Fachbereich Erziehungswissenschaft und Psychologie

der Freien Universität Berlin

Asynchronous Video Interviews in Selection: A Systematic Review and Five Empirical

Investigations

Dissertation

zur Erlangung des akademischen Grades

Doctor of Philosophy (Ph.D.)

vorgelegt von

Dipl.-Psych. Falko S. Brenner

Berlin, Juni 2019

ASYNCHRONOUS VIDEO INTERVIEWS

Primary Advisor: Prof. Dr. Stefan Krumm

(Erstgutachter)

Secondary Advisor: Prof. Dr. Klaus Melchers

(Zweitgutachter)

Date of Defense: 15.11.2019

(Tag der Verteidigung)

Table of Contents

Abstract

It has become very popular to conduct employment interviews using new digital technologies, including synchronous and even asynchronous video interviews. In contrast to this wide usage in practice, less is known about how these technologies influence psychometric properties and selection outcomes. Findings gained from traditional face-to-face interviews cannot easily be transferred to digital interview formats. In addition, scholars have recently called for increased theory development to overcome the numerous limitations of cross-media comparisons and explain why and how technology should influence selection outcomes. While a variety of theories on media usage, media choice and media adaption exist, their applicability and explanatory value for technology usage in the area of employment interviews is rarely addressed. The present dissertation addresses this notable gap in the literature through both a systematic review of conceptional frameworks on technology usage in employment interviews as well as empirical results on understudied digital interview formats. In doing so, it identifies promising avenues for future research and provides information for HR practitioners about how to design their selection systems.

More specifically, the current dissertation encompasses a comprehensive review of technology usage in employment interviews and five empirical studies on the specific format of asynchronous video interviews. The review integrates several theoretical perspectives on the topic, including the unitary perspective on technology-enhanced interviews, major theoretical directions in media research, and research on differences in psychometric properties and selection outcomes due to technology, into a comprehensive working model. The empirical portion of the dissertation presents five exploratory studies on asynchronous video interviews that explore research questions on blind spots in the literature or address urgent issues concerning the use of technology in today's selection practice. The first study addresses interrater agreement and the importance of structured evaluation formats in a non-applicant sample of $N = 111$ participants. The second study investigates the influence of social

bandwidth on the accuracy of interview ratings in a sample of non-applicants with $N = 279$ participants. The third study provides first results on the validity of asynchronous video interviews in the field of high-stakes selection with $N = 899$ real applicants. The fourth study assesses the impact of personalized communication via video messages on applicant reactions in asynchronous video interviews with a non-applicant sample of $N = 98$ participants. Finally, the fifth study explores rating inflation due to preparation time in a non-applicant sample of $N = 51$ participants.

This dissertation contributes to the literature in several ways: The review provides an up-to-date, multi-perspective overview of the field and integrates several previous research strands into a single framework on technology usage in employment interviews. The empirical studies provide promising initial results concerning the psychometric properties of asynchronous video interviewing, specifically with respect to reliability and validity, but also highlight possible pitfalls--like rating inflation--that might appear when preparation time is introduced into the process design. In addition, the studies further highlight the extraordinary importance of structure in interviews, even though further study is required to better understand the exact nature of the relationship between structure and technology in employment interviews.

Keywords: technology in selection, asynchronous video interviews, digital interviewing, employment interview

List of Abbreviations

AC – assessment center

ATIC – ability to identify criteria

AVI – asynchronous video interview

FtF – face-to-face

IM – impression management

ML – maximum likelihood

MRT – media richness theory

MST – media synchronicity theory

PEoU – perceived ease of use

PU – perceived usefulness

RJP – realistic job preview

SJT – situational judgment test

TAM – technology acceptance model

TI – telephone interview

UIT – unproctored Internet testing

VI – video interview

List of Figures

List of Tables

1. General Introduction

1.1. Focus and goals

1.1.1. Current technological advances in personnel selection

The question of what employees to recruit and select has generated such consistent and extensive interest among academics and practitioners in industrial and organizational (I/O) psychology that some scholars have even called it the *supreme problem* in the field (Ployhart, Schmitt, & Tippins, 2017). In recent years, technology has become the most important force and game changer in the selection and recruitment domain, with hundreds of new applications coming onto the market every year (Chamorro-Premuzic, Winsborough, Sherman, & Hogan, 2016; Tippins, 2015). Considering that digital technology infrastructure is just as important today as electricity was a century ago, there is surprisingly little research on how the digital revolution is affecting today's workplace (Cascio & Montealegre, 2016). Digital interviews, voice profiling[1], Big Data and social media analysis, gamification[2], professional networks, and crowdsourced reputation/peer ratings[3] are currently changing the way organizations select and onboard new staff (Chamorro-Premuzic et al., 2016). Due to the pace of these developments and the lack of scientific research on recent innovations in employee selection, HR practitioners are left without credible evidence or even feasible guidelines on the utility of these new technologies. The present work contributes to rectifying this situation by addressing one such technological advance in selection, the use of (asynchronous) video technology for employment interviews, and how it impacts selection outcomes.

---

[1] Technology to extract affective state or personality traits automatically from a voice sample
[2] Use of design principles from computer games in computer-based assessments
[3] Peer evaluations of management behavior on a public website (e.g., Glassdoor)

Table 1.1. *Comparison of old and new selection and assessment methodologies by Chamorro-Premuzic et al. (2016, p. 627)*

Old methods	New tools	Dimension assessed
Interviews	Digital interviews Voice profiling	Experise, social skills, motivation, and intelligence
Biodata Supervisory ratings IQ	Big data (internal)	Past performance Current performance Intelligence, job-related knowledge
Situational judgment tests Self-reports	Gamification	knowledge, and Big Five personality traits or minor traits
Self-reports	Social media analytics	Big Five personality traits and values (identity claims)
Resumes References	Professional socia networks (LinkedIn)	Experience, past performance, and technical skills and qualifications
360s	Crowdsourced reputation/peer-rating	Any personality trait, competencies, and reputation reputation

Despite the substantial technological progress in selection and assessment methods depicted in Table 1.1, employment interviews continue to be the most popular and most widely used tool for selecting new employees (Levashina, Hartwell, Morgeson, & Campion, 2014). Furthermore, they are often the only method used to evaluate a candidate's suitability for a given job. Traditionally, the employment interview was defined as a face-to-face interaction conducted to gain insights about a candidate's qualifications for a particular job (Huffcutt & Youngcourt, 2007). Ample research has attested to the favorable psychometric properties of structured face-to-face employment interviews (Conway, Jako, & Goodman, 1995; Huffcutt & Youngcourt, 2007; Levashina et al., 2014; Macan, 2009; Moscoso, 2000; Posthuma, Morgeson, & Campion, 2002). However, the traditional face-to-face (FtF) interview has evolved over the last two decades, driven by technological progress as well as

the emergence of a more globally connected workforce, the advancing digitalization of recruitment and selection systems, and organizations' ongoing need for effective recruitment processes. Specifically, several technology-enhanced variations of face-to-face employment interviews have been developed, including telephone interviews (Blackman, 2002; Oliphant, Hansen, & Oliphant, 2008; Silvester & Anderson, 2003; Silvester, Anderson, Haddleton, Cunningham-Snell, & Gibb, 2000), interviews conducted via videoconferencing technology (Chapman & Rowe, 2001, 2002; Sears, Zhang, Wiesner, Hackett, & Yuan, 2013; Straus, Miles, & Levesque, 2001), and asynchronous video interviews, in which decision makers evaluate recorded, time-delayed video segments in which candidates answer job-related interview questions without any direct interaction between interviewers and interviewees (Brenner, Ortner, & Fay, 2016; Gorman, Robinson, & Gamble, 2018; Langer, Konig, & Krause, 2017; Torres & Mejia, 2017).

However, our understanding of the impact of technology on selection outcomes in employment interviews is still very limited because the results obtained for face-to-face interviews cannot be easily transferred to technology-enhanced interview formats. Indeed, studies directly comparing face-to-face interviews and recorded interviews have reported significant differences in reliability and construct validity (Van Iddekinge, Raymark, Roth, & Payne, 2006). Moreover, initial research comparing two synchronous and asynchronous video interview formats found significant differences in interview scores and applicant reactions, with candidates in Skype interviews receiving lower ratings and asynchronous video interviews perceived less positively by candidates than synchronous interviews (Langer et al., 2017). Thus, there are systematic differences in interview outcomes resulting from the medium used to conduct the interview. In addition, there is little research on the psychometric properties of the emerging asynchronous video interview formats.

Nevertheless, ever since communication technology became common in the workplace, researchers have been interested in the systematic impacts of media characteristics on

communication outcomes, organizational performance, and selection decisions. Thus, several theoretical pathways and conceptual frameworks on media choice, media impacts and human-media interaction have been developed in recent decades to explain why and how specific outcomes are affected by the use of a given communication technology and its medial configuration. For instance, studies conducted in the seventies found that people are perceived as less *socially present* when they use telecommunications technology (telephone, videoconferencing) rather than face-to-face communication, which in turn has a wide range of behavioral impacts, such as a more problem-oriented communication style and the use of less open questions (Short, Williams, & Christie, 1976). The *media richness theory* tradition proposes that communication media differs with respect to distinct attributes or characteristics, such as interactivity or transparency, which impacts the choice or performance of a given medium for a given organizational task like judging a candidate's qualifications (Daft & Lengel, 1986; Dennis, Fuller, & Valacich, 2008; Potosky, 2008). The *technology acceptance model* suggests that the two core constructs of perceived usefulness and perceived ease of use can explain whether a user adopts or rejects a new selection technology (Davis, 1989; Davis, Bagozzi, & Warshaw, 1989). Recently, Arthur, Keiser, and Doverspike (2018) introduced an *information processing*-based framework identifying structural characteristics which are proposed to affect test scores when test-takers use different devices. Finally, the *socio-technical* or social material perspective emphases the interaction between technology and human beings as an important factor for understanding technological impacts (Cooper & Foster, 1971; Orlikowski, 2007). Media impact theorists anticipated media effects on selection outcomes with varying degrees of explicitness. While Potosky's (2008) and Arthur et al.'s (2018) frameworks were explicitly developed for the selection domain, the others are formulated more generally, making them equally applicable to the selection domain and other domains. While Potosky (2008) formulated some general propositions on the impact of media attributes on selection outcomes, she did not strive to elaborate her conceptional framework

for different forms of employment interviews. Moreover, other strands of media theory might also be relevant for understanding the impact of technology on employment interview outcomes. Furthermore, Lievens and Sackett (2017) argue that the unitary perspective on selection procedures, i.e., "telephone interviews" or "video interviews" in general, is no longer appropriate because the applications currently on the market have reached maturity and they are highly modular with many possible combinations of features. Thus, the existing body of empirical findings must be reevaluated; a new synthesis of media theories and research on selection outcomes in technology-enhanced interviews could assist in gaining a more modular understanding of technological changes such as asynchronous interview formats and making it easier for both academics and practitioners to make predictions.

