International Journal of Selection and Assessment

WILEY

International Journal of SELECTION AND ASSESSMENT

RESEARCH ARTICLE

Social Desirability Tendency in Personality-Based Job Interviews—A Question of Interview Format?

Valerie Schröder[1] | Anna Luca Heimann[1] | Pia Ingold[2] | Nicolas Roulin[3] | Marianne Schmid Mast[4] | Manuel Bachmann[5] | Martin Kleinmann[1]

[1]Department of Psychology, University of Zurich, Zurich, Switzerland | [2]Department of Psychology, University of Copenhagen, Copenhagen, Denmark | [3]Department of Psychology, Saint Mary's University, Halifax, Canada | [4]Faculty of Business and Economics, University of Lausanne, Lausanne, Switzerland | [5]Department of Social Work, University of Applied Sciences Bern, Bern, Switzerland

Correspondence: Valerie Schröder (v.schroeder@psychologie.uzh.ch)

Received: 19 October 2023 | **Revised:** 14 February 2025 | **Accepted:** 19 February 2025

Funding: The research project reported in this paper was supported by a grant from the Swiss National Science Foundation (Schweizer Nationalfonds zur Förderung der Wissenschaftlichen Forschung; Grant Number 179198).

Keywords: asynchronous video interviews | personality-based interviews | social desirability tendency | written interviews

ABSTRACT

Today's variety of interview formats raises the question of their interchangeability. For personality interviews, a crucial question is whether different formats are comparably robust against applicants' social desirability tendency (SDT) to ensure an accurate measurement. Using a within-subjects design in a simulated selection setting with 211 participants, this study examined how SDT affects personality scores in a face-to-face, asynchronous video, and written interview—all with similar interview questions designed to measure personality. Relationships between interview scores and SDT were weakest in the face-to-face format and strongest in the written format and differed depending on which personality trait was assessed. The findings highlight the suitedness of different interview formats for measuring personality with important implications for interview design and personality assessment.

Interviews are valid selection tools and part of almost every selection process (Sackett et al. 2022, 2023). They are used to assess various constructs, including personality traits (Huffcutt et al. 2011; Salgado and Moscoso 2002). In recent years, the ways in which interviews are conducted have become more diverse. In particular, organizations increasingly use asynchronous video interviews, in which applicants record their responses to interview questions on video (Brenner et al. 2016; Lukacik et al. 2022; Suen et al. 2019). Asynchronous video interviews are becoming more popular because they are easy to administer and less cost-intensive compared to traditional face-to-face interviews, making them particularly suitable as a preselection tool (Basch et al. 2022). Written interviews, in which candidates are presented with written questions and asked to provide written answers, promise to be an even more accessible option (Whetzel et al. 2003). Given this variety of different interview formats, it is critical to understand the extent to which different interview formats function in the same way (i.e., whether they can be used interchangeably) or have different strengths and weaknesses.

One way in which interview formats may function differently is their susceptibility to being influenced by applicants' social desirability tendency (SDT). SDT describes a trait-like, conscious, or unconscious motivation to appear to oneself and others as a person with positive attributes (Paulhus 1984). SDT can be considered problematic because it remains unclear how much of a measurement is shaped by an applicant's SDT and how much is due to the construct a selection tool was originally designed to measure. As such, SDT is a potential threat to accurate measurement in selection. SDT is often studied in the context of personality measurement, focusing on personality

Summary

- Social desirability tendency is the tendency to appear particularly positively.
- When measuring personality, social desirability tendency can interfere with an accurate measurement of the personality traits in question, causing problems, especially in personnel selection settings.
- Using job interviews to assess personality traits may help to reduce the impact of social desirability tendency as compared to personality inventories.
- This study compared social desirability tendency in personality-based interviews presented in three formats: A face-to-face interview, an asynchronous video interview, and a written interview.
- Among the three interview formats, the relationship with social desirability tendency was lowest for the face-to-face interview and highest for the written interview.
- Reducing the relationship with social desirability tendency can be achieved with any interview format compared to a contextualized personality inventory.

inventories (Ellingson et al. 1999, 2001; e.g., Holden 2007; Li et al. 2013). There is reason to believe that interviews, which are often more complex and cognitively demanding than personality inventories, may be less susceptible to SDT influences. However, this may vary between different interview formats, and using an interview format that captures SDT could prove problematic if SDT is not the trait considered relevant to performance in the job being interviewed for.

To deepen the understanding of different interview formats, this study examines the extent to which interviewees' SDT affects personality scores in a synchronous face-to-face interview, an asynchronous video interview, and an asynchronous written interview. The first goal of this study is to examine whether scores from different interview formats (and a traditional personality inventory) are related differently to interviewees' SDT. To this end, we use a within-subjects design in which the same group of interviewees completes three interviews of different formats, which are designed in parallel to measure the same personality constructs. This design allows for a systematic comparison of interview formats by keeping the measured content constant. Comparing the relationships between SDT and personality scores for different interview formats provides insight into which formats are more susceptible to (vs. robust against) capturing interviewees' SDT.

The second goal is to examine to what extent interviewees' SDT affects the assessment of different personality traits (as measured in interviews of different formats). Among the Big Five personality traits (extraversion, agreeableness, conscientiousness, emotional stability, openness/intellect; Goldberg 1992), SDT is considered to be most relevant when measuring conscientiousness and emotional stability because these two traits describe characteristics that are particularly desirable in applicants as they are central to job performance across a variety of jobs (Klehe et al. 2012; Schmit and Ryan 1993). This study will generate knowledge about how the measurement of different traits is differentially affected by interviewees' SDT and whether this differs across interview formats.

Taken together, this study will inform research and practice about the interview format(s) that are least conducive to capturing interviewee SDT to identify which formats are best suited to accurately measure the intended, job-relevant constructs they were designed to measure. Such comparative research on interview formats is required to gain insight into the functioning of different measurement approaches. It will also inform practitioners if they can freely choose an interview format that is most suitable for their purposes or if they need to consider that the interview format of their choice may tap into SDT more than other available interview formats.

1 | SDT in Personnel Selection

SDT describes a conscious or unconscious motivation to create a positive image of oneself that aligns with societal norms and expectations (Paulhus 1984). In the context of selection, the created image can also revolve around norms and expectations by the organization or the respective job. SDT is considered a relatively stable individual characteristic (Paulhus 1984). As such, SDT is different from specific behaviors that have been studied in selection research, such as impression management (Bolino et al. 2016) or faking (Snell et al. 1999). Impression management behavior is defined as behavior that aims to create a particular image of oneself (Ellis et al. 2002) and is often studied in the context of job interviews (e.g., Arseneault and Roulin 2024; Bourdage et al. 2018, 2020). It can be either honest or deceptive (Bourdage et al. 2018). Faking behavior has been studied in both job interviews (Melchers et al. 2020) and personality inventories (e.g., Furnham 1990; Griffin et al. 2004) and refers to "an intentional distortion or falsification of responses" (Levashina and Campion 2006, 300). Impression management and faking describe specific, deliberate, and conscious behaviors that applicants might engage in when completing a specific selection tool. They may be used strategically in a given situation, for instance, to increase the perceived fit with the job or the organizational culture (e.g., Roulin and Krings 2020). In contrast, SDT is a disposition to present oneself in a positive way, not only to others but also to oneself (Ellingson et al. 2001; Ones et al. 1996). As such, SDT can manifest in socially desirable behaviors (Zettler et al. 2015), and this also includes socially desirable responding in assessment instruments (Kanning and Kuhne 2006), that is, responding in a way that aligns with societal, organizational, or situational expectations. Yet, such socially desirable responding is not necessarily conscious and differs from impression management and faking behaviors in that the behavior or response can originate from a positive but potentially exaggerated self-view.

SDT can cause problems in the context of personality assessment because, as a potentially confounding variable, it can interfere with the accurate measurement of personality traits (e.g., Edwards 1957; Li and Bagger 2006). Limiting the variance caused by confounding variables is critical to accurately estimating the true score for a given personality trait. An applicant's SDT may introduce variance that is potentially unrelated

to the personality trait of interest and obfuscate the true score (Ellingson et al. 2001; Nederström and Salmela-Aro 2014; Paunonen and LeBel 2012). For example, if we want to measure conscientiousness in a selection context because it is relevant to job performance in a particular role, the measurement of this trait should reflect as closely as possible the actual, true expression of an applicant's conscientiousness to assess their suitability for the role. If applicants' SDT, as a trait-like disposition, introduces variance unrelated to their conscientiousness trait, this measurement may become inaccurate.

As a potential source of measurement error, SDT can also contribute to unfairness in the selection context by indirectly altering the rank order of applicants (Mueller-Hanson et al. 2003). Given that SDT differs across individuals, applicants with particularly high levels of SDT may unfairly score higher than their peers on a selection tool if their responses shift toward the more desirable end of a scale (Ellingson et al. 2001; King and Bruner 2000).

There is also concern that SDT may interfere with the criterion-related validity of personality assessments. Findings on the effect of SDT on criterion-related validity are mixed, with some showing no effect (Li and Bagger 2006) and others providing support that controlling for SDT can increase the prediction of job performance (Berry et al. 2007; Li et al. 2013). For academic performance, Xue et al. (2023) found a moderating effect of SDT on criterion-related validity, such that personality traits were negatively related to academic performance when scores on an SDT scale were high (defined as one standard deviation above the mean), but positively related when SDT was low. An effective way of ensuring an accurate measure of personality to reduce the risk that applicants with particularly high levels of SDT are (unfairly) more likely to receive a job offer (Hough 1998; Rosse et al. 1998), and to reduce potential negative effects on criterion-related validity, is to use selection tools that are more robust to applicants' tendency toward socially desirable responses. This means that a selection tool does not (or barely) capture SDT.

