EUROPEAN JOURNAL OF WORK AND ORGANIZATIONAL PSYCHOLOGY
https://doi.org/10.1080/1359432X.2025.2517592

Assessing biasing factors in asynchronous video interviews: applicant completion decisions, video background, and evaluation format

Nicolas Roulin[a], Antonis Koutsoumpis[b], Shahad Abdulrazaq[a], Janneke K. Oostrom[b], Yiyu Xie[c] and Alexander MacIntosh[c]

[a]Department of Psychology, Saint Mary's University, Halifax, NS, Canada; [b]Department of Social Psychology, Tilburg School of Social and Behavioral Sciences, Tilburg University, Tilburg, The Netherlands; [c]Acuity Insights

ABSTRACT

Asynchronous video interviews (AVIs) have become popular selection methods due to their flexibility and cost savings but might introduce new forms of bias. For instance, interviewees often complete them from home, their surroundings might signal personal or protected statuses, and technology issues might distort the information provided. This paper leverages two complementary studies to examine (a) the AVI completion decisions, recording quality, and background elements present in high- and low-stakes job interviews, (b) to what extent these AVI-specific elements and interviewees' characteristics can bias performance ratings, and (c) whether evaluation standardization can help mitigate such biases. Study 1 used mock interviews with ($N = 626$) Prolific participants evaluated by professional hiring managers. Study 2 involved high-stakes interviews with ($N = 523$) real applicants for competitive education programmes evaluated by trained raters using either standardized or unstandardized approaches. AVI elements (attire, room tidiness, technical issues, background) were coded in both studies. Results showed that completion decisions depended on AVI stakes and could influence evaluations. Issues with recording quality were rare and modestly related to AVI evaluations. AVI backgrounds signalling personal or protected statuses were very rare and unrelated to evaluations. Evaluations standardization reduced bias only in relation to sex-based differences, but not other interviewee characteristics.

ARTICLE HISTORY
Received 10 December 2024
Accepted 4 June 2025

KEYWORDS
Asynchronous video interviews; technology-mediated interviews; bias; structured interview

Virtual interviewing technology has rapidly advanced over the last few years, with asynchronous video interviews (AVIs) becoming a popular selection tool (e.g., Dunlop et al., 2022). Unlike in-person or video-conference interviews, AVIs present applicants with pre-selected interview questions and require them to record their answers via webcam and microphone (Lukacik et al., 2022). These video recordings are later evaluated by hiring managers or, in some cases, automatically rated using machine-learning models (Hickman et al., 2022; Koutsoumpis et al., 2024; Liff et al., 2024). Using AVIs helps organizations narrow down the number of applicants they may choose to invite to the next step of the selection process – typically an in-person interview, which can be very costly for both the organization and the applicant. As a result, AVIs provide organizations with advantages in terms of cost savings, time efficiency, and greater scheduling flexibility (Brenner et al., 2016; Castro & Gramzow, 2015; Griswold et al., 2021). However, little is known in terms of the possible biasing effects that AVIs may introduce to the selection process, and even less is known about how to minimize these biases.

A recent conceptual model of AVI design proposed that applicants' completion decisions, video-recording quality, and the background elements visible in recordings could affect interviewers' perceptions and evaluations (Lukacik et al., 2022). For instance, applicants make decisions to complete their AVI, such as how they dress and where they record it. These decisions, along with factors that may be beyond the applicant's control (e.g., internet connection), will influence the content and characteristics of their recordings, such as video or audio quality, and lighting conditions. Many of these characteristics are unrelated to job qualifications, yet they could influence evaluators' judgements (McColl & Michelotti, 2019). Similarly, elements visible in the applicant's background (e.g., framed pictures, books, artwork, posters) can signal otherwise-hidden statuses (political or religious affiliation, sexual orientation, family status), and make invisible stigmas visible to hiring managers (Summers et al., 2018). Of course, those AVI-specific factors add to biasing factors that are already present in traditional in-person interviews, such as applicants'

CONTACT Nicolas Roulin nicolas.roulin@smu.ca
Supplemental data for this article can be accessed online at https://doi.org/10.1080/1359432X.2025.2517592

demographic characteristics (McCarthy et al., 2010) or appearance (Martin-Raugh et al., 2023). Empirical studies have examined a few AVI recording or background elements such as messiness (Powell et al., 2023), the type of room or background blurring (Scott & Roulin, 2024), or protected statuses (Basch et al., 2024; Roulin, Lukacik, et al., 2023). Yet, all these studies relied on experimental designs, manipulating background elements in mock interviews and asking evaluators to rate scripted responses by actors. It is thus unclear to what extent those findings generalize to real AVIs. For example, how often are video-recording quality issues or potentially biasing background elements visible in applicants' AVI recordings? And how much do recording quality or background elements influence performance evaluations?

Further, both conceptual and empirical work suggests that structured interview formats are more resistant to biasing factors (Campion et al., 1997; Levashina et al., 2014; McCarthy et al., 2010). By design, AVIs include several structure elements, such as asking the same questions to all applicants in the same order and eliminating probing or rapport building (Lukacik et al., 2022). However, the way responses are evaluated can vary extensively, ranging from very unstructured (e.g., providing a general 1–5 stars rating for a candidate) to very structured approaches (e.g., using behaviourally anchored rating scales [BARS] to evaluate responses to each question). Evidence from in-person interviews indicates that evaluation standardization is one of the least used structure elements, especially when hiring managers have received limited training (Roulin et al., 2019). Yet, research examining the unique benefits of evaluation standardization (vs. structure in general) is limited (see Lubbe & Nitsche, 2019 for a rare exception), especially in AVIs. It is thus important to examine whether using a more standardized or structured evaluation process can help reduce potentially biasing factors in AVIs.

Overall, the present research tests key theoretical propositions by Lukacik et al. (2022), and contributes to the AVI and personnel selection literature in the following ways: First, we examine to what extent AVI recordings differ in terms of (a) key completion decisions, namely attire and room tidiness, (b) recording characteristics indirectly influenced by completion decisions (e.g., video, audio, and lighting quality), as well as (c) visible background elements (i.e., cues about applicants' political or religious affiliation, sexual orientation, and family status). Second, we build on theoretical models (Derous et al., 2016; Huffcutt et al., 2011; Lukacik et al., 2022) and expand on past experimental research (e.g., Roulin, Lukacik, et al., 2023) to examine how recording and background elements impact interview performance ratings in AVIs with real interviewees. We also compare the effects of these AVI-specific elements to well-known biasing factors in interviews, namely, attractiveness and demographic characteristics (i.e., age and gender). Third, we explore whether a more standardized evaluation (e.g., using BARS) can help minimize these biases compared to unstructured scoring. Finally, although a few large-scale studies have relied on data from real applicants completing high-stakes AVIs (e.g., Griswold et al., 2021; McCarthy et al., 2021; Tilston et al., 2024), the vast majority of the AVI literature involves participants from online panels completing mock interviews. To ensure the generalizability of our findings, and to help clarify whether results from low- and high-stakes AVIs are comparable, we examine the above-mentioned elements in two independent studies: One relying on online participants completing a low-stakes mock AVI, and one with real applicants completing a high-stakes AVI as part of the admission process for selective education programmes.

Applicant completion decisions, recording quality, and background elements in AVIs

Although AVIs have been a part of the selection landscape for over a decade (Brenner et al., 2016), their use in practice grew exponentially during the COVID-19 pandemic, driven by the need for an effective selection method that allowed for social distancing (Dunlop et al., 2022; Handler, 2020). AVIs have remained popular since then, notably because of their flexibility (i.e., on-demand format), their capacity to broaden the applicant pool (i.e., the absence of a live interaction eliminate barriers associated with scheduling or time zones), and their reported reduced costs and shorter time-to-hire (e.g., Arseneault & Roulin, 2024; Lukacik et al., 2022; Torres & Mejia, 2017). Despite these benefits, AVIs have been associated with more negative applicant reactions than in-person or video-conference interviews (e.g., Griswold et al., 2021), although reactions also depend on AVI design, explanations, or applicants' demographics (Basch & Melchers, 2019; Tilston et al., 2024). In addition, technology-mediated interviews like AVIs might generate novel or unique forms of bias that are not present in traditional in-person interviews.

Lukacik et al. (2022) proposed a conceptual model delineating how the unique characteristics of AVIs, as well as specific AVI design decisions (e.g., preparation time, re-recording options), could influence applicant behaviours and outcomes. They made several predictions about how the way organizations design AVIs can influence applicants' attitudes and behaviours (e.g.,

fairness perceptions, motivation, anxiety, impression management), which should indirectly facilitate or hinder their interview performance. This side of their model has already been tested in several empirical studies, reporting mixed findings (e.g., Basch et al., 2021; Lukacik & Bourdage, 2025; Roulin, Wong, et al., 2023). Lukacik et al. (2022) also described how the content of applicants' video-recorded responses can uniquely influence (i.e., bias) evaluators' judgements. Two key elements are central to the present research: (1) applicants' AVI completion decisions can directly or indirectly impact the content or quality of recordings, and thus influence how they are evaluated; (2) when recording from a private space, applicants could (intentionally or unintentionally) provide cues about themselves that would otherwise remain hidden, which might bias evaluations. We expand on these two mechanisms in the next sections.

Applicant completion decisions, recording quality, and AVI ratings

The first mechanism Lukacik et al. (2022) proposed is that applicants' AVI completion decisions can directly or indirectly influence evaluators' judgements. More precisely, they suggest that attire (professional vs. casual) and the tidiness of the room can influence how professional the applicant is perceived to be by evaluators. In addition, the chosen time and location to record an AVI can impact lighting conditions, and internet connection speed or the device used can impact sound or image clarity, and stability. Not all applicants have the resources to make some recording decisions, such as opting to use a laptop instead of a phone or selecting a room with natural light or additional lighting sources. Yet, these elements are likely to influence interview evaluations. Lukacik et al.'s (2022) general argument is consistent with Huffcutt et al.'s (2011) theoretical model of interview performance, which argues that interviewers' evaluations depend on information processing effects. More precisely, they explain that interviewers' limited short-term memory means that they cannot process and evaluate all the information they receive during the interview. As a result, they base their decisions on only a fraction of the information available, often relying on information that most easily comes to mind (i.e., because of availability heuristics). An AVI recording highlighting a messy room or low-quality video or audio could arguably represent salient cues evaluators could focus on.