1.1.2. Goals and focus of this thesis

This thesis has two goals: The first is to synthesize the body of existing literature on media theories and employment interviews in order to develop a pragmatic conceptional framework and derive testable hypotheses concerning how technology affects the psychometric properties of job interviews. The second goal is to present the results of studies with five independent samples testing specific research hypotheses concerning asynchronous video interviews and evaluate how the results fit with the framework developed in the first part of the thesis. These empirical studies were intentionally not designed as comparative cross-media studies on the measurement equivalence between asynchronous video interviews underlying equivalence studies is questionable or arbitrary, as a new technology might incorporate unique aspects by design, meaning that no "traditional" measure for equivalence testing is available (Morelli, Potosky, Arthur, & Tippins, 2017). In other words, equivalence studies treat selection procedures as holistic, uniform solutions. This point of view is outdated given that today's interview and assessment solutions have become increasingly modular, meaning that the usage of asynchronous video interviews in selection processes can differ drastically between vendors, organizations or even within a single organization. Subsequently,

a modular approach is more appropriate for formulating hypotheses on the impact of specific features of assessment procedures (Lievens, De Corte, & Westerveld, 2015). Finally, the second, narrower goal of this thesis is to apply a modular approach to investigate the consequences of specific design principles in media usage. This is highly relevant for assessment practice as a means of responding to the call to close the research-practitioner gap in the field of assessment innovation (Gibson, Vaughn, & Hudy, 2017).

This review part is structured as follows: First, the general definition of an employment interview is discussed in light of technological progress as well as structural features and classifications. Next, the range of technologies commonly used in conducting employment interviews, namely telephone interviews and synchronous and asynchronous video interviews, will be described in the unitary manner in which they have been treated in the existing literature. In addition, similarities and differences between interviews and other common selection procedures such as video-based situational judgment tests (SJT) and video resumes will be discussed with an eye to determining the extent to which findings on these procedures are transferable to the interview domain. The second chapter provides an overview of theoretical and conceptual frameworks that have previously been applied to explain differences in outcomes with respect to media usage, while the third chapter examines existing empirical research in relation to hypotheses derived from media theories on outcomes including reliability, validity, applicant reactions, adverse impact, and utility. At the end of the theory section, in Chapter 4, a working model of a modular approach to technology in employment interviews will be presented along with research hypotheses for five initial studies investigating asynchronous forms of video interviewing.

1.1.3 Need for a review and inclusion criteria for studies

Traditional face-to-face employment interviews have been the subject of research for more than a century, leading to multiple narrative reviews (Arvey & Campion, 1982; Harris, 1989; Levashina et al., 2014; Macan, 2009; Mayfield, 1964; Posthuma et al., 2002; Schmitt,

1976; Ulrich & Trumbo, 1965; O. R. Wright, Jr., 1969) and more than a dozen meta-analyses (Conway et al., 1995; Huffcutt & Arthur, 1994; Huffcutt, Conway, Roth, & Klehe, 2004; Huffcutt, Culbertson, & Weyhrauch, 2014; Hunter & Hunter, 1984; Latham & Sue-Chan, 1999; Marchese & Muchinsky, 1993; McDaniel, Whetzel, Schmidt, & Maurer, 1994; Reilly & Chao, 1982; Schmidt & Hunter, 1998; Schmidt & Rader, 1999; Schmidt & Zimmerman, 2004; Thorsteinson, 2018; Wiesner & Cronshaw, 1988; P. M. Wright, Lichtenfels, & Pursell, 1989). In contrast to this impressive body of literature, that latest review by Levashina et al. (2014) was the first to (briefly) address the issue of technology usage. However, the authors considered this issue a minor focus did not devote a unique section to it, despite the fact that technology usage could have drastic effects on every other topic discussed in the review. The first and only meta-analysis on differences between face-to-face and technology-enhanced interviews by Blacksmith, Willford, and Behrend (2016) synthesized twelve primary studies up to 2007, finding that median interview ratings ($d = -.41$) and applicants' reactions ($d = -.36$) were lower in technology-enhanced interviews compared to face-to-face. However, the authors themselves noted that their results should be interpreted with caution and more research in this area is required due to their study's small sample size as well as the fact that the most recent study included in the analysis was published more than a decade ago in 2007-- the year the first iPhone entered the market and eliminated the distinction between telephone and desktop computers as a means of conducting remote job interviews (Blacksmith et al., 2016). Thus, given the small bases of studies in this review, the significant technological progress since 2007, and the need to refine our understanding of equivalence, a new narrative review is needed to provide a sense of the current state of research on technology in employment interviews. Moreover, a narrative review can link popular media theories with evidence from the selection domain and synthesize the existing literature on technology in employment interviews in order to provide a comprehensive and up-to-date summary of the

phenomenon. In addition, extensions to existing theory should be suggested in order to transfer Potosky's (2008) modular approach to the specific case of employment interviews.

In the process of drafting the review section, a number of steps were taken to identify relevant research studies on employment interviews and technology. This included Boolean keyword searches (i.e., interview AND technology, employment interview AND video) of the PsycINFO, ProQuest ABI/INORM, ScienceDirect and Google Scholar databases. In addition, the reference sections of all reviews and meta-analyses after the year 2000 were screened for relevant articles. In addition, manual article-by-article searches were conducted in the following journals: *Journal of Applied Psychology*, *Personnel Psychology*, *International Journal of Selection and Assessment*, *Journal of Management*, *Journal of Work and Occupational Psychology*, and *Computers in Human Behavior*. Nevertheless, given the broad scope of the fields of technology, employment interviews, and technology in selection in general, including unproctored Internet testing (UIT), this qualitative review will be integrative but also selective.

1.2. Definition of an interview and interview components

1.2.1.1. Definition

In the literature, the definition of an employment interview was long considered to be limited to a direct, synchronous face-to-face interaction or at least synchronous interactions via telephone and later videoconferencing systems (Huffcutt & Youngcourt, 2007). Levashina et al. (2014) defined an employment interview as "a personally interactive process of one or more people asking questions orally to another person and evaluating the answers for the purpose of determining the qualifications of that person in order to make employment decisions" (p. 243). Further, "[…] the modality of the interview can be variable, so long as there is still interpersonal interaction and communication between the interviewer and interviewee. Moreover, the interpersonal interaction might include immediate (or synchronous) and delayed (or asynchronous) interaction" (p. 244). Importantly, this definition

allows the term *interview* to be used for technologies that record answers to interview questions without an interviewer and interviewee engaging in a direct personal interaction as long as one or more interviewers provide interview questions as a stimulus and receive answers for the purpose of a selection decision. Nevertheless, Levashina et al. (2014) excluded written interviews from their definition, although other researchers claim that providing written answers to interview question should also be considered a form of structured employment interview (Whetzel, Baranowski, Petro, Curtin, & Fisher, 2003). In the modular perspective on employment interviews taken in this review, written responses to tailored job specific interview questions like the format described by Whetzel et al. (2003) will be considered employment interviews, as the media format of the response can take several forms, including face-to-face, video, telephone, audio-visual or written, as might be the case on instant messaging applications.

It is important to note that interviews vary in whether their primary purpose is as a selection instrument or a recruiting device, which significantly affects the way they are conducted (Harris, 1989). The primary goal of the interview as selection device is to identify actually suitable candidates out of a pool of potentially suitable candidates. The same purpose applies to screening interviews in a multi-round selection process, which have the goal of reducing the number of potential options in a first step (negative selection) for the sake of better-decision making (Beach, 1993). In its recruitment function, the goal of the interview is to persuade candidates to see the hiring organization as an attractive choice in their job search process. The selection and recruiting functions are in partial conflict when the interviewer's goal of attracting applicants negatively impacts the predictive validity of his or her interview judgments, which was recently demonstrated in a laboratory study (Marr & Cable, 2014). The use of phone or video interviews is quite prevalent in early-stage screening interviews due to its increasing cost-benefit ratio in cases with a high volume of candidates and a low selection ratio.

1.2.1.2. ***Components of interview structure***

Although multiple meta-analyses have found structured interviews to be more valid than unstructured ones (e.g., Huffcutt et al., 2014; McDaniel et al., 1994; Schmidt & Zimmerman, 2004), there is still no consensus on what a structured interview means exactly. In the literature, structure is conceptualized in several different ways. Campion, Palmer, and Campion (1997) summarized in their review that the empirical evidence unanimously finds that structure increases the psychometric properties of employment interviews. Campion et al. (1997) define structure very broadly as "any enhancement of the interview that is intended to increase psychometric properties by increasing standardization or otherwise assisting the interviewer in determining what questions to ask or how to evaluate responses" (p. 676). Campion and colleagues proposed a typology of 15 different components of structure and clustered these dimensions into broader categories of components that either affect *interview content* or enhance the *standardization of evaluation*. Table 1.2 provides Campion et al.'s 15 components of interview structure plus three additional components suggested later by Levashina et al. (2014).

Components in the category of interview content include conducting a job analysis, asking each candidate the same set of questions, limiting prompting, using sophisticated questions like situational or past behavior questions, longer interviews with more questions (because a higher number of questions increases reliability when random errors are uncorrelated), controlling interviewers' access to ancillary information like resumes or cognitive test scores, and limiting candidates' questions to the end of the interview.

Table 1.2. *Overview of interview components (Campion et al., 1997; Levashina et al., 2014)*

Nr.	Components	Example / Elaboration
	Interview content (Campion et al., 2014)	
1	Job analysis	Using critical incident analysis to derive questions
2	Same questions to each candidate	Provide interview guides to all interviewers
3	Limit promting	Explanation: Prompting is seen as potential form of bias
4	Better questions	Use of situational or past behavior questions
5	Longer interview	More questions reduce uncorrelated error
6	Control ancillary information	Ancillary information includes forms, resumes, test scores, etc.
7	No questions by applicants	Candidate questions are not allowed during the interview
	Interview evaluation (Campion et al., 2014)	
8	Rate each question	Rate each answer or use multiple scales
9	Anchored rating scales	To use detailed anchored rating scales
10	Detailed note-taking	Explanation: Note taking is considered to reduce memory decay
11	Use multiple interviewers	Panel or board interviews are assumed to increase reliability
12	Same interviewer(s)	Same interviewer(s) conducts the interviews for all candidates
13	No discussions between interviews	Do not discuss candidates or answers between interviews
14	Interviewer training	Providing frame of reference training
15	Use statistical prediction	Averaging across dimensional ratings
	Added by Levashina et al. (2014)	
16	Limit rapport building	Explanation: Small talk is seen a potential source of bias
17	Recording of interviews	To record the interviews (audio, video, or transcripts)
18	Interview transparency	Provide information to interviewees about the constructs or goals

Components that enhance the standardization of evaluation include rating each answer or using multiple scales, using anchored rating scales, note-taking during the interview, involving multiple interviewers and having the same interviewers interview different candidates, disallowing discussions between interviewers, providing extensive interviewer training, and statistically rather than clinically aggregating[4] the interview results. Levashina et al. (2014) asked 30 subject matter experts in the selection domain to rate the importance of Campion et al.'s (1997) 15 dimensions in the two domains of content and evaluation standardization on a five-point scale. They further included three additional dimensions based

---

[4] This last point refers to the use of mean scores from dimensional ratings or any other form of algorithm. Statistical decision-making has been found to be more valid than clinical decision-making in selection (Kuncel, Klieger, Connelly, & Ones, 2013)

on a narrative review, namely (a) rapport building (e.g. small talk), (b) recording of interviews, and (c) interview transparency. Limiting rapport building is suggested to improve the interview's psychometric properties, as small talk is seen as a potential source of bias. On the other hand, completely neglecting rapport building has been suggested to negatively impact candidates' affective reactions. Interview transparency refers to the degree candidates are informed about what the interview is intended to measure. The lowest level of transparency involves providing applicants with no information at all about the interview process, while the highest level of transparency involves providing candidates with a list of all questions that will be asked in the interview and all criteria used to evaluate answers beforehand. Transparency has been found to increase construct validity, reliability and candidates' fairness perceptions. Hence, a higher level of transparency should not decrease criterion-related validity. Recording interviews was also introduced as a potential new component of structure and was conceptualized along three levels depending on the cue richness of the recording, i.e. whether the recording captures video and voice, voice only, or no recording is made (Levashina et al., 2014, p. 251). Interestingly, the recording dimension is the only one in the entire list that depends on some form of technology, while all other dimensions are free of technology dependency. Thus, it is later suggested later that the reprocessability achieved through recording be considered as a media attribute rather than a structural component of employment interviews.