In selection (and especially in personality assessment), measures of SDT are often used to detect high-SDT applicants (Reid-Seiser and Fritzsche 2001). For this purpose, SDT is typically measured with self-report scales that include so-called improbable (or unlikely virtuous) items, which are highly socially desirable statements (Goffin and Christiansen 2003). An example statement is “I would never speak ill of a colleague or my employer” (Satow 2012). These items are designed to capture conscious or unconscious tendencies to describe oneself in a desirable way, driven by high levels of SDT (Christiansen et al. 2021). SDT scales are controversially discussed in the literature due to different understandings of what they capture (e.g., de Vries et al. 2014; Dunlop et al. 2012; Müller and Moshagen 2019). On the one side, some studies have demonstrated relationships between SDT and actual honest behavior (Müller and Moshagen 2019; Zettler et al. 2015) or desirable personality traits rated by others (de Vries et al. 2014; Dunlop et al. 2012). On the other side, prominent examples of SDT scales, such as the Marlowe-Crowne scale (Crowne and Marlowe 1960), are widely used and considered as indicators of response quality in self-report assessments in research and practice (e.g., Heerwegh 2009; Rhim et al. 2022).

1.1 | SDT and Job Interview Formats

The choice of a selection tool can make a substantial difference in how strongly an applicant's level of SDT affects their assessment scores (Kanning and Kuhne 2006; Ones et al. 1996; Richman et al. 1999). Some interview formats are likely more prone to capture SDT than others because they differ in terms of several method factors (Lievens and Sackett 2017). In face-to-face interviews, applicants interact in real time with an interviewer who is present in the same location. In asynchronous video interviews, questions are presented to an applicant on an online platform, and they record themselves answering these questions via webcam without an interviewer being present. In written interviews, applicants receive and reply to interview questions in a written format, without the presence of an interviewer and without an audio-visual presentation of themselves.

These modalities of different interview formats result in systematic differences in the *synchrony* and *response richness* of each format. The *synchrony* of a medium describes whether all communication participants use the medium at the same time (Dennis et al. 2008). The face-to-face interview is a synchronous interview format. In asynchronous interview formats, such as the asynchronous video interview or the written interview, there is one-way communication: Interviewees answer first (and alone), and their responses are evaluated afterward (Griswold et al. 2022). *Response richness* is derived from the concept of media richness (Daft and Lengel 1986), which describes the amount of information an assessment method transports. Because we focus on the information that applicants share, we will refer only to the information in the interviewee's response (i.e., response richness). The face-to-face interview has the highest response richness because, adding to audiovisual information, it transports information from the live setting, such as information on body language and movement in a room, and information collected from the interaction, such as eye contact or distance (Edinger and Patterson 1983; Guerrero 2014). Asynchronous video interviews have less response richness because the response is enriched with audiovisual information (Lievens et al. 2015), that is, voice and nonverbal behavior (Rasmussen 1984; Tu et al. 2022). Written interviews have the lowest response richness because they capture only response content. Table 1 describes the methodological features of each interview format. It also includes a comparison to the features of traditional personality inventories.

1.1.1 | SDT and Synchrony in Job Interview Formats

It seems plausible that applicants' SDT will affect their ratings more in asynchronous as compared to synchronous interview formats because synchronous interviews create a more demanding situation, reducing resources to identify what is desirable in a given situation and respond accordingly. First, synchronous interview formats are cognitively demanding (Van Iddekinge et al. 2005) because they involve real-time interactions that require a high degree of mental presence and timely reactions (Stivers et al. 2009; Templeton et al. 2022). Applicants listen to the interview question while synchronously thinking of a possible

TABLE 1 | Method factors of personality interviews in comparison with a contextualized personality inventory and hypothesized relationship with social desirability tendency.

Method factor	Modality				Hypothesized relation
	Face-to-face interview	Asynchronous video interview	Written interview	Contextualized personality inventory	
Response format	Open-ended	Open-ended	Open-ended	Close-ended	Open-end—lower influence of social desirability tendency
Synchrony	Synchronous	Asynchronous	Asynchronous	Asynchronous	Synchronous—lower influence of social desirability tendency
Response richness (media richness)	Very high: transports audiovisual information (expression, gesture, voice) and cues from interaction	High: transports audiovisual information (expression, gesture, voice)	Low: transports only the response	Very low: transports only the response	Higher response richness—lower influence of social desirability tendency

response and signaling attention, perhaps by a nonverbal reaction such as nodding. They must then formulate a convincing verbal response to the interview question on the spot. In contrast, synchronous interview formats provide applicants with extra preparation time to plan a response and find an optimal answer to a question (Basch et al. 2021; Roulin et al. 2023).

Second, synchronous interview formats may be more demanding because the presence of another person (i.e., the interviewer) can increase feelings of stress. Being observed by someone (and being aware of it because that person is present) can increase nervousness and fear of exposure (Feiler and Powell 2016; McCarthy and Goffin 2004). The affective demands of a face-to-face interview situation may reduce applicants' resources for identifying what is a desirable response to a given interview question. Thus, for applicants high in SDT, the high demands of a face-to-face interview may alleviate their ability to respond in a way that is consistent with their tendency to appear socially desirable. In contrast, asynchronous interview formats may cause less stress because there is less time pressure and less awareness that one is being observed.

1.1.2 | SDT and Response Richness in Job Interview Formats

It seems plausible that applicants' SDT will affect their ratings more in interview formats with low response richness because, similar to the effect of synchrony, a higher response richness is more demanding, reducing an applicant's capacity to identify the demands of the situation. Higher response richness, such as in a face-to-face interview, requires applicants to demonstrate more behavioral skills (Ingold et al. 2015). Specifically, in simpler formats, applicants have to express themselves in only one communication channel, as compared to a richer format in which their behavior, expression, and voice can be evaluated (Daft and Lengel 1986). In other words, they must create consistency between verbal, paraverbal, and nonverbal information that is sent through their response in a face-to-face or asynchronous video interview. Again, this added cognitive effort may reduce the resources needed to identify what is socially desirable in a given interview situation.

We propose that applicants' SDT will affect asynchronous interview formats with low response richness to a greater extent. If the interview format is asynchronous and has a lower level of response richness, interviewees high in SDT can invest more cognitive effort in identifying the demands and expectations in the interview, such that they can respond to interview questions in alignment with their SDT (see also Table 1). Therefore, we posit:

Hypothesis 1: *SDT will affect personality scores as rated by interviewers in personality-based interviews, and this effect will be strongest for the written interview, followed by the asynchronous video interview and the face-to-face interview.*

1.2 | SDT and Personality Traits

Applicants' SDT will likely have different relationships with personality-based interview rating scores depending on which

personality trait is assessed. Not only SDT but also the desirability of a certain personality item affects the resulting personality score (Cui et al. 2022). In the selection context, items measuring conscientiousness (describing the tendency to act dutifully, disciplined, and achievement-striving) and emotional stability (describing the tendency for positive emotions, self-consciousness, and to remain calm) can be considered the most relevant among the Big Five personality traits (Goldberg 1990) for predicting job performance across a broad range of occupations (Sackett et al. 2022). At the same time, these traits have shown the strongest relationships with measures of SDT (Ones et al. 1996; Smith and Ellingson 2002). Ratings on these two traits also differ more than other traits, such as agreeableness, openness, and extraversion, when examined in an applicant versus a nonapplicant sample (Hu and Connelly 2021). This implies that applicants understand the desirability of different traits for the job context and (consciously or unconsciously) present themselves as particularly conscientious or emotionally stable. What follows is that SDT may show particularly strong relationships with interview questions targeting personality traits that are particularly valued by organizations (i.e., conscientiousness and emotional stability). Accordingly, we hypothesize:

Hypothesis 2: *SDT will affect interview rating scores of conscientiousness and emotional stability to a stronger extent than interview scores of the other Big Five personality traits* (*i.e., extraversion, agreeableness, and openness/intellect*).

2 | Methods

2.1 | Sample and Setting

Using a within-subjects design, a sample of 211 employed individuals (45% women) completed a simulated selection procedure. To plan our sample size, we performed a power analysis using G*Power (Faul et al. 2007), originally planning for a repeated-measures, between-factors analysis of variance (ANOVA) and an estimated effect size of 0.25 for differences between interview formats based on prior studies comparing outcomes across interview formats (Basch et al. 2020; Van Iddekinge et al. 2006). The power analyses suggested a sample size of 165. We recruited additional participants because we expected a dropout of at least 10% based on experiences in prior studies.

Similar to prior interview research (Bourdage et al. 2020; Roulin and Powell 2018; Swider et al. 2016), we advertised the study as training for future job applications. We recruited participants through different professional networking channels, university career services, social media, and local advertisements. Participants were, on average, 30.94 years old ($SD = 7.90$) and had, on average, 8.06 years of job experience ($SD = 8.09$). The biggest industry was education and research (43.27%), followed by service (9.48%), manufacturing (8.06%), finance and insurance (6.62%), administration and public service (6.16%), health and social work (6.16%), and others (19%). The majority of participants (80%) held a university degree (Bachelor, Master, or Doctoral degree). Many worked in temporary jobs, which motivated their participation in the selection training. Prior interview experience was evenly distributed across participants, with the average participant having had three to four prior job interviews.