In addition, low recording quality is likely to result in incomplete information, which can lead to more negative evaluations. For example, Jaccard and Wood (1988) suggested that incomplete information increases uncertainty in the estimation of a target's attributes. This uncertainty often results in evaluators assigning these attributes a subjectively determined average level or incorporating additional "devaluation parameters". As such, candidates without clear non-verbal signals expressing their job motivation may be perceived as below average in this attribute. The importance of recording decisions and quality has been described in an observation study of high-stakes video interviews (McColl & Michelotti, 2019). This small study ($n = 30$) illustrated technical issues (e.g., audio or video disruptions) and setting issues (e.g., suboptimal lighting or background) in video interviews. These authors further suggested (but did not directly test) that these issues can distort the signal sent by applicants, and ultimately impact evaluations. Moreover, several authors have provided recommendations about how applicants should complete their AVIs. For instance, applicants are advised to avoid poor lighting or recording the interview via a mobile phone (Mejia & Torres, 2018). They are also told to carefully choose a location to record their responses that looks professional (Lee et al., 2021) or neutral (Davis et al., 2020). These suggestions, however, seem to be based on weak or nonexistent empirical evidence.

Experimental work has also shown how some completion decisions can influence AVI outcomes, despite some mixed findings. For instance, Suen et al. (2019) showed that applicant appearance in AVIs influenced raters' initial impressions and final interview ratings, confirming results from in-person interviews (Barrick et al., 2009; Martin-Raugh et al., 2023). Powell et al. (2023) found that completing the interview in a messy room was associated with lower evaluations than when it was done in a clean room, both when the interview took place in an office and home settings. However, Scott and Roulin (2024) found that AVI ratings were unimpacted by the type of room (home office vs. bedroom) or by using background blurring. Finally, Basch and Melchers (2024) examined webcam positioning leading to deviations in eye contact. They found that horizontal deviations (e.g., from using a dual-screen setup, with the camera installed on the side screen) negatively impacted evaluations, whereas vertical deviations (e.g., from a webcam installed above a computer screen) did not.

In summary, applicants' completion decisions, which directly or indirectly impact the content or quality of recordings, have the potential to influence AVI outcomes. Yet, the available evidence is largely limited to experimental studies manipulating such elements and a small-scale observation study. We thus know little about differences in recording choices and quality

among real interviewees completing AVIs and how they impact actual AVI performance evaluations. Building on Lukacik et al. (2022), Huffcutt et al. (2011), and McColl and Michelotti's (2019) propositions, we examine the following research question and hypothesis:

RQ1. What is the prevalence of optimal vs. suboptimal applicant completion decisions, such as casual vs. professional attire, room tidiness, or quality of lighting, sound, and video, in AVIs?

H1. Optimal applicant completion decisions, including (a) wearing professional attire, (b) recording from a tidy room, or good quality of (c) lighting, (d) sound, and (e) video, are associated with higher AVI performance evaluations.

Background elements and AVI Ratings

The second mechanism Lukacik et al. (2022) proposed is that applicants recording their responses from a living room, bedroom, or home office could showcase background elements that could signal personal (and possibly legally protected) statuses, which would result in more biased AVI performance evaluations. Such a prediction was derived from prior theoretical work on interview bias, such as Derous et al.'s (2016) dual-process framework. In short, that framework argues that information collected during an interview is related to judgements and decisions via two processes: Type 1 (i.e., automatic, heuristically driven) processes lead interviewers to form initial impressions that can be biased by applicant stigmatized features, and then to engage in information processing strategies during the interview to confirm such first impressions. In contrast, Type 2 (i.e., conscious, cognitively-demanding) processes are engaged to correct initial judgements (e.g., overrule impulses to form biased impressions based on stigmas) or to update such judgements based on job-relevant information collected during the interview.

Lukacik et al. (2022) highlighted that recording an AVI from one's home might make stigmatized features, which would not be visible in in-person interviews, available to evaluators. For example, framed pictures could signal applicants' marital status, toys could showcase parental status, posters could reveal political affiliation, artwork could indicate sexual orientation, or books could divulge religious affiliation. If AVIs make stigmatized features more visible, it would facilitate Type 1 processes and increase the risk that evaluators would form biased first impressions. Lukacik e al. (2022) further argued that the asynchronous nature of AVIs might make it easier for evaluators to use Type 1 confirmatory strategies, while reducing the chance that Type 2 processes are engaged to correct such biased judgements. This argument was substantiated by empirical evidence that raters' initial impressions in AVIs (e.g., after watching one answer) can lead to skipping subsequent responses (Torres & Mejia, 2017) and are strongly associated with final interview decisions (Suen et al., 2019; Torres & Gregory, 2018).

Several studies have empirically tested Lukacik et al.'s (2022) predictions, reporting mixed evidence. Roulin, Lukacik, et al. (2023) conducted a series of three experiments manipulating AVI background elements signalling (a) parental status, (b) sexual orientation, or (c) political affiliation. They found evidence that political affiliation cues (i.e., posters and objects identifying the applicant as a Democrat vs. Republican) influenced AVI ratings, such that applicants were evaluated more negatively if they supported a different political party than the evaluator. Parental status cues played a small role, albeit opposite predictions (i.e., more positive ratings for parents). However, they found no effect of background elements signalling sexual orientation. The absence of effect for sexual orientation elements was later replicated in a German study (Basch et al., 2024). Yet that study found some biasing effects for background cues about religion affiliation, with lower performance evaluations for an applicant displaying items signalling their affiliation with Islam compared to no religious affiliation.

In summary, applicants' background content revealing personal information seems to have mixed effects on AVI performance evaluations. However, and importantly, all these studies have relied on experimental designs with actors delivering scripted responses in front of a manipulated background. It remains unclear how prevalent such background elements are in recording from real interviewees completing AVIs. And, when visible, to what extent these elements influence AVI performance evaluations. Building on Lukacik et al.'s (2022) propositions, we examine the following research question and hypothesis:

RQ2: What is the prevalence of applicant background elements revealing personal information, such as 2SLGBTQI+, political, or religious affiliation, and family status, in AVIs?

H2. The presence of applicant background elements revealing personal information, such as (a) 2SLGBTQI+, (b) political, or (c) religious affiliation, and (d) family status, are associated with lower AVI performance evaluations.

The role of evaluation standardization in AVI ratings

The personnel selection literature has long established the benefits of structured over unstructured employment interviews (Campion et al., 1997; Levashina et al., 2014). Structured interviews are now regarded as one of the most valid selection methods (Sackett et al., 2022), and they demonstrate criterion-related validity for both task and contextual performance (Wingate et al., 2025). There is also ample evidence that more structured interviews can help reduce various forms of bias or discrimination, including gender bias (Kith et al., 2022), race bias (Dahlke & Sackett, 2017), or demographic similarity effects more broadly (McCarthy et al., 2010). The level of interview structure can be determined by how many structure elements are included. Levashina et al.'s (2014) review identified 18 unique structure elements, with some elements being used more frequently than others in in-person interview practice (Chapman & Zweig, 2005; Roulin et al., 2019).

One of the strengths of AVIs is that they incorporate several structure elements by default (Lukacik e al., 2022). For example, because of their asynchronous and one-way nature, AVIs impose that the same questions are asked in the same order to all applicants for a given job, that all responses are recorded, and prevent rapport building or prompting attempts by interviewers, or questions by applicants. In theory, those elements should make AVIs more reliable, valid, and bias-resistant (Lukacik et al., 2022). Yet, despite these benefits inherent to all AVIs, the way a particular AVI is designed or implemented can integrate or omit other structure elements. Most importantly, the level of evaluation standardization can vary extensively. Levashina et al. (2014) recommended that each question be rated individually by multiple evaluators, using BARS, with descriptive notes being taken, and statistically combining scores to make decisions. AVIs can facilitate the use of all these elements, as video recordings of each individual's responses are available to be rated by multiple evaluators. Many AVI providers and organizations have even pushed the standardization further by automatizing it via machine-learning algorithms (Hickman et al., 2022; Koutsoumpis et al., 2024; Liff et al., 2024). However, it is still possible for AVIs to be evaluated more holistically, for instance, by asking evaluators to watch all recorded responses and then provide an overall evaluation of the applicant's performance.

Although decision-makers tend to prefer approaches that provide them with more autonomy (e.g., holistic approaches), hiring decisions are more valid when they rely on statistical approaches (i.e., mechanical rules to combine scores) than on holistic judgements (e.g., Neumann et al., 2022, 2023). Lubbe and Nitsche (2019) also showed that using BARS helped reduce assimilation or contrast effects when evaluating recorded interviews. Yet, to the best of our knowledge, there is no empirical evidence about the benefits of a more standardized evaluation system (or a mechanical approach) specifically in AVIs. It is thus unclear if (or how) more structured AVI evaluations could help reduce the potentially biasing effects of applicant completion decisions, recording quality, and background elements.

We examine the role of evaluation standardization in several ways in the present manuscript. Our first study employed a semi-structured evaluation approach, with specific criteria evaluated at the end of the AVI only. Our second study directly compared two levels of evaluation standardization: (1) a structured evaluation, where trained raters assessed each candidate after each interview question on separate job competencies; (2) an unstructured evaluation, where trained raters assessed each candidate holistically, giving a single interview performance rating at the end of the entire interview. Based on the evidence about the benefits of structured interviews in general, we predict:

H3. AVIs with more structured (i.e., standardized) evaluations are less prone to the biasing effects of applicant completion decisions, recording quality, and background elements.