In addition to Campion et al. (1997), there is at least one other framework on structure that has received considerable attention in employment interview research. Chapman and Zweig (2005) developed a nomological network for interview structure encompassing four dimensions, namely (a) *question consistency*, (b) *evaluation standardization*, (c) *question sophistication*, and (d) *rapport building*. In a survey design with two independent samples (candidates and interviewers) who completed face-to-face on-campus interviews, a factor analytic procedure confirmed the proposed four-factor structure, although the reliabilities for

the two dimensions of question sophistication and rapport building were low. Chapman and Zweig (2005) further found that the level of structure interviewers apply depends on training (trained interviewers conduct more structured interviews) and interview focus (selection interviews are more structured than recruiting interviews). Overall, it was proposed that all four dimensions relate positively to interview reliability and validity, but negatively to applicant reactions. Empirically, there was no linear relationship between interview structure and candidate reactions, and none of Campion et al.'s 15 components were positively related to candidates' perception (Chapman & Zweig, 2005).

Turning to the relationship between interview structure and technology usage, a content analysis conducted by Levashina et al. (2014) found more components of structure in face-to-face interviews than in telephone interviews or video interviews. Hence, Levanisha and colleagues proposed that technology-enhanced interview formats benefit from the use of more structural components, with structure compensating for limitations due to lower media richness in technology-enhanced job interviews. The introduction of interview recordings as a technology-dependent additional component of interview structure leads to conceptual fuzziness in the discussion of the relationship between interview structure and technology-enhanced interviews. A more detailed conceptual differentiation of effects due to technology vs. interview structure is needed, as technology might influence interview structure in a unidirectional way, technology and structure may be independent, or interview structure and technology might interact in affecting interview outcomes. This conceptual differentiation will be provided later on in this thesis.

1.2.2. Employment interviews via telephone

In the literature, telephone interviews are defined as the direct (synchronous) exchange of information between two or more persons as a precursor to a selection decision using a telephone only, without the simultaneous use of video or other visual media. Conducting interviews via telephone is more common in early stages of selection than in later stages.

Telephone interviews differ from interactive voice response[5], in which candidates answer standardized questions in a closed-response format (Bauer, Truxillo, Paronto, Weekley, & Campion, 2004; Tourangeau, Steiger, & Wilson, 2002). They also differ from reference checks collected by telephone in which the actual applicant is not involved in the information exchange process (Taylor, Pajo, Cheung, & Stringfield, 2004). Telephone interviews are considered to increase hiring managers and candidates' flexibility and save travel time and travel expenses.

Several peer-reviewed research articles have addressed telephone interviews. A UK field study on campus recruiting interviews by Silvester et al. (2000) compared post-interview ratings between telephone and face-to-face interviews for 70 real candidates in a within-subject design. The study revealed that candidates received significantly lower ratings for telephone interviews compared to face-to-face interviews. Only the group that had telephone interviews first and face-to-face interviews second experienced improvement from the first interview to the second. Silvester and Anderson (2003) suggested that candidates should mention more personal causal attributes to reduce anonymity in telephone interviews, as they found that making more personal attributions was correlated with higher interview ratings in telephone interviews, but not face-to-face interviews. Moreover, interviewers asked more closed questions (non-significant trend) in telephone interviews than face-to-face, indicating a difference in communication styles due to the media format.

A lab study with undergraduate students by Blackman (2002) compared the accuracy of telephone and face-to-face interview formats in simulated job interviews. The study revealed that the accuracy of personality ratings assessed by means of self-interviewer and peer-interviewer agreement was significantly higher in face-to-face than in telephone interviews. There was also a main gender effect indicating higher agreement for female dyads. An item-

---

[5] Interactive Voice Response (IVR) is a technology that asks applicants to use a telephone for a self-administrated screening, e.g., by pressing "1" for yes and "0" for no to standardized questions (e.g., "Do you have a truck driver's license?")

level analysis further indicated that candidates were experienced as significantly more cheerful, warm, considerate, and social in the face-to-face condition - traits that are conveyed via nonverbal communication. Rating accuracy was higher for these socially desirable traits in the face-to-face condition as well. Moreover, the number of follow-up questions, which was correlated with rating accuracy, was higher in the face-to-face condition.

Straus et al. (2001) compared face-to-face, videoconference, and telephone interviews in a lab study with a hypothetical setting and a sample of 60 first-year students in a business master's degree program. They found that ratings on general abilities were higher in telephone interviews compared to both face-to-face interviews and videoconference interviews. Candidates were also rated as more likable in telephone interviews compared to face-to-face. Finally, the authors reported an interaction between the medium and candidates' attractiveness, with less attractive candidates receiving lower ratings and being seen as less likeable in face-to-face interviews than in telephone interviews. Straus and colleagues suggested that negativity bias may explain this effect: As negative visual cues are overweighted, the suppression of such negative visual cues in less rich media environments results in more favorable evaluations in telephone interviews than in interviews incorporating visual media.

With respect to applicant reactions, Chapman, Uggerslev, and Webster (2003) investigated differences in candidates' perceptions of face-to-face, videoconference, and telephone interviews in a campus field sample. The results of this study revealed that face-to-face interviews were perceived as significantly fairer than telephone interviews fairer than videoconference interviews. The candidates' perceived interview outcome was significantly lower for videoconference compared to telephone and face-to-face interviews. Hence, face-to-face interviews were perceived as more favorable than videoconference interviews, but not telephone interviews. Job acceptance intentions were higher for face-to-face interviews compared to telephone interviews (Chapman et al., 2003).

Finally, Bauer et al. (2004) conducted a lab study with a student sample on applicant reactions towards face-to-face interviews, telephone interviews, and interactive voice response with the same content. Using the Selection Procedural Justice Scale (SPJS) by Bauer et al. (2001) as the dependent variable, the data revealed that face-to-face interviews were rated significantly higher on the SPJS dimensions of reconsideration opportunity, openness, interpersonal treatment, and two-way interaction. No significant differences were found for organizational attractiveness. There was also no significant difference in the reactions of top candidates scoring high on cognitive abilities and conscientiousness, which indicates that self-selection bias did not occur in this sample.

In summary, primary studies concerning main effects of media format on interview ratings in studies comparing face-to-face with telephone interviews have delivered mixed results. Two studies found lower ratings for telephone compared to face-to-face interviews, while one study found effects in the other direction (Blackman, 2002; Silvester et al., 2000; Straus et al., 2001).

With respect to applicant reactions, the only study to systematically assess this phenomenon found a main effect for media format such that applicants perceive telephone interviews as less favorable than face-to-face interviews (Bauer et al., 2004). However, no existing studies assess whether there are significant differences in actual job acceptance as opposed to short-term reactions depending on the interview format used.

1.2.3. Synchronous Video Interviews

The speed and bandwidth of Internet connections in most parts of the world have strongly increased in recent decades, and high-resolution webcams on personal computers or mobile devices have become affordable to a large share of the global population. Consequently, the administration of employment interviews via consumer videoconferencing software such as Microsoft Skype, Google Hangouts or Apple FaceTime has become increasingly popular. When scholars speak of video interviews, they typically refer to the

synchronous sharing of video and audio information between interviewer and interviewee with the ability to immediate react to one another.

A first exploratory field study on campus recruiting by Kroeck and Magnusen (1997) found that both employers and candidates preferred traditional face-to-face interviews compared to synchronous video interviews. The previously introduced laboratory study by Straus et al. (2001) did not find significant differences in ratings between synchronous video interviews and FtF interviews. However, candidates were rated as less likable in synchronous video interviews than face-to-face. Another study in the field of campus recruiting found that candidates randomly assigned to synchronous video interviews received significantly better ratings than those randomly assigned to face-to-face interviews (Chapman & Rowe, 2001). There was also a main effect for structure such that ratings were lower for structured interviews compared to semi-structured or unstructured interviews. A lab study on media influences on attribution in employment interviews suggested that less rich media lead to more external attributions and therefore increased ratings (Chapman & Webster, 2001). For example, when candidates provide a poor answer in less rich media environments, this is more likely to attributed to the difficult situation resulting from the intermediating medium. Moreover, Chapman and Rowe (2002) found an interaction between interview medium and structure on organizational attractiveness, with videoconference and structured interviews leading to more favorable applicant views of the organization and synchronous video interviews and unstructured interviews lead to less favorable applicant views.

The 2003 study by Chapman and colleagues found that face-to-face interviews were perceived as fairer than synchronous video interviews, but not fairer than telephone interviews. The perceived interview outcome for synchronous video interviews was significantly lower than for face-to-face interviews and (non-significantly) lower than for telephone interviews. Face-to-face interviews were also perceived as more favorable than synchronous video interviews (Chapman et al., 2003).

Sears et al. (2013) conducted a lab study with MBA students in which they found that candidates perceived less procedural justice in synchronous video interviews compared to face-to-face interviews. Moreover, interviewers were rated as less personable, trustworthy, and competent by interviewees in the video interview condition. Conversely, interviewers evaluated the hypothetical candidates as less competent (lower overall interview performance), less likable, and were less likely to recommend hiring the person in synchronous video interviews. Finally, a between-subject comparison between synchronous and asynchronous video interviews in a hypothetical setting revealed that synchronous video interviews are viewed as less creepy, higher in two-way communication and interpersonal treatment, and lower in privacy concerns compared to their asynchronous counterparts (Langer et al., 2017). In contrast to the authors' hypothesis, candidate ratings were higher for asynchronous video interviews compared to ratings from synchronous video interviews or ratings based on Skype interview recordings (Langer et al., 2017).

In summary, studies comparing synchronous video interviews to face-to-face formats reveal a substantial trend in the direction of lower ratings and applicant reactions. In addition, Langer et al. (2017) provide the first direct comparison of synchronous and asynchronous video interviews, finding significant differences between these two video interview formats. One limitation of the generalizability of the results of older studies is the question of whether the effects of the videoconferencing systems available on the market at those points are still comparable to today's easy-to-use applications with optimized interfaces[6].

1.2.4. Asynchronous video interviews

In asynchronous video interviews, candidates are invited to record themselves answering standardized interview questions within a certain time frame in the absence of the parallel presence of any direct interaction with an interviewer or proctor. Interview questions

---

[6] The first consumer video chat platform Skype was launched in 2003. Older studies were conducted with professional systems such as CU-SeeMe (CU-SeeMe Development Team, 1995).

(stimulus material) can be received in the form of text, audio, or video via an online platform. Interviewers can watch and rate individuals interview sequences at any time and from any place once they have been recorded by the interviewees (Brenner et al., 2016).