Participants underwent a 1-day selection training. They completed several assessments during the day, including the three different personality-based interview formats and a contextualized personality inventory, in randomized order. Each interview was followed by an exercise or break to ensure that participants never completed two parallel interviews directly after one another. At the beginning of the day, participants received a hypothetical job description to create a common frame of reference. The job description detailed requirements that targeted the same personality traits that were measured in the interviews (i.e., the Big Five traits) to signal to participants that these traits were important to their success in the simulated selection situation. At the end of the training day, participants received extensive feedback on their performance and advice on how to prepare for future interviews.

To create a realistic selection situation that allowed us to measure personality in a high-stakes setting, we asked interviewees to dress as they would for an actual job selection procedure and awarded the best participant of the day (the one that would have gotten a job offer) with a cash prize (equivalent to $100) as an incentive. Similar simulated selection procedures have demonstrated ecological validity in the past (Ingold et al. 2016; Klehe et al. 2008; Swider et al. 2016). In line with this, participants in the present study indicated that they perceived the simulated selection setting as realistic ($M = 5.18$, $SD = 1.16$), could adapt to the role as an applicant ($M = 5.45$, $SD = 1.02$), felt as if they were a real applicant ($M = 5.15$, $SD = 1.22$), and behaved as they would in a real selection setting ($M = 5.57$, $SD = 1.02$; with all responses on a scale from 1–*completely disagree* to 7–*completely agree*). All participants agreed to their data being used for research purposes. The study was not pre-registered. The quantitative data are available from the first author of the study.

2.2 | Measures

2.2.1 | SDT

To measure SDT, we used a scale with seven items by Satow (2012). The scale contains highly socially desirable statements that are rated on a 7-point Likert Scale (1–*completely disagree* to 7–*completely agree*). A sample item is "I would never speak ill of a colleague or my employer." We excluded one item from the scale ("I would never let a doctor put me on sick leave without me actually being") because it reduced the scale's reliability. The remaining six items had an internal consistency of $\alpha = 0.72$, which is in line with the 0.70 reported for this scale in prior research (Satow 2012). This scale has been shown to predict positive self-presentation in a personality inventory ($r = 0.29$; Satow 2012). SDT was measured at the end of the day and in a research rather than an applicant-like setting, which should produce a more accurate trait measurement because it reduces distortions observed in high-stakes settings on such measures (Christiansen et al. 2021).

2.2.2 | Personality Inventory

As a personality inventory, we used the 50 personality items from the international personality item pool representing markers for the Big-Five factor structure (Goldberg 1992). All original items were contextualized (similar to Heimann et al. 2021; Lievens et al. 2008) by adding the term "at work" at the beginning of each item to create similarity to the job interviews. An example item for conscientiousness was "At work, I am always prepared." Reliabilities were $\alpha = 0.80$ for extraversion, $\alpha = 0.71$ for agreeableness, $\alpha = 0.81$ for conscientiousness $\alpha = 0.85$ for emotional stability, and $\alpha = 0.81$ for openness/intellect, which is similar to past research (between $\alpha = 0.70$ and $\alpha = 0.86$; International Personality Item Pool 2019).

2.2.3 | Personality-Based Job Interviews

To allow for direct comparisons of different interview formats, we developed new sets of personality-based interview questions specifically for this study. Each interview contained 15 questions, with three interview questions for each personality trait. The objective was to create parallel versions of the same personality-based interview to be conducted as (a) face-to-face interview, (b) asynchronous video interview, and (c) written interview. We decided to develop patterned behavior description interview questions (i.e., BDI questions; Janz 1982) because personality can be expressed in behavior if a situation contains trait-relevant cues (Tett et al. 2021). BDI questions have been used to measure personality in prior work (Heimann et al. 2021; Van Iddekinge et al. 2005).

Questions were developed following the same procedure used in prior studies on personality-based interviews (Heimann et al. 2021; Van Iddekinge et al. 2005). First, as source material, we consulted Big Five personality inventories (Costa and McCrae 2008; Goldberg 1992) and selected three items per trait. We chose items that covered different facets of a personality trait (DeYoung et al. 2007) as source material. For example, for conscientiousness, we ensured that items referred to behaviors indicating industriousness (e.g., "I am exacting in my work") and to behaviors indicating orderliness (e.g., "I am always prepared"). We chose items that were relevant to work settings or could be adapted to the workplace and that had a strong behavioral expression.

Then, for each item, we generated three trait-relevant situations that were suited for BDI questions, making sure that they typically occurred in almost everyone's past work experience. Situations were trait-relevant in that individuals with a high and low expression on the item would behave differently (Tett et al. 2021). For example, individuals who score differently on the conscientiousness item "I am always prepared" are likely to behave differently in a situation in which they are scheduling a meeting with someone. In this way, we created three *different but parallel* situations in which individuals may differ in their preparation behavior. For example, in parallel to the meeting preparation question, we generated two additional situations, one related to preparing job application materials and another related to preparing for a recurring event, such as a lecture. For each situation, we created behavior-anchored rating scales to guide interviewers in assessing the target trait. Interviewers rated the expression of the trait on a scale from 1–*low expression* to 5–*high expression*, with anchors provided for low, medium, and high levels. An example of an interview question along with the corresponding rating scale can be found in the appendix to this paper.

An independent group of graduate students in I/O psychology reviewed the interview questions and behavioral anchors for comprehensibility. They also wrote responses to each question as a pretest, which were then reviewed by the first author to ensure that the parallel questions elicited the intended responses, that responses were comparable in content and length, and that they fit the rating scales. Based on the results of this pretest, the interview questions and BARS were revised again. The parallel questions were then randomly assigned to the interview formats.

In each interview format, each interview question was rated on the corresponding personality trait by two trained I-O psychology students who were also interviewers in the face-to-face interviews. The raters were trained in conducting standardized interviews and rating personality but were blind to the hypotheses of the manuscript. For all interview formats, the independent raters took notes during the interview and made their individual ratings immediately after completion. After both raters made their independent ratings, they compared their ratings. If their ratings differed by two points or more on the 5-point rating scale, raters discussed their ratings, explaining how they each arrived at their ratings using the notes they took during the interview. This ensured that no interviewer overlooked important information that contributed to their rating. After the discussion, they were allowed to adjust their ratings but did not have to agree on the same final rating. For the face-to-face interviews, the comparison and final scoring were done at the end of the day after all interviews were completed. The asynchronous video interviews and the written interviews were rated during breaks in the simulated selection day and the comparison and final scoring were done after raters completed a given set of interviews at a time. For the analyses in this paper, we averaged the ratings of both raters.

To assess inter-rater (i.e., interviewer) reliability, we computed one-way random effects intraclass coefficients (ICC). Mean ICCs after discussion between raters were 0.83 (ranging from 0.81 to 0.85) for the face-to-face interviews, 0.83 (ranging from 0.79 to 0.84) for the asynchronous virtual interview, and 0.86 (ranging from 0.83 to 0.90) for the written interview. In line with common practice in interview research, ICCs were calculated for interview scores after discussion between raters (e.g., Klehe and Latham 2006; Van Iddekinge et al. 2005) and differed slightly from ICCs before discussion (on average 0.07 higher after discussion). The values after discussion are in line with ICCs reported in earlier studies (Heimann et al. 2021; Van Iddekinge et al. 2005). An analysis of the construct variance of the personality-based interviews can be found in the Supporting Information Material.

3 | Results

3.1 | Hypotheses Testing

Hypothesis 1 stated that SDT would affect interview rating scores in personality-based interviews and that this effect would be strongest for the written interview format, followed by the asynchronous video interview format and the face-to-face interview format. To test this hypothesis, we compared the magnitude of the correlations between SDT and interview scores across interview formats. In research and practice, interviewees are most often evaluated based on an interview overall score (e.g., Basch et al. 2020; Kluemper et al. 2015; Martín-Raugh et al. 2023). Accordingly, we tested Hypothesis 1 using an interview overall score for each interview format (i.e., we averaged ratings across interviewers and all interview questions for each interview format). The correlations between SDT and interview scores are displayed in Table 2. SDT correlated positively and significantly with the overall score from the written interview ($r = 0.27$, $p < 0.001$) and, in line with Hypothesis 1, this correlation was significantly larger than the correlations for the asynchronous virtual interview ($r = 0.13$, $p = 0.052$; $z = 1.79$, $p = 0.036$) or the face-to-face interview ($r = 0.11$, $p = 0.126$; $z = -2.13$, $p = 0.017$). Although the direction of the effect was as hypothesized, Hypothesis 1 was partially supported because we found a significant relationship between SDT and the written interview format but not between SDT and the other formats.