Overview of studies

We present below two complementary empirical studies. Study 1 relies on a sample of 634 U.S. residents recruited from an online panel, completing a mock AVI. Their video responses were coded for completion decisions, recording quality, and background elements by trained research assistants, and (semi-structured) performance evaluations were provided by a group of professional hiring managers. This first study was used to provide an initial examination of Research Questions 1–2 and Hypotheses 1–2, as well as explore the effect of other applicant characteristics examined in traditional interviews (i.e., attractiveness, age, and sex). Study 2 provides further testing of Research Questions 1–2 and Hypotheses 1–2, as well as an exploration of the effects of sex, using a sample of 523 North American applicants completing a high-stakes AVI as part of the selection

process to enter competitive health sciences programmes (e.g., medical school). In addition, two independent groups of trained research assistants provided performance evaluations using an unstructured vs. structured approach, thus allowing us to test Hypothesis 3.

Study 1 - applicant completion decisions, recording quality, and background elements, and performance evaluations in low-stakes AVIs

Method

Participants and procedure

Participants ($N = 634$) were recruited through the online platform Prolific (Peer et al., 2017) to take part in a simulated selection interview for a fictitious management traineeship position (after excluding 159 participants with incomplete responses, who failed attention checks, did not take the interview seriously, with corrupted video files, etc.). The simulated interview was conducted using an AVI through a new interview platform that was developed for the purposes of another research project. The sample included 305 men, 301 women, and 26 non-binary participants. Average age was 36.69 ($SD = 11.87$). Participants were residing in the United States and were mainly White ($n = 452$), Black or African American ($n = 74$), Hispanic or Latino ($n = 47$), Asian ($n = 31$), another ethnicity ($n = 25$), or did not disclose ethnicity ($n = 5$). Of these participants, 6 had not finished high-school, 75 were high-school graduates, 213 were college graduates, 229 had a Bachelor's degree, 88 had a Master's degree, 16 had a doctorate, and 7 did not disclose education level. Participants had on average 15.93 years of working experience ($SD = 11.32$), they had applied to 18.76 jobs on average during the past 2 years ($SD = 55.9$), and they had participated on average in 1.12 prior video interviews ($SD = 2.51$).

The AVI included six questions, all presented in text format. Participants had as much time as they wished to prepare their answers, their recorded responses had to be between 30 s and 3 min in length, with only one attempt per question (i.e., no re-recording). During the interview, participants responded to six past behaviour interview questions, two of which were generic (but selected based on their frequency of use by Dutch hiring professionals, and their relevance to assess personality across multiple jobs judged by four personality experts) and four were related to personality traits. Once the video interviews were completed, a group of trained raters assessed the recording content and background information (e.g., professional attire, quality of audio, presence of awards, signs of family status), as well as attractiveness. Another group of (junior) professional hiring managers assessed the job competencies and overall performance evaluations of participants. All raters completed a five-h training session, which included 2 h of group training, 2 h rating 10 sample participants, and a final 1-h follow-up training. The content of the training mainly covered the task itself, the definitions of the job competencies, as well as best practices in rating behaviour.[1]

Measures

Completion decisions, recording quality, and background information. A group of three trained raters (two men, one woman; all Master's students in Psychology at a Dutch university) coded the content of the videos based on thin slices. Raters watched four random thin slices (2–3 s) from the six videos. One thin slice was based on the first video, another was based on the last video, and the remaining two thin slices were based on a random selection of the remaining videos. Coding was performed for the following variables: In terms of AVI completion decisions and recording quality, we coded for (a) *Attire*, that is, whether participants were properly dressed up for the interview (4-point scale; 0 = Really unprofessional; 1 = Casual; 2 = Business casual; 3 = Business formal; note that we did not request Prolific participants to dress up for the interview); (b) *Room tidiness*, that is, how tidy or untidy was the place that participants were taking the interview from (3-point scale: 0 = Very messy room; 1 = Slightly messy room; 2 = Tidy room); (c) *Light quality* (4-point rating scale: 0 = Low quality lighting; 1 = Sub-par quality lighting; 2 = Good-quality lighting; 3 = Great-quality lighting); (d) *Video quality* (3-point scale: 0 = Very low-quality video; 1 = Sub-optimal quality video; 2 = Good-quality video; note that the resolution of each video was 640 × 480 pixels); and (e) *Audio quality* (3-point scale: 0 = Very poor audio quality; 1 = Sub-optimal audio quality; 2 = Good audio quality).

Further, we coded for the presence of AVI background elements that signalled (f) *sexual orientation*, that is, whether raters could visibility tell that participants belonged to the 2SLGBTQI+ community through participants' clothing or background items (dichotomous scale: 0 = No, 1 = Yes); (g) *Political affiliation*, that is, whether raters could visibly tell that participants belonged to a political party (0 = No, 1 = Yes); (h) *Religious affiliation*, that is, whether raters could visibly tell that participants belonged to a religious group (0 = No, 1 = Yes); and (i) *Family status*, that is, whether raters could visibly tell that participants had

children (0 = No, 1 = Yes). When coding dichotomous variables (i.e., sexual orientation, political and religious affiliation, family status), variables were coded as "present" when the majority of raters (two out of three) coded for the presence of the relevant affiliation. Table 1 presents the frequencies for all background variables, and Supplementary Table S2 provides ICCs for each category.[2]

Attractiveness. A group of five trained raters (the three raters who also rated background information as well as two research assistants) assessed physical attractiveness based on thin slices (similar to background information). Raters used a 5-point scale (1 = Very unattractive; 2 = Low; 3 = Average; 4 = high; 5 = Very attractive), allowing to register up to one decimal point (e.g., "3.2" was possible). The average attractiveness was 2.85 ($SD = 0.44$), and inter-rater agreement was $ICC_{2,5} = 0.79$.

Performance evaluations. Another group of five raters (i.e., professional hiring managers; two men, three women; all held a Master's degree; average age = 27.40, $SD = 5.04$; average recruitment experience 2.2 years, $SD = 1.6$ years) assessed AVI performance. More specifically, raters assessed four job competencies (taken from the manual of a Dutch consultancy company and associated with the four personality traits the questions were designed to assess) and an overall performance score. The job competencies were the following: (a) *Integrity*, that is, the extent to which the candidate inspired trust, displayed integrity in their interaction with others, treated others fairly, and adhered to high ethical standards; (b) *Collegiality*, that is, the extent to which the candidate was open to and showed an interest in others and was willing to adapt their own activities to help others in their work; (c) *Social versatility*, that is, the extent to which the candidate had the ability to adapt their behaviour in a wide range of social situations in order to function effectively in different types of companies; and (d) *Development orientation*, that is, the extent to which the candidate was willing to exert themselves in order to broaden and deepen knowledge and skills and to gain new experiences in order to grow professionally and increase the quality of their work. *Overall performance evaluation* was defined as the extent to which the candidate would be able to fulfil the requirements of the management traineeship position. Evaluations were provided after the raters had watched the responses to all six interview questions.

Due to the high workload, raters assessed a varying number of participants. However, in total, each participant received two independent performance evaluations (one-fourth of participants received three).

Table 1. Frequencies for all coded completion decisions and background information.

			Study 1		Study 2	
Variable	Levels	Level description	Frequency	Percentage	Frequency	Percentage
AVI completion decisions and recording quality						
Attire	0	Unprofessional	4	0.6	4	0.8
	1	Casual	621	97.9	177	33.8
	2	Business casual	8	1.3	247	47.2
	3	Business formal	1	0.2	95	18.2
Room tidiness	0	Very messy	40	6.3	5	1
	1	Slightly messy	133	21.0	9	1.7
	2	Tidy	461	72.7	509	97.3
Lighting quality	0	Low	7	1.1	1	0.2
	1	Sub-par	46	7.3	36	6.9
	2	Good	275	43.4	154	29.4
	3	Great	306	48.3	332	63.5
Video quality	0	Very low	1	0.2	16	3.1
	1	Sub-optimal	65	10.2	145	27.7
	2	Good	568	89.6	362	69.2
Audio quality	0	Very poor	1	0.2	4	0.8
	1	Sub-optimal	22	3.5	31	5.9
	2	Good	611	96.4	488	93.9
AVI background elements						
2SLGBTQI+	0	No	634	100.0	517	98.9
	1	Yes	0	0	6	1.1
Political affiliation	0	No	634	100.0	523	100.0
	1	Yes	0	0	0	0
Religious affiliation	0	No	631	99.5	514	98.3
	1	Yes	3	0.5	9	1.7
Family status	0	No	630	99.4	516	98.7
	1	Yes	4	0.6	7	1.3

Responses were provided using BARS (see Supplementary Table S3) and responses were given on a 5-point scale (1 = Very low; 2 = Low; 3 = Average; 4 = high; 5 = Very high), allowing to register up to one decimal point (e.g., "3.2" was possible). Results of an exploratory principal component analysis (PCA) confirmed that the four job competencies and the overall performance evaluation loaded on a single component, which explained 77.09% of the variance. As a result, we retained the average of the five variables as an overall "performance evaluation" score. For overall performance, the inter-rater agreement was $ICC_{1,2} = .55$ (for the sub-sample of $n = 149$ participants who received three interview performance evaluations, $ICC_{2,3} = .60$; see Supplementary Table S2).

Results and discussion

Table 2 presents the descriptive statistics and intercorrelations among all variables (see also Supplementary Table 4 for information about skewness or kurtosis). In relation to RQs 1–2, the videos were by and large of good quality (91.6% had good or great light, 89.6% good video, 96.4% good audio), and participants recorded their responses in mostly tidy rooms (72.7%), yet were largely casually dressed (97.9%). The AVIs contained almost no visible signs of group affiliation (2SLGBTQI+, political, religious), or family status (all less than 1%). Overall, the results suggest that interviewees generally made positive AVI completion decisions, ensuring that they recorded good-quality videos in a generally tidy environment, with a neutral background. While most interviewees dressed casually, this is likely due to the use of a Prolific sample, and the absence of specific instructions to dress professionally for the study. Study 2 will explore this issue further with high-stakes AVIs.