Asynchronous video interviews are commonly conducted via web-based software platforms developed by specialist boutique firms whose software solutions vary in functionality and are modular in terms of the available features. Today, numerous commercial providers of platforms and software for asynchronous video interviews exist, mainly specialist boutique firms (e.g., Hirevue.com, Launchpadrecruits.com, Sonru.com, viasto.com), but also test publishers who have elected to complement their test offerings with asynchronous video interview platforms (cut-e, 2019) as well as providers of applicant tracking or talent management software (Cornerstone OnDemand Ltd., 2019). A web search conducted by Langer et al. (2017) revealed more than 70 companies offering such software solutions; the same web search in 2018 uncovered more than 80 results (Software Advice, 2018). The most prominent providers include HireVue in the US market and viasto in the German-speaking countries, with customers of the latter coming from a wide range of industries including public-sector organizations (viasto GmbH, 2019). The terms 'digital interview' (Chamorro-Premuzic et al., 2016; Langer et al., 2017), 'prerecorded interview' and 'one-way interview' (Blacksmith & Poeppelman, 2014) have been used interchangeably to describe the same principle. The main difference between such systems and synchronous video interviews is the *absence of any immediate, temporally synchronized communication between interviewers and candidates;* all other components of employment interviews can be introduced into the asynchronous interview process. Nevertheless, the software applications used for asynchronous interviews vary with respect to numerous features and structural components. For example, some vendors offer video formats such as pitch interviews for the first step of the recruitment process, which bear a stronger similarity to video resumes, while others offer more tailored processes for screening candidates in the second step of the selection process

(Cammio, 2019). There are also differences in the extent to which the evaluation is standardized, ranging from dichotomous overall ratings (pass – no pass) to behaviorally anchored rating scales. Some vendors also combine the recording of interview answers with coding tests or SJTs (e.g., HireVue, 2019). Just as in other areas, the technology has already advanced so far that some vendors of video interview platforms (e.g., HireVue) apply advanced analytic technologies like voice profiling to create automated rankings with no human involvement. Thus, how asynchronous video interviews are applied in concrete situations can vary drastically, supporting Lievens and Sackett's (2017) recommendation that assessment procedures such as (video) interviews be viewed as modular systems.

Similarly, research on asynchronous video interviews as a holistic procedure remains sparse, but some empirical studies do exist: A first experimental laboratory study with non-applicants on potential adverse impacts of asynchronous video interviews found no evidence for discrimination in asynchronous video interview ratings based on gender or minority status (Kroll & Ziegler, 2016). An initial study exploring applicant reactions revealed that the overall favorability ratings of asynchronous video interviews are mostly comparable to those of cognitive tests (Brenner et al., 2016). As previously mentioned, Langer et al. (2017) provided the first direct comparison of applicant reactions to synchronous and asynchronous video interviews, finding that (hypothetical) candidates did prefer synchronous video interviews, but synchronous video interviews produced lower interview ratings than asynchronous video interviews. A first investigation of construct validity using the taxonomy by Huffcutt, Conway, Roth, and Stone (2001) found that ratings from asynchronous video interviews on the construct dimensions of mental ability, knowledge and skills, applied social skills, and the personality trait of conscientiousness were correlated with self-rated job performance (Gorman et al., 2018).
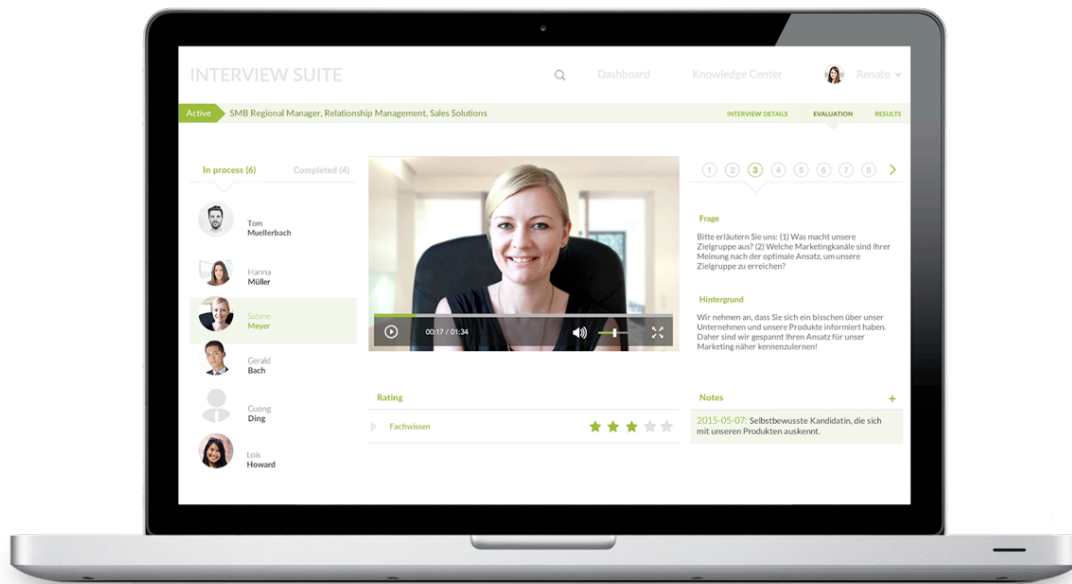
Figure 1.1. Asynchronous video interview platform (www.viasto.com)

1.2.5. Video resumes and video-based situational judgment tests

As previously noted when addressing the question of whether written interviews can be considered interviews, it is important to discuss the distinction between interviews and other pre-hiring assessments that share some commonalities with them. The two most prominent such assessments are video resumes and video-based or multimedia SJTs. Video resumes are short, pre-recorded and edited video clips of applicants which are submitted to potential employers instead of traditional text-based resumes (Hiemstra, Derous, Serlie, & Born, 2012). The use of video resumes is argued to be beneficial to candidates as a way of demonstrating their interpersonal skills, abilities, or other job-related characteristics to employers and a more personalized way of applying for a given position (Blacksmith & Poeppelman, 2014). Video resumes differ from video interviews because an interviewer's presentation of tailored, job-related interview questions is considered an essential characteristic of the communication process in interviews according to Levashina et al.'s (2014) definition. In video-based SJTs, video is used as stimulus material rather than as the response mode; instead, candidates'

responses are collected in standardized response formats (Lievens & Sackett, 2006). For example, the webcam test by Oostrom, Born, Serlie, and van der Molen (2010) presents a number of work-related situations. Candidates are asked to respond to this stimulus material as if they were real work situations and record their responses using a webcam. Such tests also differ from interviews due to the absence of a communication process characterized by tailored interview questions presented by an interviewer. Nevertheless, as already mentioned in the paragraph above describing asynchronous video interviews, these procedures share several constitutive elements with asynchronous video interviews, with the differences only becoming clear when they are viewed from a holistic perspective.

1.2.6. Future directions in technology usage in employment interviews

Technological progress is currently accelerating rather than slowing down, and innovation cycles are constantly becoming shorter. Thus, several further developments to employment interviews can be expected in the foreseeable future. This section briefly presently a few selected technological trends that are already being controversially discussed among academics and practitioners or for which first applications are already on the market. *Chatbots*: Chatbots are applications that use artificial intelligence to communicate with a human user. Chatbots are already quite common in customer service, but first applications in human resources also exist, e.g. as a means of guiding potential candidates to relevant job postings on a career site or chatbots that can answer frequently asked questions (FAQs) about the hiring process (Shaw, 2012). There are also applications that can schedule interviews with candidates without any human needing to be involved (e.g. https://x.ai/). Chatbots will become relevant for the interview domain as soon applications that interact with candidates in the early stages of the selection process hit the market. For example, such applications might ask candidates about their formal qualification or vocational interests.
Voice profiling: A person's word usage is related to both their personality and their actual behavior (Fast & Funder, 2008). Thus, personality profiles of candidates can quickly and

easily be calculated via automated quantitative analyses of text drawn from recorded interviews and speech-to-text applications. Initial applications for this purpose like Precire (Greb & Linnenbürger, 2013) are already on the market. Although empirical results on the psychometric properties and criterion validity of such applications are scarce, the technology has been discussed controversially in popular media, with a strong focus on applicant reactions and potential adverse impact[7]. From a research perspective, the most important question should be whether such technology possesses criterion or incremental validity or is merely a more efficient way of collecting the same information traditionally offered by personality inventories (which are used to train the technology), which have been found to be only marginally predictive of actual job performance (Morgeson et al., 2007). Nevertheless, peer-reviewed research on the use of voice profiling in digital employment interviews and its psychometric properties do not yet exist to the author's best knowledge.

Nonverbal signals: Research in the domain of "thin slices" has shown that both people and automated cue detection applications can make valid judgments about a stranger's states or traits based on a behavioral sample of just a few seconds (Ambady & Rosenthal, 1992; Batrinca, Mana, Lepri, Pianesi, & Sebe, 2011; Carney, Colvin, & Hall, 2007). Previous research has found automated cue detection to be predictive of hireability impressions (Nguyen & Gatica-Perez, 2015), and commercial providers of video interview platforms have already implemented solutions comparing a range of signals to an organization's training sample of high performers (www.hirevue.com). Finally, applying nonverbal cues for automated lie detection in selection on the basis on micro-impression databases has been discussed (Chamorro-Premuzic et al., 2016).

Avatars and Virtual Reality: Avatars are digital figures that represent particular persons in a virtual environment. The use of avatars for digital interview training was recently tested by

---

[7] E.g., http://www.spiegel.de/karriere/recruiting-unternehmen-setzen-verstaerkt-auf-algorithmen-a-1132633.html

Langer, Konig, Gebhard, and Andre (2016). In their study, a computer analyzed participants'
nonverbal behavior and provided real-time feedback on automatically detected non-verbal
cues. Participants who received the digital interview training achieved significantly better
scores in a mock interview as a criterion measure compared to participants from a control
group (Langer et al., 2016). Another laboratory study testing the effects of avatar appearance
on interviewer ratings in virtual employment interviews found that attractiveness stereotypes
were also applied to virtual interview characters (Behrend, Toaddy, Thompson, & Sharek,
2012). More research is needed when such applications enter the market.