Hypothesis 2 predicted that SDT would affect interview rating scores of conscientiousness and emotional stability more strongly than interview rating scores of the other assessed traits (i.e., extraversion, agreeableness, openness/intellect). The correlations between SDT and interview scores for each trait in each interview format are shown in Table 2. We found partial support for Hypothesis 2 only for the written interview format. In the written interview, the correlation between SDT and conscientiousness ($r = 0.20$, $p = 0.003$) and the correlation between SDT and emotional stability ($r = 0.21$, $p = 0.002$) were significantly larger than the correlation between SDT and agreeableness ($r = 0.03$, $p = 0.620$; $z = 1.88$, $p = 0.030$ compared to conscientiousness; $z = 1.96$, $p = 0.025$ compared to emotional stability), but this was not the case for extraversion ($r = 0.18$, $p = 0.009$; $z = 0.23$, $p = 0.409$ compared to conscientiousness; $z = 0.36$, $p = 0.359$ compared to emotional stability) or openness/intellect ($r = 0.16$, $p = 0.023$; $z = 0.49$, $p = 0.313$ compared to conscientiousness; $z = 0.59$, $p = 0.279$ compared to emotional stability). For the other interview formats, we did not find that SDT correlated positively and more strongly with either conscientiousness or emotional stability as compared to the other assessed traits.

As an alternative approach to test Hypothesis 2, we calculated composite scores of conscientiousness and emotional stability and compared them to a composite score of the remaining traits. Although the results pointed in the expected direction, correlations between SDT and the composite of conscientiousness and emotional stability did not differ significantly from correlations between SDT and the composite of the remaining traits for the written interview ($r = 0.27$, $p < 0.001$, as compared to $r = 0.19$, $p = 0.005$; $z = 1.02$, $p = 0.154$), the asynchronous video interview ($r = 0.14$, $p = 0.041$, as compared to $r = 0.10$, $p = 0.141$; $z = 0.47$, $p = 0.318$), or the face-to-face interview ($r = 0.11$, $p = 0.106$, as compared to $r = 0.03$, $p = 647$; $z = 1.09$, $p = 0.138$).

3.2 | Supplementary Analyses

3.2.1 | Comparing a High SDT Group With the Remaining Interviewees

In practice, SDT scales usually recommend excluding participants with a high SDT score (e.g., Satow 2012). A high SDT score is often indicated by a score that is one standard deviation above the mean (Ellingson et al. 2001). To match how SDT

TABLE 2 | Correlations between personality ratings and social desirability tendency for different traits and modalities.

	Face-to-face interview	Asynchronous video interview	Written interview	Contextualized personality inventory
Averaged rating across all traits	0.11	0.13	0.27**	0.40**
	[−0.03, 0.24]	[−0.00, 0.26]	[0.14, 0.39]	[0.28, 0.51]
Conscientiousness	0.08	0.12	0.20**	0.37**
	[−0.06, 0.21]	[−0.01, 0.25]	[0.07, 0.33]	[0.25, 0.48]
Emotional stability	0.09	0.10	0.21**	0.38**
	[−0.04, 0.22]	[−0.04, 0.23]	[0.08, 0.34]	[0.26, 0.49]
Agreeableness	−0.11	0.00	0.03	0.17*
	[−0.24, 0.02]	[−0.13, 0.14]	[−0.10, 0.17]	[0.04, 0.30]
Openness/intellect	0.12	0.11	0.16*	0.20**
	[−0.01, 0.25]	[−0.03, 0.24]	[0.03, 0.29]	[0.06, 0.32]
Extraversion	0.06	0.11	0.18**	0.13
	[−0.08, 0.19]	[−0.03, 0.24]	[0.05, 0.31]	[−0.01, 0.26]

Note: $N = 211$.
*$p < 0.05$; **$p < 0.01$.

TABLE 3 | Means of each personality trait rating in each method.

	Face-to-face interview		Asynchronous video interview		Written interview		Contextualized personality inventory	
	High	Remaining	High	Remaining	High	Remaining	High	Remaining
Averaged rating across all traits	3.88	3.77	3.99	3.87	3.94	3.78	4.19	3.89
Conscientiousness	3.84	3.63	4.01	3.85	4.02	3.76	4.41	4.08
Emotional stability	3.93	3.76	3.94	3.91	3.99	3.77	4.17	3.75
Agreeableness	3.44	3.53	3.64	3.68	3.59	3.57	4.24	3.98
Openness/intellect	3.99	3.92	4.24	3.91	4.00	3.91	4.19	3.99
Extraversion	4.21	3.99	4.09	3.83	4.09	3.89	3.94	3.67

Note: $N = 211$, with $n = 36$ participants in the high SDT group.

scales are applied in practice, we conducted supplementary analyses in which we systematically compared the mean interview rating scores across interview formats for interviewees with particularly high levels of SDT (i.e., one standard deviation above the mean) and the remaining interviewees. To this end, we computed a repeated measure ANOVA with SDT (high SDT vs. moderate/low SDT) as a between-subjects factor and with interview format (face-to-face interview vs. asynchronous video interview vs. written interview) and personality trait (extraversion vs. agreeableness vs. conscientiousness vs. emotional stability vs. openness/intellect) as within-subject factors. Table 3 shows descriptive statistics for mean trait levels in each interview format and the personality inventory for the high-SDT group and the moderate/low-SDT group separately.

The repeated measures ANOVA showed significant main effects of SDT ($F(1, 208) = 7.82$, $p < 0.001$), interview format ($F(2, 418) = 4.45$, $p = 0.001$), and trait ($F(3.74, 781) = 30.61$, $p < 0.001$) and significant interactions between SDT and trait ($F(3.74, 781.24) = 72.62$, $p = 0.004$) and between interview format and trait ($F(7.31, 1527.74) = 3.56$, $p < 0.001$). We report and discuss the main effect of the interview format and the interaction between the interview format and trait in more detail in the online supplement. Post hoc tests indicated significant differences between the high-SDT group and the moderate/low-SDT group in their interview ratings in the written interview ($p = 0.002$, Cohen's $d = 0.25$), the asynchronous video interview ($p = 0.011$, Cohen's $d = 0.23$), and the face-to-face interview ($p = 0.020$, Cohen's $d = 0.18$). This provides support for the assumption that SDT affects interview scores across different formats. It also replicates the pattern that the written interview seems to capture somewhat more SDT than the asynchronous video interview, followed by the face-to-face interview (see results for Hypothesis 1).

3.2.2 | Comparing Personality-Based Interviews With a Contextualized Personality Inventory

Given that personality traits are typically assessed using a self-report inventory, we were also interested in how the three (personality-based) interview formats compared to a traditional personality self-report inventory that is contextualized (i.e., adapted to the work context). The correlation between SDT and an overall score for the personality inventory ($r = 0.40$, $p < 0.001$) was significantly higher than correlations between SDT and the overall score for the face-to-face interviews ($r = 0.11$, $p < 0.126$; $z = -3.806$, $p = 0.038$), the asynchronous video interview ($r = 0.13$, $p = 0.052$; $z = -3.360$, $p < 0.001$), and the written interview ($r = 0.27$, $p < 0.001$; $z = -1.772$, $p < 0.001$; see Table 2).

Comparing which traits were most affected by SDT in the personality inventory revealed a clearer pattern than in the interviews. Consistent with the literature, we found significant correlations between SDT and conscientiousness ($r = 0.37$, $p < 0.001$) and between SDT and emotional stability ($r = 0.38$, $p < 0.001$). These correlations were significantly higher than those with agreeableness ($r = 0.17$, $p = 0.011$; $z = 2.59$, $p = 0.005$ compared to conscientiousness; $z = 2.62$, $p = 0.004$ compared to emotional stability), extraversion ($r = 0.13$, $p = 0.070$; $z = 3.01$, $p = 0.001$ compared to conscientiousness; $z = 3.36$, $p < 0.001$ compared to emotional stability), and openness/intellect ($r = 0.20$, $p = 0.003$; $z = 2.32$, $p = 0.010$ compared to conscientiousness; $z = 2.53$, $p = 0.006$ compared to emotional stability). An ANOVA that includes the personality inventory is presented in the Supporting Information Material.

4 | Discussion

With the variety of interviews available nowadays, it is necessary to understand the differences between interview formats to use each format to its best advantage. When measuring personality, interviews must be as robust as possible against influences of SDT to ensure an accurate measurement. The present study compared the extent to which SDT is associated with scores from personality-based interviews across three different interview formats (i.e., face-to-face interview, asynchronous video interview, and written interview).

4.1 | Main Findings and Theoretical Implications

The study's primary finding is that SDT affects personality ratings differently depending on the interview format. Higher

levels of SDT were associated with inflated personality scores, particularly in the written interview format. These findings demonstrate that interview formats are not interchangeable when it comes to measuring personality and that the choice of an interview format influences the extent to which SDT affects the assessment of personality traits. Conceptually, the findings might suggest that interview formats with higher synchrony and higher response richness have the advantage that they are less affected by interviewees' SDT.

The second key finding is that all interview formats are less strongly associated with SDT when compared to a contextualized personality inventory. Despite the systematic differences between interview formats, this comparison also implies that each interview format was still superior in its robustness against SDT when compared to a traditional self-report. Conceptually, the findings also imply that SDT can affect the measurement of personality across assessment tools and that an interview can reduce, but not eliminate, the effects of SDT on personality assessment. Nonetheless, our findings contribute to the considerable and long-standing efforts to reduce social desirability influences in selection (Cao and Drasgow 2019; Goffin and Christiansen 2003; Meehl and Hathaway 1946) and provide evidence for another advantage of using interviews as an alternative method for personality measurement.