AVI performance evaluation was not significantly correlated with any of the completion decisions nor background variables, except for video quality ($r = .10$), but was correlated with attractiveness ($r = .22$) and sex ($r = .17$). In other words, participants perceived to have good video quality, those rated as high on physical attractiveness, and women received higher evaluations. However, none of the other AVI completion decisions or background elements were related to interview outcomes. Hence, we only found some support for Hypothesis 1e, but no support for H1a-d or H2. These findings are generally aligned with preliminary evidence from experimental work showing little impact of background information (e.g., Basch et al., 2024; Roulin, Lukacik, et al., 2023). Yet, the observed effects for completion decisions were largely small compared to experimental studies (e.g., Powell et al., 2023; Suen et al., 2019), except for attractiveness and video quality. In sharp contrast with meta-analytical findings for in-person interviews (e.g., $r = .36$ in Barrick et al., 2009; and $r = .54$ in; Martin-Raugh et al., 2023), we also found that performance evaluations were unrelated to how professionally people dressed for their mock AVI. Interestingly, although significant, the effect of attractiveness was smaller compared to meta-analytical findings for in-person interviews (e.g., $r = .37$ in Barrick et al., 2009). Although women were evaluated slightly more

Table 2. Descriptive statistics and correlations between performance evaluations, completion decisions and background information (Study 1).

Variable	M	SD	1	2	3	4	5	6	7	8	9	10	11	12
1. Performance evaluation	3.22	0.45	–											
AVI completion decisions and recording quality														
2. Attire	0.99	0.13	.05	–										
3. Room tidiness	1.62	0.67	.03	−.02	–									
4. Lighting quality	2.29	0.79	.04	.04	.00	–								
5. Video quality	1.95	0.23	.10*	.02	−.01	.26***	–							
6. Audio quality	1.98	0.15	.06	−.19***	.00	.06	−.01	–						
AVI background elements														
7. 2SLGBTQI+[†]	0.00	0.00	–	–	–	–	–	–	–					
8. Political affiliation[†]	0.00	0.02	–	–	–	–	–	–	–	–				
9. Religious affiliation	0.00	0.05	−.05	.00	.00	−.01	.02	.01	–	–	–			
10. Family status	0.01	0.07	.01	.00	−.12**	−.02	.03	.02	–	–	−.01	–		
Other applicant characteristics														
11. Sex	1.50	0.50	.17***	−.09*	.03	−.01	.01	.06	–	–	.07	.08*	–	
12. Age	36.69	11.87	.05	.02	−.01	−.01	−.03	−.06	–	–	.04	.01	.06	–
13. Attractiveness	2.85	0.44	.22***	.05	.06	.11**	.13***	.01	–	–	−.04	−.01	−.05	−.35***

$N = 634$. [†]Pearson r correlation coefficients were not calculated due to zero variation (all values were coded as "0", that is, absence of 2SLGBTQI+ and political affiliation information). Sex coded as man = 1, woman = 2 (excluding non-binary in this table) See Table 1 for descriptions of video coding categories (variables 2–10); correlations with sex have been calculated using Spearman's rho. *$p < .05$; **$p < .01$; ***$p < .001$.

positively than men, age was unrelated to AVI outcomes, which is in line with previous AVI studies (e.g., Koutsoumpis et al., 2024).

Overall, Study 1 findings suggest little cause for concern in terms of AVI bias caused by background elements, which is potentially reassuring for organizations. That said, practitioners still need to be aware of potential issues associated with video quality or applicant demographics. The small effects can also be partly explained by generally "good" completion decisions by participants and the very low prevalence of background cues signalling personal information. Study 2 will examine whether these promising findings replicate in high-stakes settings.

Study 2 – applicant completion decisions, recording quality, background elements, and (un)structured performance evaluations in high-stakes AVIs

Method

Participants and procedure

Data were obtained from 523 applicants, who completed a 3-question high-stakes AVI and consented for their videos to be used for research purposes (out of 3,087 applicants contacted). The AVI questions were all presented in text format. Applicants had 30 s to prepare their answers and 120 s to record it, with only one attempt per question (i.e., no re-recording). All applicants were U.S. or Canadian residents who completed the AVI called "*Snapshot*" as part of their application process for various health sciences programmes (e.g., medicine, dentistry, nursing) managed by Acuity Insights. Demographic information (sex, age) was not collected as part of the application process. One of the co-authors thus coded participants (perceived) sex (346 men, 176 women, 1 non-identified) based on the videos. Due to the higher likelihood of errors in coding age, we chose not to include this demographic variable in this study. The AVI questions were designed to help assess applicants' communication skills, self-reflection, and motivation, and varied slightly depending on the programme the applicant was interviewing for. Example questions include "*Describe your ideal learning environment or academic program. Why are you well suited for it?*"; "*Consider one experience that has most impacted where you are in your life today. Why was it so impactful?*"; "*What is one thing about yourself that you are working on, and one thing you are proud of?*" (see Online Supplement for all questions).

Different raters reviewed the video responses and evaluated different elements. One of the co-authors assessed background information (e.g., professional attire, quality of audio, signs of family status). Unlike in Study 1, applicant attractiveness was not rated. This decision was made because of the high-stakes nature of the video material and the fact that asking real applicants to consent for their attractiveness to be rated for research purposes could (mistakenly) suggest that such information was used as part of the decision-making process. A group of two trained raters provided structured performance ratings, whereas another group of three raters provided unstructured performance ratings. None of these ratings were used for the actual selection process.

Measures

Completion decisions, recording quality, and background information. The content of applicants' videos was assessed similarly to Study 1. For instance, the same variables were rated identically to Study 1: *Attire, Room tidiness*, *Lighting quality, Video quality, Audio quality, Sexual orientation*, *Political affiliation*, *Religious affiliation*, and *Family status*. However, given the objective nature of the elements to code, the coding was performed by one individual (i.e., one of the co-authors) who watched the full video of the first response from each applicant (and then used a similar thin slice approach to sample/check content from responses to the other two questions). Given that applicants were interviewing for admission into health sciences education programmes, we coded those dressed in scrubs (10 cases) as "casual". While it could be considered as somewhat professional attire in that context, it did not represent formal interview attire. Table 1 presents the frequencies for all background variables.[3]

Structured AVI performance evaluations. Two research assistants (both graduates from a BA with honours in Psychology) acted as raters for structured performance evaluations. They first participated in a training session to ensure that they were familiar with the *Snapshot* AVI process, they paid attention to the content and delivery of the responses by applicants (but ignored other elements), and they became acquainted with the BARS created for each question (see Supplementary Table 6). Each BARS was designed to evaluate responses from 1 (e.g., poor communication) to 7 (e.g., exceptional communication), and included detailed behavioural descriptions for each level, as well as examples of responses for levels 1, 4, and 7. During the training, the two raters also evaluated a series of practice video responses together with one of the co-authors, justified

their scores, and discrepancies were discussed. Then both raters independently rated all video responses. They were instructed to evaluate videos in small batches to avoid fatigue, and inter-rater consistency was checked regularly. $ICCs_{2,2}$ based on the entire sample were good for all questions (i.e., .80 for Q1, .83 for Q2, .82 for Q3). Scores across the three questions were ultimately aggregated to create a structured interview performance evaluation score (with an overall $ICC_{2,2} = .87$).

Unstructured AVI performance evaluations. Another group of three research assistants (two PhD students in I/O Psychology, one graduate from a BA with honours in Psychology) acted as raters for unstructured performance evaluations. They also participated in a training session similar to the structured ratings presented above. The key difference is that they were instructed to evaluate applicants' performance on one overall 1–7 scale for the entire AVI. The scale included brief anchors for each level (i.e., poor to exceptional communication, self-reflection and motivation) and broad descriptions for levels 1, 4, and 7 (see Supplementary Table 7). The three raters were assigned a first batch of the same 20 applicants (to check for inter-rater consistency). Then, they were assigned different batches of applicants, ensuring that all applicants were evaluated by two raters, with 60 cases evaluated by all three in total. Overall $ICCs_{2,3}$ based on the 60 cases coded by all three raters was .76 (and when looking at pairs of raters, with $Ns = 205$ to 232, $ICCs_{2,2}$ ranged from .58 to .63).

Results and discussion

The frequencies for the various AVI video coding categories are presented in Table 1. In relation to RQ1, findings were mostly consistent with the results of Study 1: Most applicants' recorded responses had good-quality audio (93.9%), video (69.2%), and lighting (92.9%). In contrast to Study 1 results, the majority of real applicants dressed in business casual (47.2%) or business formal (18.2%) attire, and they recorded their responses in a tidy room (97.3%). These differences could be due to different levels of motivation between low- and high-stakes AVIs. It further emphasizes that actual applicants tend to make careful AVI completion decisions. Regarding RQ2, background elements signalling otherwise-hidden statuses were rare. For instance, only 1.1% could be identified as visibly part of the 2SLGBTQI+ community, 1.3% had family status elements, 1.7% religious elements, and none displayed any political affiliation in the background of their AVI. Largely in line with the initial results from the online sample from Study 1, Study 2 confirmed that only a small proportion of applicants have elements signalling protected group status (e.g., religious affiliation, sexual orientation, family status) visible in their AVI background in practice too. The vast majority of applicants made sure they recorded their AVI with a tidy and neutral background.

