative Artificial Intelligence (GenAI), enabling researchers to create virtual interviewers varying in their visual appearance and speech characteristics (Höhne, Neuert, & Claassen, 2024). Since virtual interviewers can be created quickly and at low costs, GenAI may render video-based web surveys more feasible. In a next step, it would be worthwhile to investigate if respondents' willingness to participate in video-based web surveys is dependent on whether questions are delivered by pre-recorded (human) interviewers or AI-generated virtual interviewers. Importantly, video-based web surveys, both with pre-recorded and virtual interviewers, hold immense potential to emulate crucial aspects of face-to-face interviews while remaining highly efficient in terms of time and cost. Moreover, the self-administration mode offers respondents a level of flexibility that traditional face-to-face (including live video) interviews cannot match. Consequently, there is no waste of interviewer time. Additionally, the elimination of interviewer travel reduces fieldwork efforts and renders data collection more environmentally sustainable. Hence, it is prudent to begin contemplating the integration of these communication forms into web surveys and determining the most effective methods for their incorporation.

References

Bauer, I., Kunz, T., & Gummer, T. (2025). Plain language in web questionnaires: Effects on data quality and questionnaire evaluation. *International Journal of Social Research Methodology, 28*(1), 57–69. https://doi.org/10.1080/13645579.2023.2294880

Bergmann, J., & Wimmer, H. (2008). A dual-route perspective on poor reading in a regular orthography: Evidence from phonological and orthographic lexical decisions. *Cognitive Neuropsychology, 25*(5), 653–676. https://doi.org/10.1080/02643290802221404

Bosch, O. J., Revilla, M., Qureshi, D. D., & Höhne, J. K. (2022). A new experiment on the use of images to answer web survey questions. *Journal of the Royal Statistical Society Series A*: *Statistics in Society, 185*(3), 955–980. https://doi.org/10.1111/rssa.12856

Callegaro, M., Lozar Manfreda, K., & Vehovar, V. (2015). *Web survey methodology*. Sage Publications.

Conrad, F. G., Schober, M. F., Hupp, A. L., West, B. T., Larsen, K. M., Ong, A. R., & Wang, T. (2023). Video in survey interviews: Effects on data quality and respondent experience. *Methods, Data, Analyses, 17*(2), 135–170. https://doi.org/10.12758/mda.2022.13

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

Conrad, F. G., Schober, M. F., Nielsen, D., & Reichert, H. (2020). Virtual interviewers, social identities, and survey measurement error. In K. Olson, J. D. Smyth, J. Dykema, A. L. Holbrook, F. Kreuter, & B. T. West (Eds.), *Interviewer effects from a total survey error perspective* (pp. 149–164). CRC Press. https://doi.org/10.1201/9781003020219

Daikeler, J., Bosnjak, M., & Lozar Manfreda, K. (2020). Web versus other survey modes: An updated and extended meta-analysis comparing response rates. *Journal of Survey Statistics and Methodology, 8*(3), 513–539. https://doi.org/10.1093/jssam/smz008

Fuchs, M. (2009). Gender-of-interviewer effects in a video-enhanced web survey: Results from a randomized field experiment. *Social Psychology, 40*(1), 37–42. https://doi.org/10.1027/1864-9335.40.1.37

Fuchs, M., & Funke, F. (2007). Video web survey: Results of an experimental comparison with a text-based web survey. In M. Trotman , T. Burrell, L. Gerrard, K. Anderton, G. Basi, M. Couper, K. Morris, D. Birks, A. J. Johnson, R. Baker, M. Rigg, S. Taylor, & A. Westlake (Eds.), *The challenges of a changing world: Proceedings of the Fifth ASC International Conference* (pp. 63–80). Association for Survey Computing.