The third key finding is that SDT affects ratings in the written interview differently depending on the trait being assessed, and this contrasts with the pattern identified for the other interview formats. Specifically, the written format was the only format for which we found the expected higher relationships between SDT and ratings for conscientiousness and emotional stability in comparison to the other Big Five traits. The fact that we did not find the same patterns for face-to-face or asynchronous video interviews could support our argument that the written interview places less cognitive demands on applicants, allowing them to put more effort into meeting specific expectations of a given situation, resulting in the effects of their SDT being more pronounced.

Beyond these main conclusions, the findings of the written interview may raise questions about its comparability with the other interview formats. In general, the variety of interview formats available today has complicated our understanding of what characterizes an interview. For example, while interactions are generally considered to be an essential characteristic of interviews (Levashina et al. 2014), the rise of AVIs has, to some degree, introduced the notion that such interactions may also be asynchronous. The written interview deviates the most from traditional face-to-face interviews, and although they are still sometimes treated as a variant of interviews (Whetzel et al. 2003), some researchers have argued that they are closer to a written test (Levashina et al. 2014). In fact, they show similarities to methods such as biodata (Breaugh 2009), which are considered written tests. Considering that the pattern of results in our study was comparable for the face-to-face interview and AVI but differed for the written interview, our findings add to these uncertainties in categorizing the written format. Ultimately, the different relationships between interview formats and SDT in this study may reflect differences beyond synchrony and response richness and may be considered as a hint that written interviews are not a classic representative of interview instruments.

4.2 | Limitations

A limitation of this study is that we collected data in a simulated selection setting, which is particularly noteworthy because SDT should be more strongly associated with assessments in high-stakes situations. Although we took several steps to ensure that the data collection closely resembled a real selection situation (e.g., providing a job description, asking participants to dress as they would for an interview), participants may still behave differently when they actually apply for a job (Watrin et al. 2022). Nevertheless, we chose to use a simulated selection setting because it allowed us to examine three interview formats in parallel using a within-subjects design, thus allowing for direct comparisons between interview formats. In support of this, interviewees reported that they perceived the training as realistic and behaved as they would in a real selection process.

In addition, and also related to the study design, the order in which the interview ratings and SDT were assessed does not allow for causal inference. In the selection setting, participants had to be unaware of the purpose of the study until after they had completed all interviews so as not to bias their responses. Therefore, SDT was measured at the end of the day and after all interviews had been completed. Although this reversed the order in which predictor and criterion constructs were measured, both SDT and the Big Five personality traits are considered relatively stable constructs that should not change within a short time span (Cobb-Clark and Schurer 2012; Paulhus 1984).

Further, it is important to acknowledge that SDT scales have been subject to criticism in the past. This criticism relates to the questions of what SDT scales measure (i.e., the subject vs. style debate; e.g., de Vries et al. 2014; Dunlop et al. 2012; McCrae and Costa 1983) and the mixed results on whether they are related to dishonest behavior or impact the prediction of job performance (Berry et al. 2007; Ellingson et al. 2001; Holden 2007; Hough 1998; Li et al. 2013; McCrae and Costa 1983; Ones et al. 1996; e.g., Xue et al. 2023). As a result of the uncertainties related to SDT scales, we want to be very clear that measuring SDT (i.e., a trait-like conscious or unconscious tendency to respond in a way that conforms with social norms) cannot be equated with measuring faking or impression management (i.e., a behavior). Instead, what the relationships with SDT in this study show is how prone an interview format is for measuring unwanted variance attributable to SDT and, therefore, inform on the suitability of these tools to accurately measure personality traits in selection.

Another limitation is that the relationship between SDT and the contextualized personality inventory can be inflated because of common method bias (Podsakoff et al. 2012). The SDT scale and the personality inventory were traditional self-report measures, whereas the interviews were rated by independent interviewers/raters. Similar methods (rating sources) may elicit similar response behaviors in respondents (e.g., response tendencies or biases; Bagozzi et al. 1991) and inflate their

relationships. Following recommendations to mitigate this effect, there was a time lag between both tools during the assessment day (Podsakoff et al. 2012). Further, personality was assessed during the simulated selection day, that is, in a high-stakes situation, whereas the SDT scale was assessed at the end of the day outside the simulated selection context, that is, in an evident research setting. These efforts may have been beneficial in reducing the impact of potential method-based inflation on the data. To balance what each method adds to the comparison, assessing SDT in a more interview-like manner (e.g., with open-ended questions about their desire to conform to social norms) might be an interesting approach in future replications.

4.3 | Future Research

This study uncovered differences in interview formats for their relationships with SDT. One possible explanation of the differences pertains to the cognitive load of applicants associated with each format. Notably, as this was not tested, we call for examining this potential explanatory mechanism along with other potential explanations in future research. For example, the degree of interaction between interview formats may not only affect the cognitive load but may also elicit different affective responses, such as interview anxiety (McCarthy and Goffin 2004), which may affect how comfortable applicants feel about sharing different desirable or undesirable information about themselves. Identifying the mechanisms underlying the effects of SDT on personality scores in interviews would help to understand and potentially address differences between formats.

Building on our comparison, future research may want to examine further consequences of using different interview formats, such as the impact of the interview format on criterion-related validity. The interview format may affect which information interviewees are willing and able to share. It might be interesting to investigate how the information each format captures is differently suited to predict job performance. To date, the effects of interviews measuring various constructs on criterion-related validity remain understudied (Wingate et al. 2023) and call for systematic comparisons.

Our findings further invite the exploration of additional, new method factors and their interaction with interviewees' inter-individual differences for selection outcomes. Building on existent theory (Daft and Lengel 1986; Lievens and Sackett 2017; Suen et al. 2019), we proposed that synchrony and response richness are method factors that distinguish the three interview formats. Yet, other factors derived from these or other theories could be examined. For example, technology-mediated selection tools come in many different formats (from phone interviews to AI-based chatbots), which may further influence applicants' cognitive demands and nervousness. How novel a method is to an applicant may also affect their interview behavior (Huffcutt et al. 2011), but such effects may also depend on their personality. For example, applicants who score higher on openness may remain calmer when faced with newer, innovative methods. Such interactions may be the subject of interesting future research.

In this study, we focused on personality-based interviews because SDT is rooted in personality research (e.g., Birkeland et al. 2006; Li and Bagger 2006; Viswesvaran and Ones 1999), and sticking to its traditional construct allowed us to compare its effect on interview formats with its effect on a personality inventory, which is most commonly researched in this area. From here, more research is needed that explores additional interview constructs and potential interactions between constructs and interview format.

4.4 | Practical Implications

For practitioners, results imply that using interviews instead of a traditional inventory for personality assessment can help to reduce the risk of capturing SDT instead of the intended personality construct, with face-to-face interviews being the interview format least affected by SDT. Our results suggest that using a written interview instead of a personality inventory already brings significant benefits for reducing such unwanted SDT influences while being comparable in its benefits (e.g., the flexibility and efficiency of asynchronous paper-pencil instruments). Given that personality assessments are often used at the preselection stage, such asynchronous formats represent a useful alternative in terms of reducing the unwanted effects of SDT.

One restriction that we want to mention at this point is that although written interviews show potential to reduce relationships with SDT, the research on them remains limited, making them a less predictable option. As outlined in the theoretical implications, written interviews are also not fully accepted as an interview format (Levashina et al. 2014). Despite the advantages they offer, the use of written interviews may, therefore, not provide all the benefits of established interview formats such as AVIs, where the implications for validity and applicant reactions are better understood. This is further aggravated by the increasing availability of advanced language models (LLMs) that provide applicants with accessible opportunities to draft written interview questions. Overall, organizations may consider relying directly on AVIs as the more thoroughly researched interview format.

Acknowledgments

The research project reported in this paper was supported by a grant from the Swiss National Science Foundation (Schweizer Nationalfonds zur Förderung der Wissenschaftlichen Forschung; Grant Number 179198).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The authors have nothing to report.

References

Arseneault, R., and N. Roulin. 2024. "Investigating Impression Management Use in Asynchronous Video Interviews Across 10 Countries." *International Journal of Selection and Assessment* 32, no. 3: 461–477. https://doi.org/10.1111/ijsa.12476.

Bagozzi, R. P., Y. Yi, and L. W. Phillips. 1991. "Assessing Construct Validity in Organizational Research." *Administrative Science Quarterly* 36, no. 3: 421–458. https://doi.org/10.2307/2393203.

Basch, J. M., F. Brenner, K. G. Melchers, et al. 2021. "A Good Thing Takes Time: The Role of Preparation Time in Asynchronous Video Interviews." *International Journal of Selection and Assessment* 29, no. 3–4: 378–392. https://doi.org/10.1111/ijsa.12341.

Basch, J. M., K. G. Melchers, and J. C. Büttner. 2022. "Preselection in the Digital Age: A Comparison of Perceptions of Asynchronous Video Interviews With Online Tests and Online Application Documents in a Simulation Context." *International Journal of Selection and Assessment* 30, no. 4: 639–652. https://doi.org/10.1111/ijsa.12403.

Basch, J. M., K. G. Melchers, A. Kurz, M. Krieger, and L. Miller. 2020. "It Takes More Than a Good Camera: Which Factors Contribute to Differences Between Face-to-Face Interviews and Videoconference Interviews Regarding Performance Ratings and Interviewee Perceptions?" *Journal of Business and Psychology* 36: 921–940. https://doi.org/10.1007/s10869-020-09714-3.