Table 3 presents the descriptive statistics intercorrelations among all variables. Structured AVI performance evaluations were significantly related to having better-quality lighting ($r = .11$, $p = .02$), better audio quality ($r = .09$, $p = .03$), and more professional attire ($r = .22$, $p < .01$). Unstructured AVI performance evaluations were significantly related to having better-quality lighting ($r = .13$, $p < .01$), better video quality ($r = .12$, $p < .01$), and more professional attire ($r = .14$, $p < .01$). However, and importantly, neither structured nor unstructured evaluations were associated with any AVI background

Table 3. Descriptive statistics and correlations between performance ratings, completion decisions, and background information (Study 2).

Variable	M	SD	1	2	3	4	5	6	7	8	9	10	11
1. Structured evaluations	4.29	0.68	-										
2. Unstructured evaluations	4.64	0.91	.53***	-									
AVI completion decisions and recording quality													
3. Attire	1.83	0.72	.22***	.14**	-								
4. Room tidiness	1.96	0.23	.04	.01	.04	-							
5. Lighting quality	2.56	0.63	.11*	.13**	.16***	.17***	-						
6. Video quality	1.66	0.53	.04	.12**	.12**	.06	.37***	-					
7. Audio quality	1.93	0.29	.09*	.05	−.06	−.04	.02	−.02	-				
AVI background elements													
8. 2SLGBTQI+[†]	0.01	0.12	−.07	−.04	−.03	.02	−.03	−.05	.03	-			
9. Political affiliation[†]	0.00	0.00	-	-	-	-	-	-	-	-	-		
10. Religious affiliation	0.02	0.13	.05	.08	−.01	.02	−.05	.03	.03	−.01	-	-	
11. Family status	0.01	0.12	−.03	.02	−.02	−.13**	.00	.07	−.03	−.01	-	−.02	-
Other applicant characteristics													
12. Sex	1.34	0.47	.00	−.12**	.08	−.10*	−.04	−.04	.02	.14**	-	−.03	−.08

$N = 522$. None of the applicants had any political affiliation in the background of the AVI so political affiliation was excluded from further analysis. Sex coded as man = 1, woman = 2 (excluding non-binary in this table). See Table 1 for descriptions of video coding categories (variables 3–11). * $p < .05$; ** $p < .01$; *** $p < .001$.

elements. In other words, we found support for H1a (i.e., in high-stakes AVIs, contrary to low-stakes AVIs in Study 1) and H1c, and partial support for H1d (in unstructured interview only) and H1e (in structured interview only). However, we found no support for H1b and H2 (consistent with Study 1). We also note that applicant (coded) sex was unrelated to structured evaluations (r = .00), while it was modestly related to unstructured evaluations ($r = -.12$, $p < .01$). Interestingly, the direction of the effect was opposite to Study 1 (i.e., lower evaluations for women in Study 2, higher evaluations in Study 1). That said, these results confirm the benefits of more structured interview approaches to reduce or even eliminate sex or gender bias (Kith et al., 2022).

To examine H3, we compared the correlations for the structured vs. unstructured evaluations, using the cocor R package tool (Diedenhofen et al., 2015). The relationship between professional attire and performance evaluations was stronger for structured evaluations compared to unstructured evaluations ($z = 1.92$, $p = .03$), contrary to our predictions. In addition, while we found stronger correlations for unstructured than structured interview evaluations for video quality ($z = 1.89$, $p = .03$), there were no differences in correlations for lighting quality ($z = 0.47$, $p = .32$) or audio quality ($z = 0.94$, $p = .82$). Overall, we found no substantial support for H3. In other words, we found largely similar effects when evaluations were structured (i.e., using a behaviourally anchored rating scale – BARS) or unstructured.

Interestingly, we found that both types of evaluations were impacted by the way applicants dressed, with stronger effects for structured evaluations. Similar to Study 1, these correlations were smaller than those typically found in in-person interviews (e.g., Barrick et al., 2009; Martin-Raugh et al., 2023). Moreover, being professionally dressed for a high-stakes AVI might signal applicant preparation and motivation, and thus represent a reliable and valid indicator or applicant quality (rather than a bias). To further explore this, we looked at the correlations between attire and performance at the question level for the structured evaluations. We found slightly larger correlations for the questions assessing motivation and self-reflection (both $rs = .18$) than for the communication question ($r = .13$). Note that applicants' self-reflection was defined as "critically analyzing their own personal attributes, strengths, and experiences" and "describing general behaviors demonstrating their qualities, values, motivations, relationships, etc." It is possible that applicants who had more self-insights were also more mindful about how to dress properly for their AVI.

General discussion

Main findings and theoretical contribution

The present research empirically examined several central elements and propositions from Lukacik et al.'s (2022) model, focusing specifically on interviewees' AVI completion decisions, recording quality, and background elements. We did so through two studies, with large samples, and including both mock/low-stakes and real/high-stakes interviews. Overall, our findings have several important theoretical implications, and they contribute to the emerging literature on AVIs and technology-mediated interviews more broadly in the following ways:

First, our results illustrate that interviewees tend to carefully consider how they complete their AVIs, particularly by striving to ensure good lighting and high-quality audio and video recordings. This suggests that the types of technical issues or "signal distortion" described in previous work (e.g., McColl & Michelotti, 2019) do emerge in AVIs, but they are likely not very prevalent. Findings related to other recording decisions (Lukacik et al., 2022) highlight that the stakes of the interview matter. For example, almost all interviewees (98%) dressed casually for the mock AVIs in Study 1, but most applicants in high-stakes admission processes wore business casual (47%) or formal (18%) attire. Similarly, although most interviewees tried to record their responses in a tidy room, this was even more pronounced in the high-stakes compared to low-stakes AVIs (97% vs. 73%). Interviewees, and particularly real applicants, thus seem to adhere to an "AVI etiquette" consistent with recommendations provided in past research (e.g., Davis et al., 2020; Lee et al., 2021; Mejia & Torres, 2018). Overall, these results also suggest that participants from online panels like Prolific take research-oriented AVIs seriously, but real applicants pay even more attention to key AVI completion decisions.

Second, our research addresses calls from prior theoretical (Lukacik et al., 2022) and empirical (Roulin, Lukacik, et al., 2023) work to examine the presence of background elements that can signal personal or legally protected statuses, such as political or religious affiliation, sexual orientation, or family status, in real AVIs. We found that very few interviewees' recordings included such background elements. For instance, political affiliation elements were never observed across our two

studies (with over 1,150 interviewees in total), and the highest prevalence of religious affiliation elements was 1.7% in Study 2. This could be because interviewees are actively inspecting their video background and considering factors that could influence evaluators' judgements before starting their AVIs. This would be consistent with recommendations from research (e.g., Davis et al., 2020; Mejia & Torres, 2018; Scott & Roulin, 2024) or, most likely, advice on social media platforms[4] to keep one's background as neutral as possible.

Third, our research explored whether AVI completion decisions, recording quality, and background elements influence interview performance evaluations (Lukacik et al., 2022). Prior empirical attempts to test such predictions have been largely limited to experimental research (e.g., Basch et al. 2024; Powell et al. 2023; Roulin, Lukacik, et al. 2023) and reported mixed results. Regarding AVI completion decisions, room tidiness was unrelated to AVI performance evaluations across both studies, in contrast to prior experimental findings (Powell et al., 2023). It might be that interviewees' recordings need to display an extremely messy room (like in Powell et al., 2023) to have meaningful impact on AVI evaluations, whereas slightly disorganized settings remain unnoticed. Or, perhaps our findings confirm that norms about what is considered to be an "acceptable" recording setting have evolved towards more leniency, similar to Scott and Roulin's (2024) findings about recording location and blurring features.

Interestingly, attire was only modestly related to performance evaluations in Study 2 but unrelated to evaluations in Study 1. As noted above, the relationships observed here are much smaller than those reported in traditional in-person interviews (Barrick et al., 2009; Martin-Raugh et al., 2023). This could represent preliminary evidence that norms differ between in-person and technology-mediated interviews like AVIs, such that business attire is not as strongly expected in the latter. This interpretation is consistent with recent findings showing that employees working remotely were viewed as more authentic and engaged when dressed in more casual "home attire" than professional attire (Bailey et al., 2022). Our findings therefore highlight that more research could explore what type of attire is considered "appropriate" (by applicants, but also by hiring professionals) in AVIs. As noted by an anonymous reviewer, some applicants might perceive wearing formal attire (suit and tie, blouse, etc.) as awkward when recording an AVI from home. But perhaps expectations differ by job, industry, culture, etc. Alternatively, it could also be that an applicant's attire is less visible or noticeable in an AVI compared to an in-person interview. Expectations for professional attire in AVIs might also be lower for students interviewing for admission to education programmes (as in our Study 2) than for job applicants.

Overall, the fact that attire was slightly more strongly associated with more structured (vs. unstructured) performance evaluations in Study 2 was somewhat surprising. Indeed, the effect of professional appearance on in-person interview evaluations tends to be smaller in more structured interviews (Barrick et al., 2009). Yet, as noted above, casual dress is arguably now the norm for virtual work, and possibly also virtual interviewing. In addition, because Study 2 took place in a high-stakes selection setting and not a controlled experiment, it could be that interviewees who chose to wear business-casual or business-formal attire were more motivated, better prepared, or simply more professional than those completing an AVI, while dressed casually. Such superior motivation and preparation might have led those applicants to provide higher quality responses and demonstrate stronger performance overall. In other words, wearing professional attire in an AVI could represent a valid signal of other job-relevant characteristics (Bangerter et al., 2012), which should be rewarded by evaluators.