Fuchs, M., & Funke, F. (2009). Die video-unterstütze Online-Befragung: Soziale Präsenz, soziale Erwünschtheit und Underreporting sensitiver Informationen [Video-assisted online surveys: Social presence, social desirability, and underreporting of sensitive information]. In N. Jackob, H. Schoen, & T. Zerback (Eds.), *Sozialforschung im Internet* [Social research on the Internet] (pp. 159–180). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-91791-7_10

Haan, M., Ongena, Y. P., Vannieuwenhuyze, J. T. A., & de Glopper, K. (2017). Response behavior in a video-web survey: A mode comparison study. *Journal of Survey Statistics and Methodology, 5*(1), 48–69. https://doi.org/10.1093/jssam/smw023

Höhne, J. K. (2023). Are respondents ready for audio and voice communication channels in online surveys? *International Journal for Social Research Methodology, 26*(3), 335–342. https://doi.org/10.1080/13645579.2021.1987121

Höhne, J. K., Kern, C., Gavras, K. & Schlosser, S. (2024). The sound of respondents: Predicting respondents' level of interest in questions with voice data in smartphone surveys. *Quality and Quantity, 58*, 2907–2927. https://doi.org/10.1007/s11135-023-01776-8

Höhne, J. K., Neuert, C., & Claassen, J. (2024). *Examining the effects of embodied interviewing agents on open narrative responses*. ResearchGate. http://doi.org/10.13140/RG.2.2.25109.95203

Höhne, J. K., Schlosser, S., Couper, M. P., & Blom, A. G. (2020). Switching away: Exploring on-device media multitasking in web surveys. *Computers in Human Behavior, 111*, Article 106417. https://doi.org/10.1016/j.chb.2020.106417

Höhne, J. K., Ziller, C., & Lenzner, T. (2023). Investigating respondents' willingness to participate in video-based web surveys. *International Journal of Market Research, 66*(1), 3–13. https://doi.org/10.1177/14707853231198788

Juhasz, B. J., Yap, M. J., Raoul, A., & Kaye, M. (2019). A further examination of word frequency and age-of-acquisition effects in English lexical decision task performance: The role of frequency trajectory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*(1), 82–96. http://doi.org/10.1037/xlm0000564

Katz, E., Blumler, J. G., & Gurevitch, M. (1973). Uses and gratifications research. *Public Opinion Quarterly, 37*(4), 509–523. https://doi.org/10.1086/268109

Keusch, F., & Conrad, F. G. (2022). Using smartphones to capture and combine self-reports and passively measured behavior in social research. *Journal of Survey Statistics and Methodology, 10*(4), 863–885. https://doi.org/10.1093/jssam/smab035

Keusch, F., Pankowska, P. K., Cernat, A., & Bach, R. L. (2024). Do you have two minutes to talk about your data? Willingness to participate and nonparticipation bias in Facebook data donation. *Field Methods, 36*(4), 279–293. https://doi.org/10.1177/1525822X231225907

Kroh, M., Lüdtke, D., Düzel, S., & Winter, F. (2016). *Response error in a web survey and a mailed questionnaire: The role of cognitive functioning* (SOEPpapers on Multidisciplinary Panel Data Research No. 888). Deutsches Institut für Wirtschaftsforschung (DIW), Berlin. https://hdl.handle.net/10419/149803

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174. https://doi.org/10.2307/2529310

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

Montag, C., Yang, H., & Elhai, J. D. (2021). On the psychology of TikTok use: A first glimpse from empirical findings. *Frontiers in Public Health, 9,* Article 641673. https://doi.org/10.3389/fpubh.2021.641673

Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C., & Kovaleva, A. (2014). Big Five Inventory (BFI-10). *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)*. https://doi.org/10.6102/zis76

Revilla, M., Couper, M. P., & Ochoa, C. (2018). Giving respondents voice? The feasibility of voice input for mobile web surveys. *Survey Practice, 11*(2). https://doi.org/10.29115/SP-2018-0007

Revilla, M., & Höhne, J.K. (2020). Comparing the participation of Millennials and older age cohorts in the CROss-National Online Survey panel and the German Internet Panel. *Survey Research Methods, 14*(5), 499–513. https://doi.org/10.18148/srm/2020.v14i5.7619