Berry, C. M., R. C. Page, and P. R. Sackett. 2007. "Effects of Self-Deceptive Enhancement on Personality–Job Performance Relationships." *International Journal of Selection and Assessment* 15, no. 1: 94–109. https://doi.org/10.1111/j.1468-2389.2007.00374.x.

Birkeland, S. A., T. M. Manson, J. L. Kisamore, M. T. Brannick, and M. A. Smith. 2006. "A Meta-Analytic Investigation of Job Applicant Faking on Personality Measures." *International Journal of Selection and Assessment* 14, no. 4: 317–335. https://doi.org/10.1111/j.1468-2389.2006.00354.x.

Bolino, M., D. Long, and W. Turnley. 2016. "Impression Management in Organizations: Critical Questions, Answers, and Areas for Future Research." *Annual Review of Organizational Psychology and Organizational Behavior* 3, no. 1: 377–406. https://doi.org/10.1146/annurev-orgpsych-041015-062337.

Bourdage, J. S., N. Roulin, and R. Tarraf. 2018. "'I (Might Be) Just That Good': Honest and Deceptive Impression Management in Employment Interviews." *Personnel Psychology* 71, no. 4: 597–632. https://doi.org/10.1111/peps.12285.

Bourdage, J. S., J. Schmidt, J. Wiltshire, B. Nguyen, and K. Lee. 2020. "Personality, Interview Performance, and the Mediating Role of Impression Management." *Journal of Occupational and Organizational Psychology* 93, no. 3: 556–577. https://doi.org/10.1111/joop.12304.

Breaugh, J. A. 2009. "The Use of Biodata for Employee Selection: Past Research and Future Directions." *Human Resource Management Review* 19, no. 3: 219–231. https://doi.org/10.1016/j.hrmr.2009.02.003.

Brenner, F. S., T. M. Ortner, and D. Fay. 2016. "Asynchronous Video Interviewing as a New Technology in Personnel Selection: The Applicant's Point of View." *Frontiers in Psychology* 7, no. 863: 863. https://doi.org/10.3389/fpsyg.2016.00863.

Cao, M., and F. Drasgow. 2019. "Does Forcing Reduce Faking? A Meta-Analytic Review of Forced-Choice Personality Measures in High-Stakes Situations." *Journal of Applied Psychology* 104, no. 11: 1347–1368. https://doi.org/10.1037/apl0000414.

Christiansen, N. D., C. Robie, G. N. Burns, R. W. Loy, A. B. Speer, and R. R. Jacobs. 2021. "Effects of Applicant Response Distortion on the Relationship Between Personality Trait Scores and Cognitive Ability." *Personality and Individual Differences* 171: 110542. https://doi.org/10.1016/j.paid.2020.110542.

Cobb-Clark, D. A., and S. Schurer. 2012. "The Stability of Big-Five Personality Traits." *Economics Letters* 115, no. 1: 11–15. https://doi.org/10.1016/j.econlet.2011.11.015.

Costa, P. T., and R. R. McCrae. 2008. "The Revised Neo Personality Inventory (Neo-Pi-r)." *SAGE Handbook of Personality Theory and Assessment* 2, no. 2: 179–198.

Crowne, D. P., and D. Marlowe. 1960. "A New Scale of Social Desirability Independent of Psychopathology." *Journal of Consulting Psychology* 24, no. 4: 349–354. https://doi.org/10.1037/h0047358.

Cui, T., C. C. S. Kam, E. H. Cheng, and M. Y. Ho. 2022. "Distinguishing Between Trait Desirability and Item Desirability in Predicting Item Scores: Is Informant Evaluation of Personality Free From Social Desirability?" *Personality and Individual Differences* 196: 111708. https://doi.org/10.1016/j.paid.2022.111708.

Daft, R. L., and R. H. Lengel. 1986. "Organizational Information Requirements, Media Richness and Structural Design." *Management Science* 32, no. 5: 554–571. https://doi.org/10.1287/mnsc.32.5.554.

Dennis, A. R., R. M. Fuller, and J. S. Valacich. 2008. "Media, Tasks, and Communication Processes: A Theory of Media Synchronicity." *MIS Quarterly* 32, no. 3: 575–600. https://doi.org/10.2307/25148857.

DeYoung, C. G., L. C. Quilty, and J. B. Peterson. 2007. "Between Facets and Domains: 10 Aspects of the Big Five." *Journal of Personality and Social Psychology* 93, no. 5: 880–896. https://doi.org/10.1037/0022-3514.93.5.880.

Dunlop, P. D., D. L. Morrison, J. Koenig, and B. Silcox. 2012. "Comparing the Eysenck and HEXACO Models of Personality in the Prediction of Adult Delinquency." *European Journal of Personality* 26, no. 3: 194–202. https://doi.org/10.1002/per.824.

Edinger, J. A., and M. L. Patterson. 1983. "Nonverbal Involvement and Social Control." *Psychological Bulletin* 93, no. 1: 30–56. https://doi.org/10.1037/0033-2909.93.1.30.

Edwards, A. L. 1957. *The Social Desirability Variable in Personality Assessment and Research*. Dryden Press.

Ellingson, J. E., P. R. Sackett, and L. M. Hough. 1999. "Social Desirability Corrections in Personality Measurement: Issues of Applicant Comparison and Construct Validity." *Journal of Applied Psychology* 84, no. 2: 155–166.

Ellis, A. P. J., B. J. West, A. M. Ryan, and R. P. DeShon. 2002. "The Use of Impression Management Tactics in Structured Interviews: A Function of Question Type?" *Journal of Applied Psychology* 87, no. 6: 1200–1208. https://doi.org/10.1037/0021-9010.87.6.1200.

Ellingson, J. E., D. B. Smith, and P. R. Sackett. 2001. "Investigating the Influence of Social Desirability on Personality Factor Structure." *Journal of Applied Psychology* 86, no. 1: 122–133. https://doi.org/10.1037/0021-9010.86.1.122.

Faul, F., E. Erdfelder, A. G. Lang, and A. Buchner. 2007. "G*Power 3: A Flexible Statistical Power Analysis Program for the Social, Behavioral, and Biomedical Sciences." *Behavior Research Methods* 39, no. 2: 175–191. https://doi.org/10.3758/bf03193146.

Feiler, A. R., and D. M. Powell. 2016. "Behavioral Expression of Job Interview Anxiety." *Journal of Business and Psychology* 31, no. 1: 155–171. https://doi.org/10.1007/s10869-015-9403-z.

Furnham, A. 1990. "Faking Personality Questionnaires: Fabricating Different Profiles for Different Purposes." *Current Psychology* 9, no. 1: 46–55. https://doi.org/10.1007/BF02686767.

Goffin, R. D., and N. D. Christiansen. 2003. "Correcting Personality Tests for Faking: A Review of Popular Personality Tests and an Initial Survey of Researchers." *International Journal of Selection and Assessment* 11, no. 4: 340–344. https://doi.org/10.1111/j.0965-075X.2003.00256.x.

Goldberg, L. R. 1990. "An Alternative 'Description of Personality': The Big-Five Factor Structure." *Journal of Personality and Social Psychology* 59, no. 6: 1216–1229.

Goldberg, L. R. 1992. "The Development of Markers for the Big-Five Factor Structure." *Psychological Assessment* 4, no. 1: 26–42. https://doi.org/10.1037/1040-3590.4.1.26.

Griffin, B., B. Hesketh, and D. Grayson. 2004. "Applicants Faking Good: Evidence of Item Bias in the NEO PI-R." *Personality and Individual*

Differences 36, no. 7: 1545–1558. https://doi.org/10.1016/j.paid.2003.06.004.

Griswold, K. R., J. M. Phillips, M. S. Kim, N. Mondragon, J. Liff, and S. M. Gully. 2022. "Global Differences in Applicant Reactions to Virtual Interview Synchronicity." *International Journal of Human Resource Management* 33, no. 15: 2991–3018. https://doi.org/10.1080/09585192.2021.1917641.

Guerrero, L. K. 2014. "Observer Ratings of Nonverbal Involvement and Immediacy." In *The Sourcebook of Nonverbal Measures*, 221–235. Psychology Press.

Heerwegh, D. 2009. "Mode Differences Between Face-to-Face and Web Surveys: An Experimental Investigation of Data Quality and Social Desirability Effects." *International Journal of Public Opinion Research* 21, no. 1: 111–121. https://doi.org/10.1093/ijpor/edn054.

Heimann, A. L., P. V. Ingold, M. E. Debus, and M. Kleinmann. 2021. "Who Will Go the Extra Mile? Selecting Organizational Citizens With a Personality-Based Structured Job Interview." *Journal of Business and Psychology* 36: 985–1007. https://doi.org/10.1007/s10869-020-09716-1.

Holden, R. R. 2007. "Socially Desirable Responding Does Moderate Personality Scale Validity Both in Experimental and in Nonexperimental Contexts." *Canadian Journal of Behavioural Science/Revue Canadienne Des Sciences Du Comportement* 39, no. 3: 184–201. https://doi.org/10.1037/cjbs2007015.

Hough, L. M. 1998. "Effects of Intentional Distortion in Personality Measurement and Evaluation of Suggested Palliatives." *Human Performance* 11, no. 2–3: 209–244. https://doi.org/10.1080/08959285.1998.9668032.