Fourth, in terms of AVI recording quality, we found null or small (and partly inconsistent) relationships across the two studies. Video quality was modestly positively related to the (semi-structured) performance evaluations in Study 1 and the unstructured (but not structured) evaluations in Study 2. Audio quality was modestly associated with structured evaluations in Study 2 only. Lighting quality was modestly related to both types of evaluations in Study 2, but not Study 1 evaluations. Overall, these findings suggest that technical issues in AVI recordings, while meaningful, are unlikely to have a major impact on interview evaluations, in line with preliminary evidence (Koutsoumpis et al., 2024) but contrary to theoretical predictions (Lukacik et al., 2022; McColl & Michelotti, 2019). That said, serious technical issues in AVIs (i.e., major signal distortions), such as recordings without any sound, could prevent evaluators from judging interviewees' responses altogether. It is unclear how such situations are addressed by evaluators, organizations, or AVI providers. For example, are interviewees evaluated only based on their other responses? Or are they offered a second opportunity to complete their AVI (or problematic questions)? How does that influence the AVI outcomes, in terms of fairness, reliability, or validity? Are such issues more likely to occur for applicants from equity-deserving groups (e.g., racial minorities or lower socio-economic status individuals, who might not have access to reliable technology or high-speed internet), and could that create adverse

impact? Those are important questions that future research should examine.

Fifth, we found no relationship between background elements signalling personal or protected statuses and AVI performance evaluations across both studies. These results complement preliminary experimental research (Basch et al., 2024; Roulin, Lukacik, et al., 2023), which reported largely mixed findings. For instance, Basch et al. (2024) found some bias against an applicant displaying signals of Islamic religious affiliation (vs. no signal of affiliation), whereas Roulin, Lukacik, et al. (2023) found strong bias against an applicant displaying support for a political party that differed from the evaluator's. These results were not replicated in our real interviews. It could be that such background elements need to be very salient or visible (like in the experiments discussed above) to influence evaluators' judgements. However, it is also important to reiterate that none of our interviews (across both studies) included signals of political affiliation, and very few included signals of religious affiliation (0.5% and 1.7% in Studies 1–2). It is thus possible that background elements can play a role in AVI evaluations, but the signal was too weak to be impactful in our studies. The use of psychology students as raters in Study 2, who might be more aware of biasing factors than hiring managers in Study 1, could have also contributed to these differences.

Finally, we examined applicant characteristics that have regularly been identified as biasing factors in traditional in-person interviews, such as a sex, age, and attractiveness. We found no effect of age on performance evaluations in Study 1 (age data was not available in Study 2). We found slightly higher evaluations for women in Study 1, but slightly lower unstructured evaluations for women in Study 2, and no sex effect for structured evaluations in Study 2. The difference in direction could be explained by the composition of the rating teams: Study 1 raters were three women and two men, whereas Study 2 were all men. Given that less structured interviews are prone to demographic similarity bias (McCarthy et al., 2010), it might explain the different results across the two studies. These findings suggest that AVIs might remain prone to sex-based biases to some extent, but a more standardized evaluation approach using BARS helps eliminate such biases. This is consistent with both the theoretical benefits of structured interviews (e.g., Levashina et al., 2014) and past in-person interview findings (e.g., Kith et al., 2022). Attractiveness was only coded in the mock AVIs in Study 1, and it was positively related to performance evaluations. Yet, the magnitude of this effect is smaller compared to in-person interview findings (Barrick et al., 2009). Overall, our findings show that although AVIs are not bias-free, they might be somewhat more resistant to bias associated with applicant characteristics than in-person interviews in general, partly because they include several interview structure components by design (Lukacik et al., 2022).

More generally, our results combined with prior research can be seen as promising about the potential validity of AVI. Indeed, we found only small relationships between applicants' characteristics, completion decision, or background elements and AVI evaluations. In addition, structured interviews demonstrate superior criterion-related validity (Sackett et al., 2022; Wingate et al., 2025) and all AVIs are structured-by-design (Lukacik et al., 2022). Moreover, although we found mixed evidence about bias reduction via standardized evaluations in our studies, general interview research suggests that mechanical predictions (e.g., Neumann et al., 2022, 2023) or using BARS (Lubbe & Nitsche, 2019) can be beneficial. AVI evaluators are required to watch recordings for each question individually. This makes AVIs particularly well-suited for using BARS to rate each recorded response and rely on mechanical predictions to make decisions. That said, we want to emphasize that direct evidence for the criterion-related validity of AVIs is very limited (see Liff et al., 2024 for a rare exception). Given their popularity (e.g., Dunlop et al., 2022), more research examining the validity of AVIs is necessary. Such research could explore AVI validity in general, whether it differs depending on design decisions (e.g., preparation time, re-recording – Lukacik et al., 2022), as well as how it compares to in-person or other technology-mediated interviews.

Practical implications

The findings from our studies have valuable implications for both job applicants and hiring organizations. Background information signalling protected status was largely absent and had no relationship with AVI performance evaluations. These results should be reassuring for both parties because they suggest that these elements are unlikely to unfairly impact applicants because of their religious beliefs or political views. That said, our findings might be due to the very low prevalence of protected status cues in AVI backgrounds. Because prior research has found such factors to be a potential source of bias (e.g., Roulin, Lukacik, et al., 2023), we would still encourage job applicants to carefully consider the risks of displaying such signals in the background of their AVIs. Similarly, our findings suggest that most AVI completion decisions are weakly related to interview evaluations. Yet, they can still play a small role in AVI performance evaluations, and they could possibly

be the difference between moving to the next stage of the selection process or not. For example, the results of Study 2 showed that applicants wearing formal attire obtained substantially higher performance evaluations than those wearing casual attire, in both unstructured (*M*s = 4.94 vs. 4.57) and structured evaluations (*M*s = 4.52 vs. 4.13). While room tidiness did not play a major role in our studies, it did in prior work (Powell et al., 2023). Recording quality elements also played a modest role. Taken together, we recommend that applicants carefully consider how they dress, where they record their AVI, and (if possible) the technology they use to do so.

For hiring organizations, our findings generally showed that AVI completion decisions, recording quality, background elements, or applicant characteristics were either unrelated or weakly related to interview evaluations. Overall, this should provide some comfort for organizations using (or considering using) AVIs as part of their hiring process, although it does not mean that AVIs are bias-free and other methods (e.g., phone interviews) might still have value. For instance, the biasing effects observed for factors like sex or attractiveness were largely smaller than those reported for in-person interviews (Barrick et al., 2009; Martin-Raugh et al., 2023), which confirms the inherent benefits of highly-structured-by-design AVIs (Lukacik et al., 2022). We did not find large differences between the more or less standardized evaluations in Study 2 overall. Yet, we still found small but meaningful benefits for the more structured BARS approach, for instance, to help eliminate differences in ratings between men and women. We would thus encourage organizations to use more standardized evaluations, ideally with question-level BARS. Beyond bias reduction, a more standardized evaluation approach can have additional benefits in terms of interview reliability, validity, or legal defensibility (Chapman & Zweig, 2005; Levashina et al., 2014).

Limitations and future research directions

This research has several limitations that need to be acknowledged and could help pave the way for future research. First, Study 1 relied on mock AVIs with online panel participants. This limitation was partly addressed by using high-stakes AVIs in Study 2. However, in Study 2, we were not able to obtain applicant data (e.g., age, sex – although we coded the latter) or to rate attractiveness. Therefore, we encourage future research to examine whether these applicant characteristics play a role in high-stakes AVIs too.

Second, we obtained ratings from hiring professionals in Study 1, but evaluations were provided by trained psychology student raters in Study 2. This was necessary to examine the differences between higher and lower levels of evaluation standardization. But it means that while the interviews in Study 2 were high stakes, the ratings were not. We were able to obtain evaluations used in the actual admission process (from health science programmes) for a sub-sample of applicants (n = 133). These scores were significantly correlated with both unstructured (r = .42, p < .001) and structured (r = .37, p < .001) evaluations from our trained raters, providing evidence of the external validity of our findings. That said, future research could examine the benefits of standardized evaluation in AVIs with hiring managers in charge of making actual screening or selection decisions.

Third, the ICCs for performance evaluations in Study 1 were substantially lower compared to evaluations obtained for Study 2 (overall $ICC_{1,2}$ = .55 in Study 1, $ICC_{2,3}$ = .76 for unstandardized evaluations and $ICC_{2,2}$ = .87 for standardized evaluations in Study 2). This could be because Study 1 interviewees responded to questions primarily designed to measure personality traits, but raters were asked to evaluate performance on four job-related competencies. While the personality traits and competencies were conceptually related (e.g., honesty-humility and integrity; extraversion and collegiality), raters possibly did not have enough information to infer job competencies from the responses. In addition, ICC(1), which we used in Study 1, typically returns lower values compared to ICC(2), which we used in Study 2. Yet, we encourage future research to ensure that AVI questions and BARS are created together and designed to assess the same job-related competencies.

Fourth, the AVIs in both studies did not incorporate all key components of structured interviews, as recommended in the literature (e.g., Levashina et al., 2014). For instance, interview questions were not based on in-depth job analyses, and not all questions were among the "better" question types recommended in the literature (e.g., Campion et al., 1997). More precisely, Study 1 included two initial questions that were more generic. In addition, although the last four questions were behavioural in nature, they asked interviewees how they typically behaved in a general context (e.g., when they joined a new team) rather than about one specific behaviour in a particular situation (e.g., asking them to describe a unique experience when they joined a specific team, which is how "better" types of past behavioural questions are usually formatted; Campion et al., 1997). All this might have also contributed to the low ICC, because performance ratings were based on the entire interview. Similarly, the AVI questions used in Study 2 were not behavioural or situational, but they were the actual questions used as part of the admission

process. Overall, we encourage future research to focus on "better" question types when examining biases in AVIs, as well as to compare biases between past-behaviour and future-oriented questions. Relatedly, the BARS used in Study 2 were created ad-hoc, for the purpose of this research. Our results, and especially the lack of support for H3, could thus be due to using BARS to evaluate responses to non-behavioural interview questions. Future studies examining AVI bias should rely on a combination of behavioural questions and associated BARS.