Schober, M. F. (2018). The future of face-to-face interviewing. *Quality Assurance in Education, 26*(2), 290–302. https://doi.org/10.1108/QAE-06-2017-0033

Schober, M. F., Okon, S., Conrad, F. G., Hupp, A. L., Ong, A. R., & Larsen, K. M. (2023). Predictors of willingness to participate in survey interviews conducted by live video. *Technology, Mind, and Behavior, 4*(2). https://doi.org/10.1037/tmb0000100

Schupp, J., & Gerlitz, J.-Y. (2008). Big Five Inventory-SOEP (BFI-S). *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)*. https://doi.org/10.6102/zis54

Schwarz, H., Revilla, M., & Weber, W. (2020). Memory effects in repeated survey questions: Reviving the empirical investigation of the independent measurements assumption. *Survey Research Methods, 14*(3), 325–344. https://doi.org/10.18148/srm/2020.v14i3.7579

Silber, H., Danner, D., & Rammstedt, B. (2019). The impact of respondent attentiveness on reliability and validity. *International Journal of Social Research Methodology, 22*(2), 153–164. https://doi.org/10.1080/13645579.2018.1507378

Snowling, M., Dawes, P., Nash, H., & Hulme, C. (2012). Validity of a protocol for adult self-report of Dyslexia and related difficulties. *Dyslexia, 18*(1), 1–15. https://doi.org/10.1002/dys.1432

Sun, H., Conrad, F. G., & Kreuter, F. (2021). The relationship between interviewer-respondent rapport and data quality. *Journal of Survey Statistics and Methodology, 9*(3), 429–448. https://doi.org/10.1093/jssam/smz043

Wenz, A., Jäckle, A., Couper, M. P. (2019). Willingness to use mobile technologies for data collection in a probability household panel. *Survey Research Methods, 13*(1), 1–22. https://doi.org/10.18148/srm/2019.v1i1.7298

West, B. T., Ong, A. R., Conrad, F. G., Schober, M. F., Larsen, K. M., & Hupp, A. L. (2022). Interviewer effects in live video and prerecorded video interviewing. *Journal of Survey Statistics and Methodology, 10*(2), 317–336. https://doi.org/10.1093/jssam/smab040

Wilson, L., & Dickinson, E. (2022). *Respondent centred surveys. Stop, listen and then design*. Sage Publications.

Zuell, C., Menold, N., & Körber, S. (2015). The influence of the answer box size on item nonresponse to open-ended questions in a web surveys. *Social Science Computer Review, 33*(1), 115–122. https://doi.org/10.1177/0894439314528091

Appendix A

Screenshots of the Willingness and Probing Questions (Study 1)

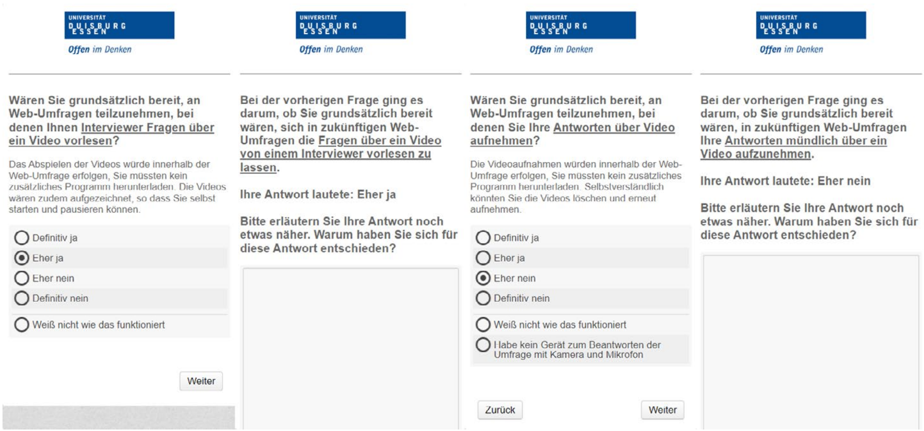


Figure A1 Exemplary smartphone screenshots of the two willingness questions and the two open probing questions in the first experimental group (no motivational statement) in Study 1. The first set (on the left) shows the questions on video-based questions and the second set shows the questions on video-based answers (on the right).