Hu, J., and B. S. Connelly. 2021. "Faking by Actual Applicants on Personality Tests: A Meta-Analysis of Within-Subjects Studies." *International Journal of Selection and Assessment* 29, no. 3–4: 412–426. https://doi.org/10.1111/ijsa.12338.

Huffcutt, A. I., C. H. Van Iddekinge, and P. L. Roth. 2011. "Understanding Applicant Behavior in Employment Interviews: A Theoretical Model of Interviewee Performance." *Human Resource Management Review* 21, no. 4: 353–367. https://doi.org/10.1016/j.hrmr.2011.05.003.

Ingold, P. V., M. Kleinmann, C. J. König, and K. G. Melchers. 2015. "Shall We Continue or Stop Disapproving of Self-Presentation? Evidence on Impression Management and Faking in a Selection Context and Their Relation to Job Performance." *European Journal of Work and Organizational Psychology* 24, no. 3: 420–432. https://doi.org/10.1080/1359432X.2014.915215.

Ingold, P. V., M. Kleinmann, C. J. König, and K. G. Melchers. 2016. "Transparency of Assessment Centers: Lower Criterion-Related Validity But Greater Opportunity to Perform?" *Personnel Psychology* 69, no. 2: 467–497. https://doi.org/10.1111/peps.12105.

International Personality Item Pool. 2019. "Big-Five Factor Markers." https://ipip.ori.org/newBigFive5broadKey.htm.

Janz, T. 1982. "Initial Comparisons of Patterned Behavior Description Interviews Versus Unstructured Interviews." *Journal of Applied Psychology* 67, no. 5: 577–580. https://doi.org/10.1037/0021-9010.67.5.577.

Kanning, U. P., and S. Kuhne. 2006. "Social Desirability in a Multimodal Personnel Selection Test Battery." *European Journal of Work and Organizational Psychology* 15, no. 3: 241–261. https://doi.org/10.1080/13594320600625872.

King, M. F., and G. C. Bruner. 2000. "Social Desirability Bias: A Neglected Aspect of Validity Testing." *Psychology & Marketing* 17, no. 2: 79–103. https://doi.org/10.1002/(SICI)1520-6793(200002)17:2<79::AID-MAR2>3.0.CO;2-0.

Klehe, U.-C., M. Kleinmann, T. Hartstein, et al. 2012. "Responding to Personality Tests in a Selection Context: The Role of the Ability to Identify Criteria and the Ideal-Employee Factor." *Human Performance* 25, no. 4: 273–302. https://doi.org/10.1080/08959285.2012.703733.

Klehe, U.-C., C. J. König, G. M. Richter, M. Kleinmann, and K. G. Melchers. 2008. "Transparency in Structured Interviews: Consequences for Construct and Criterion-Related Validity." *Human Performance* 21, no. 2: 107–137. https://doi.org/10.1080/08959280801917636.

Klehe, U.-C., and G. Latham. 2006. "What Would You Do—Really or Ideally? Constructs Underlying the Behavior Description Interview and the Situational Interview in Predicting Typical Versus Maximum Performance." *Human Performance* 19, no. 4: 357–382. https://doi.org/10.1207/s15327043hup1904_3.

Kluemper, D. H., B. D. McLarty, T. R. Bishop, and A. Sen. 2015. "Interviewee Selection Test and Evaluator Assessments of General Mental Ability, Emotional Intelligence and Extraversion: Relationships With Structured Behavioral and Situational Interview Performance." *Journal of Business and Psychology* 30, no. 3: 543–563. https://doi.org/10.1007/s10869-014-9381-6.

Levashina, J., and M. A. Campion. 2006. "A Model of Faking Likelihood in the Employment Interview." *International Journal of Selection and Assessment* 14, no. 4: 299–316. https://doi.org/10.1111/j.1468-2389.2006.00353.x.

Levashina, J., C. J. Hartwell, F. P. Morgeson, and M. A. Campion. 2014. "The Structured Employment Interview: Narrative and Quantitative Review of the Research Literature." *Personnel Psychology* 67, no. 1: 241–293. https://doi.org/10.1111/peps.12052.

Li, A., and J. Bagger. 2006. "Using the BIDR to Distinguish the Effects of Impression Management and Self-Deception on the Criterion Validity of Personality Measures: A Meta-Analysis." *International Journal of Selection and Assessment* 14, no. 2: 131–141. https://doi.org/10.1111/j.1468-2389.2006.00339.x.

Li, A., J. Bagger, and W. Friske. 2013. "Social Desirability in the Selection Process: New Insights From a Novel Context." *Asia Pacific Journal of Human Resources* 51, no. 1: 45–62. https://doi.org/10.1111/j.1744-7941.2012.00032.x.

Lievens, F., W. De Corte, and E. Schollaert. 2008. "A Closer Look at the Frame-of-Reference Effect in Personality Scale Scores and Validity." *Journal of Applied Psychology* 93, no. 2: 268–279. https://doi.org/10.1037/0021-9010.93.2.268.

Lievens, F., W. De Corte, and L. Westerveld. 2015. "Understanding the Building Blocks of Selection Procedures: Effects of Response Fidelity on Performance and Validity." *Journal of Management* 41, no. 6: 1604–1627. https://doi.org/10.1177/0149206312463941.

Lievens, F., and P. R. Sackett. 2017. "The Effects of Predictor Method Factors on Selection Outcomes: A Modular Approach to Personnel Selection Procedures." *Journal of Applied Psychology* 102, no. 1: 43–66. https://doi.org/10.1037/apl0000160.

Lukacik, E.-R., J. S. Bourdage, and N. Roulin. 2022. "Into the Void: A Conceptual Model and Research Agenda for the Design and Use of Asynchronous Video Interviews." *Human Resource Management Review* 32, no. 1: 100789. https://doi.org/10.1016/j.hrmr.2020.100789.

Martín-Raugh, M. P., H. J. Kell, J. G. Randall, C. Anguiano-Carrasco, and J. T. Banfi. 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, no. 1: 132–156. https://doi.org/10.1002/job.2670.

McCarthy, J., and R. Goffin. 2004. "Measuring Job Interview Anxiety: Beyond Weak Knees and Sweaty Palms." *Personnel Psychology* 57, no. 3: 607–637. https://doi.org/10.1111/j.1744-6570.2004.00002.x.

McCrae, R. R., and P. T. Costa. 1983. "Social Desirability Scales: More Substance Than Style." *Journal of Consulting and Clinical Psychology* 51, no. 6: 882–888. https://doi.org/10.1037/0022-006X.51.6.882.

Meehl, P. E., and S. R. Hathaway. 1946. "The K Factor as a Suppressor Variable in the Minnesota Multiphasic Personality Inventory." *Journal of Applied Psychology* 30, no. 5: 525–564. https://doi.org/10.1037/h0053634.

Melchers, K. G., N. Roulin, and A. K. Buehl. 2020. "A Review of Applicant Faking in Selection Interviews." *International Journal of Selection and Assessment* 28, no. 2: 123–142. https://doi.org/10.1111/ijsa.12280.

Mueller-Hanson, R., E. D. Heggestad, and G. C. Thornton Iii. 2003. "Faking and Selection: Considering the Use of Personality From Select-in and Select-Out Perspectives." *Journal of Applied Psychology* 88, no. 2: 348–355. https://doi.org/10.1037/0021-9010.88.2.348.

Müller, S., and M. Moshagen. 2019. "True Virtue, Self-Presentation, or Both?: A Behavioral Test of Impression Management and Over-claiming." *Psychological Assessment* 31, no. 2: 181–191. https://doi.org/10.1037/pas0000657.

Nederström, M., and K. Salmela-Aro. 2014. "Self-Other Agreement of Personality Judgments in Job Interviews: Exploring the Effects of Trait, Gender, Age and Social Desirability." *Scandinavian Journal of Psychology* 55, no. 5: 520–526. https://doi.org/10.1111/sjop.12154.

Ones, D. S., C. Viswesvaran, and A. D. Reiss. 1996. "Role of Social Desirability in Personality Testing for Personnel Selection: The Red Herring." *Journal of Applied Psychology* 81, no. 6: 660–679. https://doi.org/10.1037/0021-9010.81.6.660.

Paulhus, D. L. 1984. "Two-Component Models of Socially Desirable Responding." *Journal of Personality and Social Psychology* 46, no. 3: 598–609. https://doi.org/10.1037/0022-3514.46.3.598.

Paunonen, S. V., and E. P. LeBel. 2012. "Socially Desirable Responding and Its Elusive Effects on the Validity of Personality Assessments." *Journal of Personality and Social Psychology* 103, no. 1: 158–175. https://doi.org/10.1037/a0028165.

Podsakoff, P. M., S. B. MacKenzie, and N. P. Podsakoff. 2012. "Sources of Method Bias in Social Science Research and Recommendations on How to Control It." *Annual Review of Psychology* 63, no. 1: 539–569. https://doi.org/10.1146/annurev-psych-120710-100452.

Rasmussen, K. G. 1984. "Nonverbal Behavior, Verbal Behavior, Resumé Credentials, and Selection Interview Outcomes." *Journal of Applied Psychology* 69, no. 4: 551–556. https://doi.org/10.1037/0021-9010.69.4.551.

Reid-Seiser, H. L., and B. A. Fritzsche. 2001. "The Usefulness of the NEO PI-R Positive Presentation Management Scale for Detecting Response Distortion in Employment Contexts." *Personality and Individual Differences* 31, no. 4: 639–650. https://doi.org/10.1016/S0191-8869(00)00168-9.