   Finally, to conclude this chapter, I would like to provide an outlook on technologies that
will very likely find first applications in the selection and interview context in the coming
years. For this purpose, the Gartner Group consultancy provides a useful framework for
classifying new technology according to its maturity of productivity (Fenn & Raskino, 2008).
Gartner's Hype Cycle describes the emergence of new technology over five phases: an
innovation trigger, a peak of inflated expectations, a trough of disillusionment, a slope of
enlightenment and a plateau of productivity. Applying Gartner's Hype Cycle to the
recruitment and interview domain suggests that new technologies such as machine learning
for automatically scoring digital interviews or virtual assistants and avatars will reach their
plateau of productivity in recruiting in the next two to ten years.
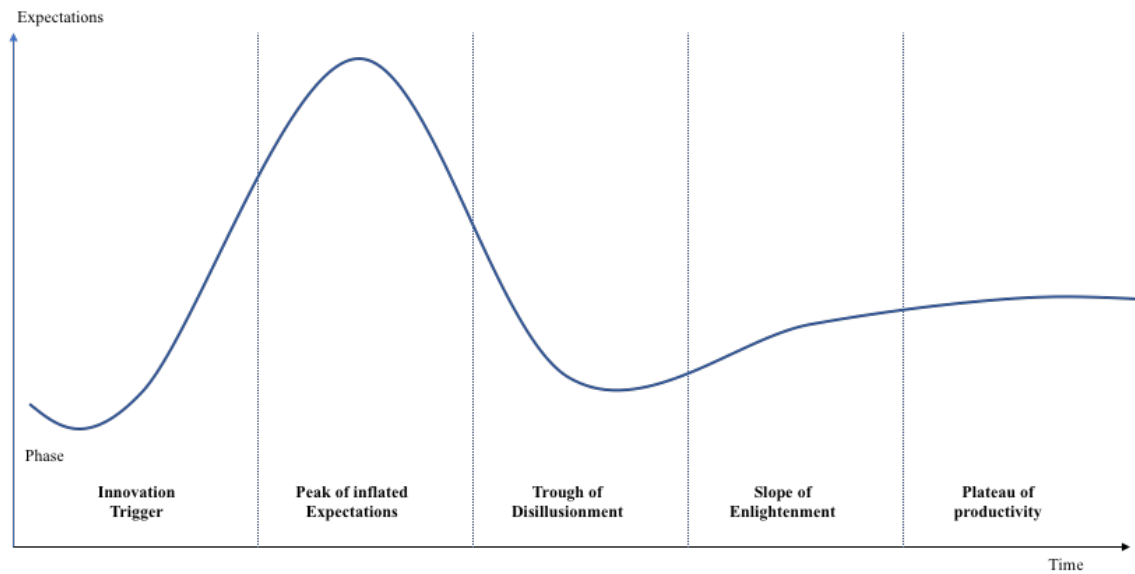
Figure 1.2 . The five phases of Gartner's Hype Cycle (Fenn & Raskino, 2008)

1.2.7. Summary

The usage of various technology-enhanced employment interview formats continues to

expand globally. In contrast to this deep practical relevance, our knowledge of such formats is

limited to only a small number of studies examining the equivalence of interviews using two

or three different media formats. Research on newer interview modes such as asynchronous

video interviews is even more scarce. This existing body of research is summarized

comprehensively in Table 1.3. As depicted in the table, the results of the few existing studies

on technology usage in employment interviews sometimes conflict with respect to the

direction of the observed effects. Furthermore, interpretations of how to explain these

conflicting results vary from no explanation at all to references to various subdisciplines of

psychological research, betraying a serious lack in theory development in the field. It is true

that a first meta-analysis summarizing the few existing studies before 2007 indicated that both

interview scores and applicants' reactions are lower for technology-enhanced compared to

face-to-face interviews (Blacksmith et al., 2016). However, the partially conflicting results of

the primary studies, the methodological limitations of the only existing meta-analysis on the

issue, and the lack of theory development provide at least three good rationales for taking a closer look at media theory and the empirical studies presented above.

This review has focused so far on comparing interview scores or applicant reactions between different procedures considered holistically in order to build on the existing literature. Nevertheless, this comparison is problematic because such a holistic view implies that different interview formats could be interchangeable, e.g. that organizations might think about replacing telephone interviews wholescale with asynchronous video interviews for screening. In contrast, many new software solutions have modular decisions, and the probability that face-to-face interviews will be replaced by asynchronous video interviews as a final selection step is very low, as the latter are optimized for preselection. Thus, as also noted by Morelli et al. (2017), equivalence comparisons in the absence of conceptual frameworks regarding the role of technology in selection fail to answer the appropriate questions. Moreover, theory is essential for predicting how interviewers' and interviewees' psychological processes interact with various technologies, which is in the end what ultimately impacts selection outcomes, including psychometric properties.

Table 1.3. *Overview of research on technology-enhanced employment interviews.*

Author(s)	Research topic	Technology	N	Dependent variables	Key finding
Kroeck & Magnusen (1997)	Candidate & recruiter reactions to SVI vs. FtF interviews	FtF vs. SVI	159	Candidate reactions Interview ratings	General preference for FtF versus SVI interviews, no significant differences on interview performance ratings
Silvester et al. (2000)	Measure equivalance between FtF and TI	FtF vs. TI	70	Interview ratings	Significant lower ratings for TI vs. FtF interviews
Chapman & Rowe (2001)	Measure equivalance between FtF and SVI	FtF vs. SVI	92	Interview ratings	Significant higher ratings for SVI vs. FtF interview
Chapman & Webster (2001)	Process of rating inflation in SVI vs. FtF interviews	FtF vs. SVI	92	Interview ratings	Significant higher ratings for SVI vs. FtF interviews due to interviewers' external attribution in interview formats with low media richness
Straus et al. (2001)	Measure equivalance between FtF, SVI and TI	FtF vs. SVI vs. TI	60	Interview ratings	Significant higher ratings in TI vs. FtF interviews, no significant differences between FtF vs. SVI; moderation of candidates' physical attractiveness on the magnitude of mean difference between TI vs. FtF interviews
Blackman (2002)	Measure equivalance between FtF vs. TI	FtF vs. TI	120	Personality judgments (accuracy)	Significant lower average self-interviewer and peer-interviewer agreement correlations for TI vs. FtF interviews
Chapman & Rowe (2002)	Equivalence of applicants' perception between FtF vs. SVI and the role of interview structure	FtF vs. SVI	92	organizational attractiveness interviewer friendliness interviewer performance self-ratings	Interaction between interview format and interview structure on organizational attractiveness (FtF > SVI in low structured interviews vs. SVI > FtF in high structured interviews); Significant main effect for interviewer performance (FtF > SVI); Significant main effect for self-ratings (FtF>SVI) moderated by structure (FtF > SVI in low structured interviews vs. SVI > FtF in high structured interviews)

Chapman et al. (2003)	Examination of differences in applicant reactions between FtF vs. SVI vs. TI among real applicants	FtF vs. SVI vs. TI	802	Perceived fairness Expectation of favorable outcome Job acceptance intention Self-monitoring	Higher fairness perception and higher job acceptance intention for FtF vs. SVI or TI. Higher perceived interview outcome perception for FtF & TI > SVI. Moderation of self-monitoring on the relationship between interview medium and perception of fairness
Silvester & Anderson (2003)	Investigation on interviewers' questioning strategies and applicants' causal attribution in FtF vs. TI	FtF vs. TI	62	No. of personal vs. global causal attribution and their relationship to interview performance	Greater number of personal causal attributions of TI vs. FtF interviews. Correlation between number of personal causal attributions and interview ratings in TI, and negative correlation between global causes attributions and interview ratings in FtF interviews
Bauer et al. (2004)	Comparative investigation on the differences in applicant reactions between FtF, TI and IVR	FtF vs. TI vs. IVR	153	Procedural justice Conscientiousness Cognitive abilities	Significant main effects of interview format on procedural justice in the dimensions reconsideration opportunity, openness, interpersonal treatment, and two-way communication (FtF > IVR & TI); no significant interaction of conscientiousness or cognitive ability
Van Iddekinge et al. (2006)	Comparative investigation on measurement equivalence between FtF and videotapes of the same interviews	FtF vs. videotapes of the same interviews	113	Interview ratings	Significant higher ratings for FtF vs. videotapes of the same interviews. Smaller correlation between FtF and videotaped interviews compared to interviewer ratings in panel interviews; higher reliability of videotaped versus FtF interviews
Sears et al. (2013)	Comparative investigation on differences between FtF vs. SVI	FtF vs. SVI	52	Procedural justice perceived interviewer characteristics Interviewer evaluation Affect toward applicant and perceived competence Overall hiring recommendation	Significant lower perception of chance to perform, availability of selection information, job-relatedness of SVI vs. FtF. Significant lower ratings of interviewer characteristics in SVI vs. FtF interviews. Significant lower interview ratings in SVI vs. FtF interviews.

Tiller et al. (2013)	Investigation on measurement equivalance between FtF vs. SVI	FtF vs. SVI	864	Interview ratings	No significant mean difference between FtF vs. SV
Brenner et al. (2016)	Investigation on technology acceptance antecedents on applicant reations in AVI	AVI	106	Procedural justice Process favorability	Perceived usefulness and perceived ease of use significantly predict applicant reactions in AVI
Kroll & Ziegler (2016)	Investigation on the suspectability of discrimination in AVI due to gender or ethnicity	AVI	208	Interview ratings	No evidence for discimination due to gender or ethnicity
Horn & Behrend (2017)	Investigation on the impact of picture-in-picture on interview performance	SVI	113	Applicant reactions Interview performance Self-awareness	No significant effect of picture-in-picture on interview performance
Langer et al. (2017)	Comparative investigation on differences between SVI vs. AVI.	SVI vs. AVI	113	Applicant reactions Interview performance Creepy ambiguity	Significant higher scores of creepiness and privacy concerns, and sigificant lower scores for two-way communication and interpersonal treatment finAVI vs. SVI interview; significant higher interview ratings for AVI vs. SVI
Gorman et al. (2018)	Investigation on the validity of AVI	AVI	75	Job perforamnce (self-rated) Organizational tenure	Significant correlation between interview ratings and job performance as well as organizational tenure

Note. AVI = asynchronous video interviewing, FtF = face-to-face, IVR = interactive voice response, SVI= synchronous video interviewing, TI = telephone interviewing

2.   Theoretical Directions

Advancing theory is essential for better understanding technology-induced effects in employment interviews. Keeping up with the pace of technological advances is challenging: Innovation cycles in the recruitment and assessment domain have become so quick since the start of the Internet revolution that technology use among practitioners is often ahead of scientific advances in the field (Bartram, 2000). This is especially true with respect to the publication of peer-reviewed studies concerning the (alleged) equivalence between a new technology and the traditional administration of the same procedure (Morelli et al., 2017). Highly important for overcoming this challenge are theory and conceptual frameworks that contribute to our understanding of *why* or *how* technology affects psychometric properties and selection outcomes, including interview ratings, performance appraisals, applicant reactions or the economic utility of selection systems as a whole (Morelli et al., 2017). Studies of *measurement equivalence*, which refers to the degree to which two assessment instruments produce comparable results on the intended construct, are not sufficient, as sometimes an allegedly equivalent "traditional" instrument does not exist. For instance, most computer-administrated tests do have any paper-and-pencil counterpart. This might also be true for technologies such as asynchronous video interviews that can be designed in a highly modular manner, as described above, and may or may not contain voice analyses, which may also be critical for outcomes like applicant reactions[8]. Thus, the benefit of theories and conceptual frameworks on technology is that they make it possible to generate hypotheses about specific modules and predict the impact of a new technology in the selection context even when it is not comparable to any "analog" technology antecedent. Thus, the following Chapter presents and discusses common working definitions of technology in the selection context and provides a comprehensive overview of several theories and conceptual frameworks

---

[8] Chamorro-Premuzic et al. (2016) argue that voice analyses are partially the counterparts of intuitive analyses from traditional interviews.

considered helpful for gaining a better understanding of the role of technology in technology-enhanced employment interviews in personnel selection.

It is important to note upfront that theory can differ in its scope of application and coverage. Specifically, media theories have been developed in at least three different directions: (1) theories of media choice, which aim to *explain which technology should be preferred* or used by either recruiters or applicants; (2) theories of *media performance*, which aim to explain why different technologies differ in their outcomes (e.g., differences in accuracy/validity due to error induced by the technology); and (3) systemic theories such as socio-material or socio-technical systems theories, which *embed technology usage and performance in a larger ecosystem* and consider possible interactions between technologies, users and the environment. Before discussing these theories in detail, some definitional issues related to 'technology' and 'medium' first be addressed. Afterwards, these three theoretical directions are reviewed, beginning with *media richness theory* (Daft & Lengel, 1986), proceeding to *media synchronicity theory* (Dennis et al., 2008; Dennis & Valacich, 1999), the conceptional frameworks by Potosky (2008) and Arthur et al. (2018), and the family of theories in the technology acceptance model tradition (Davis, 1989; Davis et al., 1989; Venkatesh & Bala, 2008; Venkatesh & Davis, 2000; Venkatesh, Morris, Davis, & Davis, 2003), and ending with Orlikowski's (2007) socio-material theory.