Fifth, when examining the role played by background elements revealing personal information (e.g., sexual orientation, political or religious affiliation, and family status), we only focused on interviewees' characteristics. However, how the presence of such elements impacts AVI evaluations might depend on whether the evaluator shares the same characteristics as the applicant (e.g., the negative effect would be larger if they support different political parties – see Roulin, Lukacik, et al., 2023). Unfortunately, our group of trained raters was too small (i.e., five per study) to examine such effects. This is a valuable avenue for future research to explore.

Finally, our studies were focused on examining the relationship between AVI completion decisions, recording quality, or background elements and human ratings of performance. However, many AVI providers and organizations rely on automated assessments based on machine-learning algorithms. There is evidence for the reliability and validity of automated methods (Hickman et al., 2022; Koutsoumpis et al., 2024; Liff et al., 2024), what type or quantity of data is required to obtain reliable automated assessments (Hickman, Liff et al., 2024), and theoretical discussions about different sources of bias that could be associated with artificial intelligence or machine-learning models used in assessments (e.g., Landers & Behrend, 2022; Tay et al., 2022; Tippins et al., 2021). But much less research has examined whether automated assessments in AVIs are more or less prone to bias, with some exceptions looking specifically at personality assessments in AVIs (e.g., Koutsoumpis et al., 2024) or bias in automated speech recognition based on race or accent (e.g., Hickman, Langer, et al., 2024). Future research could explore whether the AVI elements examined in the present study are associated with different types of automated assessments (e.g., relying on different machine-learning algorithms or large language models).

Conclusion

This research empirically examined theoretical propositions (Lukacik et al., 2022; McColl & Michelotti, 2019) suggesting that AVI completion decisions, recording quality, or background elements could influence (i.e., bias) performance evaluations across two complementary studies. We found that interviewees completion decisions depend on the stakes of the AVIs, with recordings done in more formal attire and tidier rooms for high-stakes AVIs. Such decisions can influence AVI evaluations, with more formal attire being rewarded in high-stakes interviews, possibly because it signals higher motivation or preparation. Signal distortions related to recording quality were fairly rare and were only modestly related to AVI evaluations. Very few interviewees recorded their responses in front of backgrounds signalling personal or protected statuses, and the presence of those elements was unrelated to AVI evaluations. More structured (i.e., standardized) evaluations were not systematically associated with less bias, although it did help reduce the impact of interviewees' demographic characteristics (i.e., sex). Yet, more research is needed to examine potential AVI bias with high-stakes decisions or when more automated assessments approaches are used.

Notes

1. Training materials can be accessed online: https://osf.io/gsj46/?view_only=4e928d5f05334144908dd8a6983fa429
2. We performed a principal component analysis (PCA) to test whether the background information loaded on one or multiple components. 2SLGBTQI+and political affiliation were excluded because they had no variance. While a robust PCA would account for skewed distributions (Hubert et al., 2009), which was the case for most of the background variables, the model failed to converge. Results from a regular PCA are presented in Supplementary Table 1 of our online supplement (see: https://osf.io/cfs7d/?view_only=4b46fa7c86854b2b85b7e41d4215f7dc). They suggested four components explaining 58.18% of the variance: (1) light and video quality; (2) audio quality and attire (reverse scored); (3) room tidiness; and (4) awards, religious affiliation, and family status. Yet, component 2 seemed counter-intuitive, awards and family status had secondary loadings on component 3, and 2SLGBTQI+and political affiliation were excluded. Thus, we decided to treat each variable separately in subsequent analyses, and we excluded awards – which did not clearly fit a type of element.
3. Similar to Study 1, we performed a PCA on the background (except political affiliation). Results are presented in Supplementary Table 5. They suggested four components explaining 54.82% of the variance: (1) dress and video quality; (2) family status (reverse scored) and room tidiness; (3) religious affiliation and audio quality; and (4) awards and sexual orientation. Since most components seemed counter-intuitive, we decided to treat each variable separately in subsequent analyses – like in Study 1.

4. For instance, many YouTube videos providing advice about video interviews list controlling one's surroundings as a key success factor, including very popular videos (e.g., over 1.4 million views for this one, as of Oct 2024: https://www.youtube.com/watch?v=J2VnJOw5Cd0&t=243s).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was supported by a Mitacs Accelerate Grant (IT30679). The authors thank Damian Canagasuriam, Kaiheng Liang, Zhixin Liu, Mehdi Salimian Rizi, and Mehak Tekchandani for their help with data scoring.

ORCID

Nicolas Roulin http://orcid.org/0000-0003-2797-7359

Data availability statement

Datasets and supplementary materials for this project are available in OSF: https://osf.io/cfs7d/?view_only=4b46fa7c86854b2b85b7e41d4215f7dc

References

Arseneault, R., & Roulin, N. (2024). Examining discrimination in asynchronous video interviews: Does cultural distance based on country-of-origin matter? *Applied Psychology*, *73*(1), 185–214. https://doi.org/10.1111/apps.12471

Bailey, E. R., Horton, C. B., & Galinsky, A. D. (2022). Enclothed harmony or enclothed dissonance? The effect of attire on the authenticity, power, and engagement of remote workers. *Academy of Management Discoveries*, *8*(3), 341–356. https://doi.org/10.5465/amd.2021.0081

Bangerter, A., Roulin, N., & König, C. J. (2012). Personnel selection as a signaling game. *Journal of Applied Psychology*, *97*(4), 719–738. https://doi.org/10.1037/a0026078

Barrick, M. R., Shaffer, J. A., & DeGrassi, S. W. (2009). What you see may not be what you get: Relationships among self-presentation tactics and ratings of interview and job performance. *Journal of Applied Psychology*, *94*(6), 1394–1411. https://doi.org/10.1037/a0016532

Basch, J. M., Brenner, F., Melchers, K. G., Krumm, S., Dräger, L., Herzer, H., & Schuwerk, E. (2021). A good thing takes time: The role of preparation time in asynchronous video interviews. *International Journal of Selection and Assessment*, *29*(3–4), 378–392. https://doi.org/10.1111/ijsa.12341

Basch, J. M., & Melchers, K. G. (2019). Fair and Flexible?! Explanations can improve applicant reactions toward asynchronous video interviews. *Personnel Assessment and Decisions*, *5*(3), Article 2. https://doi.org/10.25035/pad.2019.03.002

Basch, J. M., & Melchers, K. G. (2024). Here's looking at you: Does eye contact in video interviews affect how applicants are perceived and evaluated? *Journal of Business & Psychology*, *40*(3), 669–685. https://doi.org/10.1007/s10869-024-09981-4

Basch, J. M., Roulin, N., Gläsner, J., Spengler, R., & Wilhelm, J. (2024). How different backgrounds in video interviews can bias evaluations of applicants. *International Journal of Selection and Assessment*, *32*(4), 535–543. https://doi.org/10.1111/ijsa.12487

Brenner, F. S., Ortner, T. M., & Fay, D. (2016). Asynchronous video interviewing as a new technology in personnel selection: The applicant's point of view. *Frontiers in Psychology*, *7*, 863. https://doi.org/10.3389/fpsyg.2016.00863

Campion, M. A., Palmer, D. K., & Campion, J. E. (1997). A review of structure in the selection interview. *Personnel Psychology*, *50*(3), 655–702. https://doi.org/10.1111/j.1744-6570.1997.tb00709.x

Castro, J. R., & Gramzow, R. H. (2015). Rose colored webcam: Discrepancies in personality estimates and interview performance ratings. *Personality and Individual Differences*, *74*, 202–207. https://doi.org/10.1016/j.paid.2014.10.038

Chapman, D. S., & Zweig, D. I. (2005). Developing a nomological network for interview structure: Antecedents and consequences of the structured selection interview. *Personnel Psychology*, *58*(3), 673–702. https://doi.org/10.1111/j.1744-6570.2005.00516.x

Dahlke, J. A., & Sackett, P. R. (2017). The relationship between cognitive-ability saturation and subgroup mean differences across predictors of job performance. *Journal of Applied Psychology*, *102*(10), 1403–1420. https://doi.org/10.1037/apl0000234

Davis, M. G., Haas, M. R. C., Gottlieb, M., House, J. B., Huang, R. D., Hopson, L. R., & Coates, W. C. (2020). Zooming in versus flying out: Virtual residency interviews in the era of COVID-19. *AEM Education and Training*, *4*(4), 443–446. https://doi.org/10.1002/aet2.10486

Derous, E., Buijsrogge, A., Roulin, N., & Duyck, W. (2016). Why your stigma isn't hired: A dual-process framework of interview bias. *Human Resource Management Review*, *26*, 90–111. https://doi.org/10.1016/j.hrmr.2015.09.006

Diedenhofen, B., Musch, J., & Olivier, J. (2015). Cocor: A comprehensive solution for the statistical comparison of correlations. *PLOS ONE*, *10*(4), e0131499. https://doi.org/10.1371/journal.pone.0121945

Dunlop, P. D., Holtrop, D., & Wee, S. (2022). How asynchronous video interviews are used in practice: A study of an Australian-based AVI vendor. *International Journal of Selection and Assessment*, *30*(3), 448–455. https://doi.org/10.1111/ijsa.12372

Griswold, K. R., Phillips, J. M., Kim, M. S., Mondragon, N., Liff, J., & Gully, S. M. (2021). Global differences in applicant reactions to virtual interview synchronicity. *The International Journal of Human Resource Management*, *33*(15), 2991–3018. https://doi.org/10.1080/09585192.2021.1917641

Handler, C. (2020). *COVID-19 talent acquisition benchmarking survey*. https://rocket-hire.com/2020/06/17/covid-19-talentacquisition-benchmarking-report/

Hickman, L., Bosch, N., Ng, V., Saef, R., Tay, L., & Woo, S. E. (2022). Automated video interview personality assessments: Reliability, validity, and generalizability investigations.