Rhim, J., M. Kwak, Y. Gong, and G. Gweon. 2022. "Application of Humanization to Survey Chatbots: Change in Chatbot Perception, Interaction Experience, and Survey Data Quality." *Computers in Human Behavior* 126: 107034. https://doi.org/10.1016/j.chb.2021.107034.

Richman, W. L., S. Kiesler, S. Weisband, and F. Drasgow. 1999. "A Meta-Analytic Study of Social Desirability Distortion in Computer-Administered Questionnaires, Traditional Questionnaires, and Interviews." *Journal of Applied Psychology* 84, no. 5: 754–775. https://doi.org/10.1037/0021-9010.84.5.754.

Rosse, J. G., M. D. Stecher, J. L. Miller, and R. A. Levin. 1998. "The Impact of Response Distortion on Preemployment Personality Testing and Hiring Decisions." *Journal of Applied Psychology* 83, no. 4: 634–644. https://doi.org/10.1037/0021-9010.83.4.634.

Roulin, N., and F. Krings. 2020. "Faking to Fit In: Applicants' Response Strategies to Match Organizational Culture." *Journal of Applied Psychology* 105, no. 2: 130–145. https://doi.org/10.1037/apl0000431.

Roulin, N., and D. M. Powell. 2018. "Identifying Applicant Faking in Job Interviews: Examining the Role of Criterion-Based Content Analysis and Storytelling." *Journal of Personnel Psychology* 17, no. 3: 143–154. https://doi.org/10.1027/1866-5888/a000207.

Roulin, N., O. Wong, M. Langer, and J. S. Bourdage. 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 and Organizational Psychology* 32, no. 3: 333–345. https://doi.org/10.1080/1359432X.2022.2156862.

Sackett, P. R., C. Zhang, C. M. Berry, and F. Lievens. 2022. "Revisiting Meta-Analytic Estimates of Validity in Personnel Selection: Addressing Systematic Overcorrection for Restriction of Range." *Journal of Applied Psychology* 107, no. 11: 2040–2068. https://doi.org/10.1037/apl0000994.

Sackett, P. R., C. Zhang, C. M. Berry, and F. Lievens. 2023. "Revisiting the Design of Selection Systems in Light of New Findings Regarding the Validity of Widely Used Predictors." *Industrial and Organizational Psychology* 16, no. 3: 283–300. https://doi.org/10.1017/iop.2023.24.

Salgado, J. F., and S. Moscoso. 2002. "Comprehensive Meta-Analysis of the Construct Validity of the Employment Interview." *European Journal of Work and Organizational Psychology* 11, no. 3: 299–324. https://doi.org/10.1080/13594320244000184.

Satow, L. 2012. "SEA. Skala Zur Erfassung Von Testverfälschung Durch Positive Selbstdarstellung Und Sozial Erwünschte Antworttendenzen." In *Open Test Archive*, edited by L.-I. F P. (ZPID). ZPID. https://doi.org/10.23668/psycharchives.6890.

Schmit, M. J., and A. M. Ryan. 1993. "The Big Five in Personnel Selection: Factor Structure in Applicant and Nonapplicant Populations." *Journal of Applied Psychology* 78, no. 6: 966–974. https://doi.org/10.1037/0021-9010.78.6.966.

Smith, D. B., and J. E. Ellingson. 2002. "Substance Versus Style: A New Look at Social Desirability in Motivating Contexts." *Journal of Applied Psychology* 87, no. 2: 211–219. https://doi.org/10.1037/0021-9010.87.2.211.

Snell, A. F., E. J. Sydell, and S. B. Lueke. 1999. "Towards a Theory of Applicant Faking: Integrating Studies of Deception." *Human Resource Management Review* 9, no. 2: 219–242. https://doi.org/10.1016/S1053-4822(99)00019-4.

Stivers, T., N. J. Enfield, P. Brown, et al. 2009. "Universals and Cultural Variation in Turn-Taking in Conversation." *Proceedings of the National Academy of Sciences of the United States of America* 106, no. 26: 10587–10592.

Suen, H.-Y., M. Y.-C. Chen, and S.-H. Lu. 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.

Swider, B. W., M. R. Barrick, and T. B. Harris. 2016. "Initial Impressions: What They Are, What They Are Not, and How They Influence Structured Interview Outcomes." *Journal of Applied Psychology* 101, no. 5: 625–638. https://doi.org/10.1037/apl0000077.

Templeton, E. M., L. J. Chang, E. A. Reynolds, M. D. Cone LeBeaumont, and T. Wheatley. 2022. "Fast Response Times Signal Social Connection in Conversation." *Proceedings of the National Academy of Sciences of the United States of America* 119, no. 4: 1–8. https://doi.org/10.1073/pnas.2116915119.

Tett, R. P., M. J. Toich, and S. B. Ozkum. 2021. "Trait Activation Theory: A Review of the Literature and Applications to Five Lines of Personality Dynamics Research." *Annual Review of Organizational Psychology and Organizational Behavior* 8: 199–233. https://doi.org/10.1146/annurev-orgpsych-012420-062228.

Tu, M. H., E. K. Gilbert, and J. E. Bono. 2022. "Is Beauty More Than Skin Deep? Attractiveness, Power, and Nonverbal Presence in Evaluations of Hirability." *Personnel Psychology* 75, no. 1: 119–146. https://doi.org/10.1111/peps.12469.

Van Iddekinge, C. H., P. H. Raymark, and P. L. Roth. 2005. "Assessing Personality With a Structured Employment Interview: Construct-Related Validity and Susceptibility to Response Inflation." *Journal of Applied Psychology* 90, no. 3: 536–552. https://doi.org/10.1037/0021-9010.90.3.536.

Van Iddekinge, C. H., P. H. Raymark, P. L. Roth, and H. S. Payne. 2006. "Comparing the Psychometric Characteristics of Ratings of Face-to-Face and Videotaped Structured Interviews." *International Journal of Selection and Assessment* 14: 347–359.

Viswesvaran, C., and D. S. Ones. 1999. "Meta-Analyses of Fakability Estimates: Implications for Personality Measurement." *Educational and Psychological Measurement* 59, no. 2: 197–210.

de Vries, R. E., I. Zettler, and B. E. Hilbig. 2014. "Rethinking Trait Conceptions of Social Desirability Scales: Impression Management as an Expression of Honesty-Humility." *Assessment* 21, no. 3: 286–299. https://doi.org/10.1177/1073191113504619.

Watrin, L., L. Weihrauch, and O. Wilhelm. 2022. "The Criterion-Related Validity of Conscientiousness in Personnel Selection: A Meta-Analytic Reality Check." *International Journal of Selection and Assessment* 31, no. 2: 1–16. https://doi.org/10.1111/ijsa.12413.

Whetzel, D. L., L. E. Baranowski, J. M. Petro, P. J. Curtin, and J. L. Fisher. 2003. "A Written Structured Interview by Any Other Name Is Still a Selection Instrument." *Applied H.R.M. Research* 8, no. 1–2: 1–16.

Wingate, T., J. S. Bourdage, and P. Steel. 2023. "Valid for What? A Meta-Analysis of Construct-Specific Interview Criterion-Related Validity." *Academy of Management Proceedings* 2023, no. 1: 11827. https://doi.org/10.5465/AMPROC.2023.79bp.

Xue, J., L. Zheng, X. Liu, and J. Fan. 2023. "On the Effect of Warnings on Criterion-Related Validity of Personality Scores: Testing an Indirect Moderation Model." *Personality and Individual Differences* 208: 112191. https://doi.org/10.1016/j.paid.2023.112191.

Zettler, I., B. E. Hilbig, M. Moshagen, and R. E. de Vries. 2015. "Dishonest Responding or True Virtue? A Behavioral Test of Impression Management." *Personality and Individual Differences* 81: 107–111. https://doi.org/10.1016/j.paid.2014.10.007.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Appendix A

Sample Interview Question With Behavior-Anchored Rating Scale

A sample interview question for measuring the conscientiousness item "At work, I follow the rules" is: "Think of a situation where you had to follow certain guidelines at work or in your studies that you explicitly agreed to, such as quality guidelines in a company or safety guidelines in a building. However, these guidelines hindered you at work, for example, because they complicated or prolonged processing. There was no way to modify the guidelines. Please tell us briefly in one or two sentences what the situation was. Then report how you experienced the situation in terms of compliance with these guidelines and how you acted." In this example, there is a dilemma between sticking to the guidelines and accepting the negative impact it has on one's work (higher expression on the item) or taking a more pragmatic approach and accomplishing work more efficiently, thereby bypassing the guidelines (lower expression on the item). Because of the dilemma, responses indicating low expression on the item are still socially desirable.

For the example presented above, anchors were: 1–does not find it important to comply with the agreement (does not adhere to the guidelines, does not try to comply, disobeys), is not willing to accept disadvantages, 3–feels inner resistance or guilt when ignoring the agreement, basically tries to adhere to the agreement (roughly adheres to the guidelines, deviates from them to a certain degree), accepts certain disadvantages in doing so (e.g., somewhat more complicated processing), and 5–perceives it as very important to adhere to the agreement (adheres to the guidelines), tries to adhere to it by all means, also accepts disadvantages in doing so (e.g., circumstances, more complicated work processes).