2.1. A working definition of technology in personnel selection

Due to the often interchangeable use of terms like 'technology', 'media', 'digital media', 'media format' or even 'technology usage' to refer to process automatization, the need for a pragmatic working definition of technology has sometimes been overlooked. In recognition of this, Morelli et al. (2017) suggested defining technology on a very global level as "the constellation of individual tools that assist a user with controlling or adapting to his or her environment" (p. 636). This definition implies that technologies used for employment interviews represent more than just a difference in medium (e.g., video versus telephone), but

also a constellation of tools around the medium like the specific software application (Skype versus Google Hangouts) and the hardware (laptop versus tablet) used to conduct the interview. In the present work, the terms 'technology', 'medium' and 'media format' will be used interchangeably and always in reference to a constellation of tools that mediate the communication process between interviewers and interviewees, with a transformational effect on interview stimulus, stimulus delivery and stimulus presentation, evaluation and scoring, accessibility, and interviewees' and interviewers' reactions.

Next, technology adapts to its users, and communication always involves an interaction between a technology and at least one human being. For example, when a device has low bandwidth due to a poor Internet connection, users can speak more slowly so that their interaction partners can better understand them; in doing so, they adapt to the technology. Hence, the term *technology-enhanced assessments,* which includes technology-enhanced interviews, refers to organizations' use of any of several technologies to help assessors "process more response data or response data more accurately while implementing and interpreting assessments more quickly, affordably, profitably, and accessibly" (Morelli et al., 2017, p. 637). This definition explicitly assumes that technology is used to improve or simplify an aspect of assessment, whether it be time effectiveness, cost efficiency, or the quality of decision-making.

In the media richness (Daft & Lengel, 1986) and media synchronicity theory traditions (Dennis et al., 2008; Dennis & Valacich, 1999), technology is operationalized as a medium that acts as an information carrier for a communication process; different media can be sorted on a continuum ranging from e.g., very rich to less rich (or from high to low synchronicity). Potosky (2008) further develops this communication channel perspective on media in a more nuanced conceptualization. Technology acceptance models (TAM) refer to any hardware or software with which users interact (Davis et al., 1989). Finally, Orlikowski (2007) eliminates

the distinction between human and technological agency in proposing a mutually interacting socio-material system encompassing both users and technologies.

Morelli et al. (2017) presents a perspective on technology in selection that is mainly based on the assumptions of classical test theory (Guilford, 1954; Gulliksen, 1950). From this perspective, technology is considered by default as a potential source of variance that might increase the portion of random error in a construct measurement in line with the following formula: observed score = true score + random error. In this thesis, it is suggested that viewing technology merely as a potential source of score contamination might be too narrow, as technological enrichment might in some cases increase outcomes like candidate reactions by incorporating playful elements[9] or directly affect the predictor-criterion relationship by incorporating unique technology-dependent stimuli in augmented or virtual reality applications that help to establish behavioral consistency vis-à-vis digital workplaces (Jansen et al., 2013).

2.2.   Media Richness Theory

Media richness theory (Daft & Lengel, 1986) is one of the most prominent media theories. The original version of the theory proposes that media choice and task performance involve matching media characteristics to organizational information processing needs. Specifically, richer media formats (e.g. face-to-face communication) are better suited for tasks with higher levels of uncertainty and ambiguity, where multiple interpretations are possible, while leaner media (email, chat) are preferred for tasks with low levels of uncertainty, like information issues (e.g., informing staff that the printer is out of service). The theory further postulates that media differ in their "richness", with face-to-face considered a richer medium than video, telephone, and written documents on the other end of the continuum. Media richness varies on several dimensions, including the ability of a medium to transmit multiple

---

[9] The term 'Recrutainment' has gained traction in practice to refer to this phenomenon.

cues (gestures, smiles), the degree to which feedback is given immediately, language variety, and level of personalization. Despite its popularity, empirical evidence for media richness theory is quite sparse, especially with respect to its explanatory value for media choice. For instance, the frequent and persistent use of email communication for tasks with a high level of uncertainty and ambiguity runs contrary to the theory's predictions (Dennis et al., 2008). In its pure version, media richness theory suggests that face-to-face interviews would always be more effective for conducting employment interviews, which represents highly complex tasks with high levels of uncertainty and ambiguity. However, despite the lack of empirical evidence underlying it, media richness theory has proved very fruitful in inspiring subsequent conceptualizations.

2.3.   Media Synchronicity Theory

Due to the lack of empirical evidence for media richness theory, media synchronicity was introduced as a refinement of media richness theory with the goal of overcoming the former's limitations (Dennis et al., 2008; Dennis & Valacich, 1999). Media synchronicity theory (MST) identifies *physical media capabilities* that "induce the creation of a socially developed characteristic that we term media synchronicity, which may differ from person to person and over time" (Dennis et al., 2008, p. 579). Media synchronicity theory aims further to explain communication performance in specific tasks instead of media choice, with tasks seen as underlying communication processes. In MST, two communication processes are relevant for every task: conveyance and convergence. Conveyance processes are defined as the transmission of information to enable receivers to create a mental model of the situation. Convergence processes are defined as the preprocesses required to interpret information and gain a mutual understanding. The fit between conveyance and convergence on the one hand and the information transmission and processing capabilities of the media on the other hand predicts communication performance. When people work together synchronously, their actions unfold at the exact same rate. High synchronicity is associated with lower cognitive

demands required to encode and decode messages. In media synchronicity theory, media capabilities are defined as "the potential structures provided by a medium which influence the manner in which individuals can transmit and process information" (Dennis et al., 2008, p. 583), which ultimately determine synchronicity capability along the following dimensions:

Transmission Velocity: This first dimension is defined as the speed at which a medium can transmit information to its intended recipients. It is the same concept as immediacy of interactivity feedback in media richness theory.

Parallelism: Parallelism refers to the number of simultaneous signal transmissions from different senders that can effectively take place at one time. Multiple simultaneous transmissions can reduce some of the losses that occur when messages must be transmitted sequentially. However, when different discussions take place at the same time, parallelism can reduce interactional coherence and shared focus and may therefore be a threat to synchronicity (Dennis et al., 2008, p. 585).

Symbol Sets: Symbol sets refer to the number of cues (physical, visual, and verbal symbol sets) that can be transmitted via a medium. The concept is similar to Daft and Lengel's (1986) notion of cue multiplicity in media richness theory. Different symbols can have different encoding and decoding costs. For example, it takes more time to type a message compared to speaking and more time to listen to spoken language compared to reading a text, but it may be easier than looking at an informational chart. When symbol sets are removed, there is usually a reduction in social presence (Short et al., 1976).

Rehearsability: According to Dennis et al. (2008), rehearsability refers to the extent to which a medium enables its users to refine or fine-tune a message before transmission and helps senders carefully generate the intended shared model. Rehearsability can delay the transmission process due to the additional time needed to revise a message and can reduce synchronicity when quick answers are expected.

Reprocessability: Reprocessability refers to the extent to which a medium enables its users to reexamine a message during decoding after the message has been transmitted, e.g. the ability to reread an email that one does not understand fully versus the fluid nature of the message in a face-to-face conversation. This dimension is important when the content of a message is very complex and previous messages should also be taken into account. Like rehearsability, reprocessability can reduce the level of synchronicity when immediate answers are expected because it involves additional time (Dennis et al., 2008, p. 587).
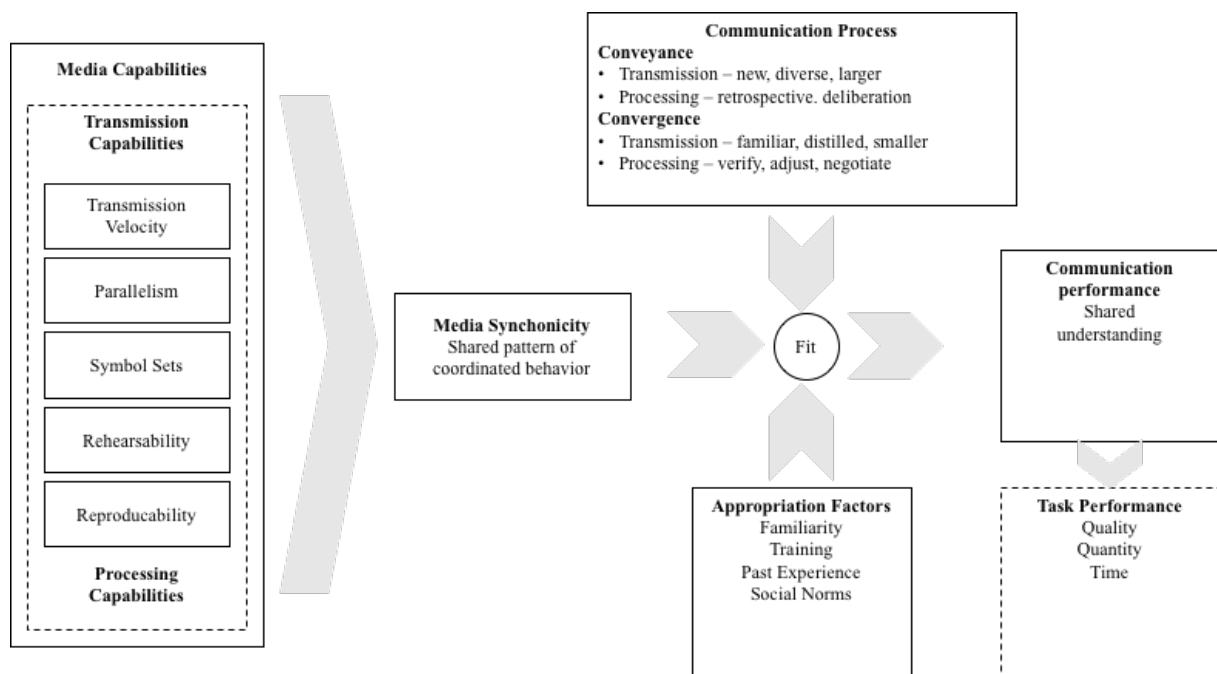
Figure 2.1. Media Synchronicity Theory (adapted from Dennis et al., 2008)

The five media capabilities from media synchronicity theory can also constrain communication. The "best" medium in terms of this theory might also be a combination of different media; the most important consideration is an optimal fit between media synchronicity, the communication process, and appropriation factors, as depicted in *Figure 2.1*. The main benefit of media synchronicity theory is that it offers a more nuanced view beyond the richness continuum. Applying media synchronicity theory's notions of dimensionality, reprocessability and parallelism could explain why some features (e.g.,

recording an interview) might reduce adverse effects or even exhibit positive effects on outcomes like reliability. Table 2.1 compares several technologies with respect to their scores on media capabilities from MST. Transferred to the interview context, face-to-face interviews are classified as high in synchronicity, while asynchronous video interviews would be considered low in synchronicity.