Appendix B

Screenshots of the Willingness Question (Study 2)

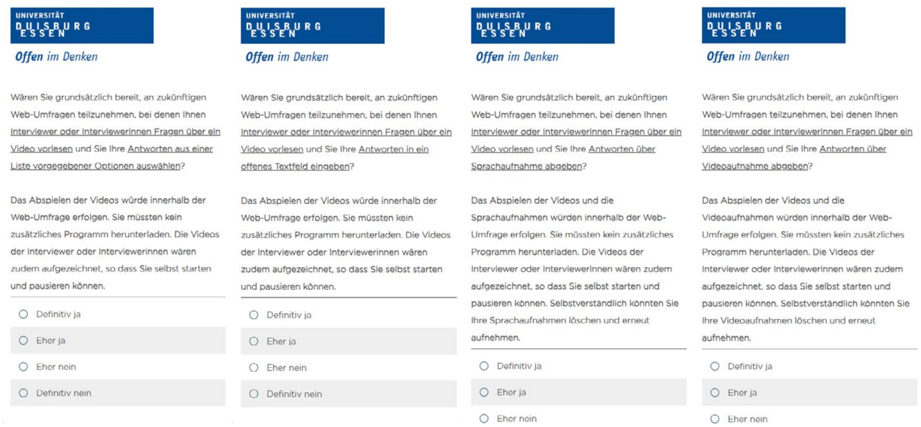


Figure B1 Exemplary smartphone screenshots of the four versions of the willingness question in Study 2. Question versions from left to right: predefined answer options (Group 1), open textual answers (Group 2), voice answers (Group 3), and video-based answers (Group 4).

Appendix C

English Translations of the Question Wordings

Belief That Technology Facilitates Communication (Studies 1 and 2)

To what extent does the following statement apply or not apply to you? New technologies make it easier for me to communicate with other people.
Answer options: 1 (*does not apply at all*) to 7 (*applies completely*)

Big Five Personality Traits (Study 1)

To what extent do the following statements apply or not apply to you?
I am someone who works thoroughly.
I am someone who is communicative and talkative.
I am someone who is sometimes a bit rude to others.
I am someone who is original, brings in new ideas.
I am someone who often worries.
I am someone who is reserved.
I am someone who is forgiving.
I am someone who is rather lazy.
I am someone who can come out of his shell, is sociable.
I am someone who appreciates artistic experiences.
I am someone who gets nervous easily.
I am someone who completes tasks effectively and efficiently.
I am someone who treats others with consideration and kindness.
I am someone who has vivid imagination and ideas.
I am someone who is relaxed and copes well with stress.
Answer options: 1 (*does not apply at all*) to 7 (*totally applies*)

Big Five Personality Traits (Study 2)

To what extent do the following statements apply to you?
I am rather restrained, reserved.
I trust others easily, believe in the good in people.
I am comfortable, tend to be lazy.
I am relaxed, do not let stress upset me.
I have little artistic interest.
I come out of my shell and am sociable.
I tend to criticize others.
I complete tasks thoroughly.
I get nervous and insecure easily.
I have an active imagination, I am creative.
Answer options: 1 (*does not apply at all*) to 7 (*totally applies*)

Survey Interest (Studies 1 and 2)

How interesting did you find it to answer the questions asked?
Answer options: 1 (*very interesting*) to 7 (*not interesting at all*)

Survey Difficulty (Studies 1 and 2)

How easy or difficult did you find it to answer the questions asked?
Answer options: 1 (*very easy*) to 7 (*very difficult*)

Topic Sensitivity (Studies 1 and 2)

How personal did you find answering the questions asked?
Answer options: 1 (*very personal*) to 7 (*not personal at all*)

General Evaluation (Study 2)

How did you like the survey overall?
Answer options: 1 (*not at all*) to 7 (*very much*)