Journal of Applied Psychology, *107*(8), 1323–1351. https://doi.org/10.1037/apl0000695

Hickman, L., Langer, M., Saef, R. M., & Tay, L. (2024). Automated speech recognition bias in personnel selection: The case of automatically scored job interviews. *Journal of Applied Psychology*, Advance online publication. https://doi.org/10.1037/apl0001247

Hickman, L., Liff, J., Rottman, C., & Calderwood, C. (2024). The effects of the training sample size, ground truth reliability, and NLP method on language-based automatic interview scores' psychometric properties. *Organizational Research Methods*, *10944281241264027*. https://doi.org/10.1177/10944281241264027

Hubert, M., Rousseeuw, P., & Verdonck, T. (2009). Robust PCA for skewed data and its outlier map. *The Fourth Special Issue on Computational Econometrics*, *53*(6), 2264–2274. https://doi.org/10.1016/j.csda.2008.05.027

Huffcutt, A. I., Van Iddekinge, C. H., & Roth, P. L. (2011). Understanding applicant behavior in employment interviews: A theoretical model of interviewee performance. *Human Resource Management Review*, *21*, 353–367. https://doi.org/10.1016/j.hrmr.2011.05.003

Jaccard, J., & Wood, G. (1988). The effects of incomplete information on the formation of attitudes toward behavioral alternatives. *Journal of Personality and Social Psychology*, *54*(4), 580–591. https://doi.org/10.1037/0022-3514.54.4.580

Kith, S., Cassidy, S. A., & Powell, D. M. (2022). The structured interview's resistance to gender discrimination under cognitive load. *International Journal of Selection and Assessment*, *30*(2), 281–301. https://doi.org/10.1111/ijsa.12361

Koutsoumpis, A., Ghassemi, S., Oostrom, J. K., Holtrop, D., van Breda, W., Zhang, T., & de Vries, R. E. (2024). Beyond traditional interviews: Psychometric analysis of asynchronous video interviews for personality and interview performance evaluation using machine learning. *Computers in Human Behavior*, *154*, 108128. https://doi.org/10.1016/j.chb.2023.108128

Landers, R. N., & Behrend, T. S. (2022). Auditing the AI auditors: A framework for evaluating fairness and bias in high stakes AI predictive models. *American Psychologist*, *78*(1), 36–49. https://doi.org/10.1037/amp0000972

Lee, T. C., McKinley, S. K., Dream, S. Y., Grubbs, E. G., Dissanaike, S., & Fong, Z. V. (2021). Pearls and pitfalls of the virtual interview: Perspectives from both sides of the camera. *Journal of Surgical Research*, *262*, 240–243. https://doi.org/10.1016/j.jss.2020.12.052

Levashina, J., Hartwell, C. J., Morgeson, F. P., & Campion, M. A. (2014). The structured employment interview: Narrative and quantitative review of the research literature. *Personnel Psychology*, *67*(1), 241–293. https://doi.org/10.1111/peps.12052

Liff, J., Mondragon, N., Gardner, C., Hartwell, C. J., & Bradshaw, A. (2024). Psychometric properties of automated video interview competency assessments. *Journal of Applied Psychology*, *109*(6), 921–948. https://doi.org/10.1037/apl0001173

Lubbe, D., & Nitsche, A. (2019). Reducing assimilation and contrast effects on selection interview ratings using behaviorally anchored rating scales. *International Journal of Selection and Assessment*, *27*(1), 43–53. https://doi.org/10.1111/ijsa.12230

Lukacik, E.-R., & Bourdage, J. S. (2025). Does design matter? The (limited) effects of six asynchronous video interview design features on impression management, reactions, and evaluations. *International Journal of Selection and Assessment*, *33*(1), e12511. https://doi.org/10.1111/ijsa.12511

Lukacik, E.-R., Bourdage, J. S., & Roulin, N. (2022). Into the void: A conceptual model and research agenda for the design and use of asynchronous video interviews. *Human Resource Management Review*, *32*(1), 100789. https://doi.org/10.1016/j.hrmr.2020.100789

Martin-Raugh, M. P., Kell, H. J., Randall, J. G., Anguiano-Carrasco, C., & Banfi, J. T. (2023). Speaking without words: A meta-analysis of over 70 years of research on the power of nonverbal cues in job interviews. *Journal of Organizational Behavior*, *44*(1), 132–156. https://doi.org/10.1002/job.2670

McCarthy, J. M., Truxillo, D. M., Bauer, T. N., Erdogan, B., Shao, Y., Wang, M., Liff, J., & Gardner, C. (2021). Distressed and distracted by COVID-19 during high-stakes virtual interviews: The role of job interview anxiety on performance and reactions. *Journal of Applied Psychology*, *106*(8), 1103–1117. https://doi.org/10.1037/apl0000943

McCarthy, J. M., Van Iddekinge, C. H., & Campion, M. A. (2010). Are highly structured job interviews resistant to demographic similarity effects? *Personnel Psychology*, *63*(2), 325–359. https://doi.org/10.1111/j.1744-6570.2010.01172.x

McColl, R., & Michelotti, M. (2019). Sorry, could you repeat the question? Exploring video-interview recruitment practice in HRM. *Human Resource Management Journal*, *29*(4), 637–656. https://doi.org/10.1111/1748-8583.12249

Mejia, C., & Torres, E. N. (2018). Implementation and normalization process of asynchronous video interviewing practices in the hospitality industry. *International Journal of Contemporary Hospitality Management*, *30*(2), 685–701. https://doi.org/10.1108/IJCHM-07-2016-0402

Neumann, M., Niessen, A. S. M., Hurks, P. P. M., & Meijer, R. R. (2023). Holistic and mechanical combination in psychological assessment: Why algorithms are underutilized and what is needed to increase their use. *International Journal of Selection and Assessment*, *31*(2), 267–285. https://doi.org/10.1111/ijsa.12416

Neumann, M., Niessen, A. S. M., Tendeiro, J. N., & Meijer, R. R. (2022). The autonomy-validity dilemma in mechanical prediction procedures: The quest for a compromise. *Journal of Behavioral Decision Making*, *35*(4), e2270. https://doi.org/10.1002/bdm.2270

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*, 153–163. https://doi.org/10.1016/j.jesp.2017.01.006

Powell, D., Kavanagh, M., Wiseman, B., & Hodgins, A. (2023). Effects of background cues on videoconference interview ratings. *Personnel Assessment and Decisions*, *9*(1), Article 3. https://doi.org/10.25035/pad.2023.01.003

Roulin, N., Bourdage, J. S., & Wingate, T. G. (2019). Who is conducting "better" employment interviews? Antecedents of structured interview components use. *Personnel Assessment and Decisions*, *5*(1), Article 2. https://doi.org/10.25035/pad.2019.01.002

Roulin, N., Lukacik, E.-R., Bourdage, J. S., Clow, L., Bakour, H., & Diaz, P. (2023). Bias in the background? The role of background information in asynchronous video interviews. *Journal of Organizational Behavior*, *44*(3), 458–475. https://doi.org/10.1002/job.2680

Roulin, N., Wong, O., Langer, M., & Bourdage, J. S. (2023). Is more always better? How preparation time and re-recording opportunities impact fairness, anxiety, impression management, and performance in asynchronous video interviews. *European Journal of Work & Organizational Psychology*, *32*(3), 333–345. https://doi.org/10.1080/1359432X.2022.2156862

Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2022). Revisiting meta-analytic estimates of validity in personnel selection: Addressing systematic overcorrection for restriction of range. *Journal of Applied Psychology*, *107*(11), 2040–2068. https://doi.org/10.1037/apl0000994

Scott, C., & Roulin, N. (2024). Does background type and blurring affect performance ratings in video interviews? *Personnel Assessment and Decisions*, *10*(1), Article 2. https://doi.org/10.25035/pad.2024.01.002

Suen, H.-Y., Chen, M. Y.-C., & Lu, S.-H. (2019). Does the use of synchrony and artificial intelligence in video interviews affect interview ratings and applicant attitudes? *Computers in Human Behavior*, *98*, 93–101. https://doi.org/10.1016/j.chb.2019.04.012

Summers, J. K., Howe, M., McElroy, J. C., Ronald Buckley, M., Pahng, P., & Cortes-Mejia, S. (2018). A typology of stigma within organizations: Access and treatment effects. *Journal of Organizational Behavior*, *39*(7), 863–868. https://doi.org/10.1002/job.2279

Tay, L., Woo, S. E., Hickman, L., Booth, B. M., & D'Mello, S. (2022). A conceptual framework for investigating and mitigating machine-learning measurement bias (MLMB) in psychological assessment. *Advances in Methods and Practices in Psychological Science*, *5*(1), 25152459211061337. https://doi.org/10.1177/25152459211061337

Tilston, O., Krings, F., Roulin, N., Bourdage, J. S., & Fetzer, M. (2024). Reactions to asynchronous video interviews: The role of design decisions and applicant age and gender. *Human Resource Management*, *63*(2), 313–332. https://doi.org/10.1002/hrm.22202

Tippins, N. T., Oswald, F. L., & McPhail, S. M. (2021). Scientific, legal, and ethical concerns about AI-based personnel selection tools: A call to action. *Personnel Assessment and Decisions*, *7*(2), 1. https://doi.org/10.25035/pad.2021.02.001

Torres, E. N., & Gregory, A. (2018). Hiring manager's evaluations of asynchronous video interviews: The role of candidate competencies, aesthetics, and resume placement. *International Journal of Hospitality Management*, *75*, 86–93. https://doi.org/10.1016/j.ijhm.2018.03.011

Torres, E. N., & Mejia, C. (2017). Asynchronous video interviews in the hospitality industry: Considerations for virtual employee selection. *International Journal of Hospitality Management*, *61*, 4–13. https://doi.org/10.1016/j.ijhm.2016.10.012

Wingate, T. G., Bourdage, J. S., & Steel, P. (2025). Evaluating interview criterion-related validity for distinct constructs: A meta-analysis. *International Journal of Selection and Assessment*, *33*(1), e12494. https://doi.org/10.1111/ijsa.12494