Table 2.1. *Comparison of selected media from media synchronicity theory (from Dennis et al., 2008, p. 589)*

Technology	Transmission Velocity	Parallelism	Symbol Sets	Rehearsability	Reprocessability	Information Transmission	Information Processing	Synchronicity
Face-to-face	High	Medium	Few-Many	Low	Low	Fast	Low	High
Video conference	High	Medium	Few-Many	Low	Low	Fast	Low	High
Telephone conference	High	Low	Few	Low	Low	Fast	Low	Medium
Synchonous instant messageging	Medium-High	Low-Medium	Few-Medium	Medim	Medium-High	Medium	Low-Medium	Medium
Synchronous electronic conferencing	Medium-High	High	Few-Medium	Medim	High	Medium	Medium	Low-Medium
Asynchonous electronic conferencing	Low-Medium	High	Few-Medium	High	High	Slow	High	Low
Asynchonous electronic mail	Low-Medium	High	Few-Medium	High	High	Slow	High	Low
Voice mail	Low-Medium	Low	Few	Low	High	Slow	Medium	Low
Fax	Low-Medium	Low	Few-Medium	High	High	Slow	High	Low
Documents	Low	High	Few-Medium	High	High	Slow	High	Low

2.4. Potosky's Conceptual Framework

Potosky (2008) presented a conceptual framework for the role of media in the personnel assessment process that considers the selection processes as an act of interpersonal communication between an organization and applicants, who gather information about and react to one another while making employment decisions. The intended evaluation of individual characteristics of a test-taker or interviewee is viewed as the message. Organizations need these evaluations to be accepted, reliable, valid, and fair. Technology is frequently used to conduct these evaluations, including face-to-face interviews, telephone or video interviews, web-based surveys, computer simulations etc. Technology in use has

structural attributes that affect the communication process in a way that accounts for systematic and unsystematic measurement variance, as depicted in *Figure 2.2*. Potosky introduces a list of four media attributes, some of which directly refer to the aforementioned theories.

Transparency: *Transparency* refers to the extent to which a medium allows the clear and smooth exchange of information. A medium with low transparency is highly salient and evident. When a medium is low on transparency, e.g. due to a disturbed Internet connection or low physical bandwidth on a Skype call, interactants can be distracted by the medium and must expend additional effort in maintaining the communication process. A medium possesses the highest level of transparency if interactants are not aware of the presence of the medium at all (Potosky, 2008).

Social Bandwidth: *Social bandwidth* describes the degree to which a communication medium can transmit social information. This concept is similar to the concept of multiple cues in media richness theory described above and refers to the number of social cues that can be included or transferred within a message. For example, video technology, which can transmit nonverbal and paralinguistic cues, has higher social bandwidth compared to telephone interviews. Social bandwidth differs from transparency in that transparency refers to the attention paid to a medium and social bandwidth to the number of social cues that can be transmitted (Potosky 2008).

Interactivity: Interactivity refers to the opportunities a medium gives to provide immediate feedback. This attribute is equivalent to the concept of immediate feedback in media richness theory. For example, face-to-face interviews are high on interactivity because interviewees can correct their answers immediately if they see that the interviewer disagrees, while interviewers can probe for further elaborations of candidates' answers. In asynchronous interview formats low on interactivity, neither supplementary questions nor probing is possible.

Surveillance: Surveillance describes the degree to which outside parties can monitor or intervene in the communication process. Hence, surveillance is a relatively objective attribute, but it also refers to users' subjective perceptions of suspected or anticipated surveillance. With respect to assessment practices, perceived surveillance is much higher when a proctor is present at the testing site compared to unproctored Internet testing. Furthermore, asynchronous video interviews are much higher on surveillance than face-to-face interviews, as interviewees might suspect that unauthorized persons will be granted access to the recorded video sequences.
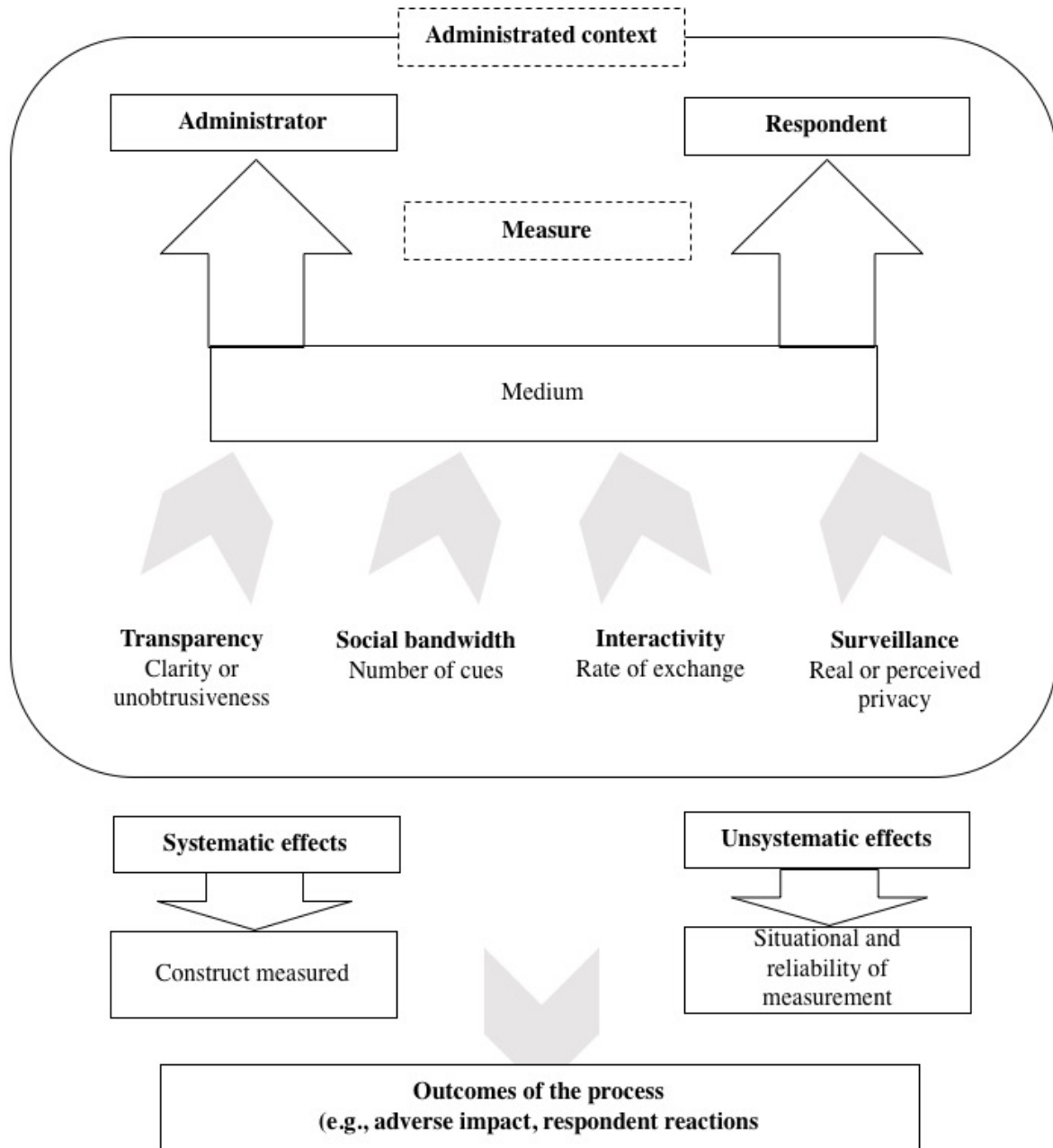
Figure 2.2. The Potosky (2008) framework

In Potosky's framework, all media are subject to a structural range defining the limits of each attribute. Assessment can only be designed within these limits. For example, an organization can choose to maximize social bandwidth and interactivity by using a high-end video solution. Nevertheless, these settings are not fixed and can dynamically change, e.g. when the Internet connection begins to buffer and a user adapts to the new circumstances by speaking more slowly.

2.5. The Information Process-Based Conceptual Framework (IPCF)

Arthur et al. (2018) proposed an information processing-based conceptual framework to address the growing variety of devices used in unproctored Internet testing (UIT) and explain score differences in such tests, which are completed by test-takers on their personal computers or mobile devices without any surveillance. In contrast to the aforementioned models, Arthur et al.'s framework describes structural components of technologies that are proposed to affect test-takers' cognitive demands. These components are screen size, screen clutter, response interface, and permissibility. The model is based on the findings of a recent meta-analysis of over twenty published and unpublished studies, which found that device type lowers test scores in cognitive tests, while reliability and factor structure are not affected (Arthur et al., 2018). A large screen size with low screen clutter, a non-challenging response interface, and low permissibility result in lower information processing demands, while small screens, high clutter, a highly-challenging response interface, and high permissibility result in higher cognitive load: demands on working memory, perceptual speed and visual acuity, psychomotor abilities (e.g. use of touch screens), and selective attention resulting from the presence of distractors are higher. It has further been suggested that the framework is also applicable to other tests formats, such as virtual role plays, immersive simulations, and gamified assessments (Morelli et al., 2017). Nevertheless, one major weakness of the model is that it does not define any "breaking points" between--for example--laptops and mobile devices, which is problematic as the boundaries between them are gradually fading away. Hence, it has been considered to have limited explanatory value for explaining differences in face-to-face versus technology-enhanced interviews (Morelli et al., 2017). Nevertheless, the information processing perspective might gain relevance for interviews in the future, as evaluating candidates in recorded interviews always involves the usage of a device, which can lead to systematic variation in psychological processes and in turn affect selection outcomes.

2.6. Technology Acceptance Model

The technology acceptance model (TAM), originally formulated by Davis and colleagues (Davis, 1989; Davis et al., 1989) and later extended and modified by Venkatesh and colleagues (Venkatesh & Bala, 2008; Venkatesh & Davis, 2000; Venkatesh et al., 2003), suggests that the acceptance of a new technology is determined by its *perceived usefulness*[10] (PU) and *perceived ease of use*[11] (PEoU). PU and PEoU predict behavioral intentions towards a technology, which in turn predicts actual system usage, as depicted in *Figure 2.3*. The later versions of the model identify a variety of antecedents of PU and PEoU, including subjective norms, computer self-efficacy, perceived enjoyment, and output quality (Venkatesh & Bala, 2008). Meta-analytic results provide robust support for the TAM and its two core components (King & He, 2006). The model has also been used to explain the acceptance of new technologies in human resources and asynchronous video interviewing (Brenner et al., 2016; Maier, Laumer, Eckhardt, & Weitzel, 2013). For technology-enhanced interviews, TAM is valuable for gaining a better understanding of applicant reactions beyond the Gilliland model, e.g. when the information given to applicants is ambiguous with regard to its perceived usefulness (Langer, Konig, & Fitili, 2018).
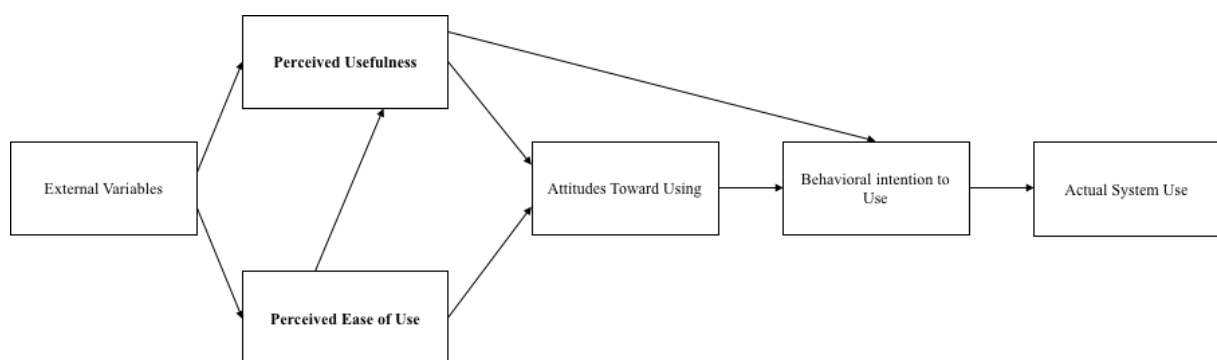
Figure 2.3. The Technology Acceptance Model

---

[10] Davis defines perceived usefulness in organizational contexts as "the prospective user's subjective probability that a specific application system will increase his or her job performance or more general „a system high in perceived usefulness (…) is one, for which a user believes in the existence of a positive use-performance relationship" (p. 320)

[11] Perceived ease of use refers to "the degree to which a person believes that using a particular system would be free of effort" (Davis, 1989, p. 320). Davis claimed that systems are more accepted when they are easier to use even if all other components are equal. In summary, the higher the perceived usefulness and ease of use, the higher the probability that a system will actually be used.

2.7. The Theory of Sociomaterial Practice

The idea that an understanding of technological effects must include an analysis of humans' interactions with machines is rather new in research on organizational behavior, but well-established in research on socio-technical systems (Cooper & Foster, 1971). Potosky (2008) acknowledged that assessments occur in a socially constructed administration context. Thus, the theory of socio-material practice from organizational science could potentially be useful for understanding this phenomenon (Orlikowski, 2007). Unlike the aforementioned frameworks, socio-material theory does not analyze individual capabilities or attributes, but rather views technology and its social usage as reciprocally "entangled" in a mutually-shaped "assemblage". In short, technology affects human behavior and human thought, and these in turn affect technology. As an anecdotal example, Orlikowski (2007) refers to Google's page rank algorithm. This algorithm which creates a weighted ranked list of websites by examining backlinks. Thus, page ranks reflect what people consider to be relevant. Page ranks are continuously updated as a result of any number of mundane changes; hence, the result list at any one moment in time is 'temporally emergent'. Therefore, "the result is a constitutive entanglement of the social and the material — 'a mangling of human and material agencies'" (Orlikowski, 2007, p. 1440). For the personnel selection domain, this theory is valuable for shifting from a static view of technology in the form of fixed assignments of media attributes or tools towards a focus on the applicant's psychological processes and behavior patterns related to technology usage during the assessment (Morelli et al., 2017). Thus, the perspective of human-technology interaction might be very valuable for analyzing how evaluation practices can be changed to better respond to their users. For example, when software applications for asynchronous video interviews force users to rate every candidate on a very detailed rating scale up to the last video sequence, users might get annoyed if they have

already made a decision[12]. As a result, fewer people will use the application, and the vendor might change the application in response to this user feedback, perhaps introducing instead a very simple rating format that enables candidates judged as not suitable to be swiped away. The validity of the application will likely suffer due to this social-material interaction. The socio-material perspective might also be quite useful for analyzing the recruiting process from a signaling game perspective (Bangerter, Roulin, & Konig, 2012), with an eye to how increasingly adaptive and personalizable technologies influence the signaling game between applicants and hiring organizations.

2.8.   Summary

This chapter has summarized several conceptual approaches to media, beginning with media richness theory (Daft & Lengel, 1986) and proceeding to media synchronicity theory (Dennis et al., 2008; Dennis & Valacich, 1999); two practitioner-oriented frameworks by Potosky (2008) and Arthur et al. (2018); the most popular family of theories for technology adaption, the *technology acceptance model* (Davis, 1989; Davis et al., 1989); and finally Orlikowski's (2007) socio-material theory.

The first four theories have in common that they describe media or technological characteristics either in terms of richness characteristics (immediacy of feedback), media capabilities, attributes, or device-oriented structural characteristics. Despite this variation in terminology, all of the theories taking the media characteristics view propose that web-based asynchronous video interviews are a less rich or less synchronous variation on the original analog format. As a result, in the classical test theoretical view taken by Morelli et al. (2017), the medium is a potential source for increased unsystematic error in an assessment, while Arthur et al.'s (2018) framework even specifies for which devices this source of potential error is larger due to increased cognitive load. Media synchronicity offers a more nuanced

---

[12] Interviewers tend to form an initial impression very early (Swider, Barrick, & Harris, 2016)

view on communication outcomes (interviews are considered to be an exchange of information and therefore a communication process) by introducing the concepts of rehearsability and reprocessability, which might lead to beneficial effects of technology on psychometric properties, for example, by allowing for multiple raters, which is sometimes not possible in face-to-face interviews due to practical limitations. The family of theories developed on the basis of the technology acceptance model (TAM) provide a framework that is not only applicable for understanding classical reliability and validity issues related to new selection technologies, but also offers insights into technology adoption by candidates and recruiters, which is a further important research issue (Anderson, 2003). If top candidates withdraw from the selection process because they do not accept the use of a given technology, its utility for the entire selection process might suffer (K. R. Murphy, 1986). Finally, Orlikowski's (2007) theory of socio-material practice offers a very different perspective on the interaction between the social and technology domains by identifying their reciprocal dependencies.

3.   Previous Research

Chapter 3 integrates the literature on traditional employment interviews and technology usage with an eye to the systematic impacts of technology on psychometric properties and selection outcomes. In light of the considerable body of literature on employment interviews in general, including numerous and recent meta-analyses and review articles (Levashina et al., 2014; McDaniel et al., 1994; Posthuma et al., 2002), I will limit this review to the main findings and focus on the integration of the technology perspective. The interview outcomes discussed in this chapter are interview reliability, validity, applicant reactions, adverse impact, fakeability and impression management as well as economic efficiency.

3.1.   Employment interview reliability

Explanation and findings regarding FTF interviews: Reliability is important for employment interviews for several reasons: According to classical test theory, reliability sets the upper limit for validity. Evidence suggests that differences in validity between structured and unstructured interviews are mainly due to differences in reliability, and reliability estimates are important for accurately estimating corrected criterion validity coefficients (Huffcutt, Culbertson, & Weyhrauch, 2013).

Research on interview reliability has focused on several indicators. Interrater agreement seeks to quantify the extent to which absolute scores gathered from two or more raters converge; interrater reliability captures the rank order stability in ratings obtained from two or more raters assessing two or more interviewees (LeBreton & Senter, 2008). Rater accuracy, on the other hand, describes how accurate ratings are compared to a given "true" score, which is typically obtained through ratings by subject matter experts (LeBreton & Senter, 2008; Sulsky & Balzer, 1988).

In the literature, there are two meta-analyses focusing explicitly on interview reliability (Conway et al., 1995; Huffcutt et al., 2013) and three additional meta-analyses on interview

validity that also address reliability (McDaniel et al., 1994; Taylor & Small, 2002; Thorsteinson, 2018; Wiesner & Cronshaw, 1988). The main finding of these meta-analyses is that reliability in face-to-face interviews mainly depends on differences in the interview's level of structure. Huffcutt et al. (2013) reported reliability coefficients ranging from .36 for the lowest level of structure to .76 for the highest level on a five-level categorization of structure in which Level 5 refers to individual ratings with complex scales. The study reported further significant differences between individual and panel interviews,[13] which had average interrater reliabilities of .44 versus .74, respectively[14]. Most recently, Thorsteinson (2018) reported reliabilities of .50 for weakly structured interviews and .84 for highly structured ones, as well as .44 for individual interviews and .75 for panel interviews.

Media impacts: In addition to structural components, the media format has been discussed as another source of systematic and unsystematic variance that might affect the interrater agreement or accuracy of employment interviews (Potosky, 2008). Several other theoretical approaches support Potosky's view that differences in interrater agreement and accuracy are affected by the media format of employment interviews. Therefore, it is not trivial to assume that technology-enhanced employment interview formats always possess the same level of agreement and accuracy as traditional forms of employment interviews, even when they have the same structure. Unfortunately, there are almost no data available about the reliability of technology-enhanced interviews. A study on the influence of picture-in-picture on performance ratings in synchronous video interviews reported an ICC(2) of .84 for overall performance ratings (Horn & Behrend, 2017). The only study addressing potential differences between different interview formats found significantly lower Cronbach's alpha values for ratings based on face-to-face interviews compared to videotaped interviews (Van Iddekinge et

---

[13] Panel interviews (also known as 'board interviews' or 'team interviews') involve multiple (at least two) interviewers who interview candidates together and aggregate their ratings into one overall score per candidate (Macan, 2009).
[14] Note that Huffcutt et al. (2013) discussed the overestimation of interrater reliability in panel interviews

al., 2006). In contrast, interrater reliabilities were found to be significantly higher for face-to-face (.73) panel interviews compared to videotaped ratings (.58) of the same interviews (p. 355).

3.2.   Employment interview validity

The validity of a score on any psychological assessment, including interview ratings, is "an overall evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions on the basis of test scores or other modes of assessment" (Messick, 1995, p. 741). The following section presents main findings about interview validity in face-to-face interviews and relevant literature regarding the impact of technology usage for criterion-related validity, incremental, and construct validity.

Criterion-related validity

Explanation and findings regarding FTF interviews: Previous research has focused extensively on the criterion-related validity of structured employment interviews. Thus, numerous meta-studies and review articles exist (Huffcutt & Arthur, 1994; Huffcutt et al., 2014; Huffcutt, Roth, & McDaniel, 1996; McDaniel et al., 1994; Schmidt & Hunter, 1998; Thorsteinson, 2018). Meta-analyses have consistently found that highly structured interviews possess higher criterion-related validity than weakly structured interviews. Most recently, Thorsteinson (2018) reported coefficients of .49 for weakly structured, .63 for medium structured, and .57 for highly structured interviews after correcting for criterion unreliability and range restriction. A few years earlier, Huffcutt et al. (2014) reported validity estimates of .20 for weakly structured interviews up to .70 for the highest category of structure. Compared interviews to other typical selection procedures, Schmidt and Hunter (1998) reported a meta-analytic average corrected correlation of .51 between overall job performance and structured interviews, .38 for unstructured interviews, .37 for assessment centers and .51 for general mental ability (GMA) testing. Validity coefficients for the most popular criterion

indicators, namely overall job performance and training performance, have been comparable in previous meta-analyses (McDaniel et al., 1994; Schmidt & Hunter, 1998).

Surprisingly, the criterion validity of employment interviews is not higher when panel interviews are used compared to individual interviewers, despite the fact that multiple interviewers rating one candidate is considered more reliable (McDaniel et al., 1994). The availability of cognitive test scores has also been found to reduce the criterion-related validity of employment interviews (McDaniel et al., 1994). With respect to question type, Levashina et al. (2014) summarized that situational interview questions have slightly lower criterion-related validity than past behavior questions, especially for jobs high in complexity.

Media impacts: Research on criterion-related validity for technology-enhanced interviews is scare. Gorman et al. (2018) reported significant correlations between asynchronous interview ratings and self-rated job performance and organizational tenure. The highest correlations were found between interview ratings of knowledge and skills and self-reported job performance (.48) and lowest for ratings of applied social skills and self-reported job performance (.26). The mean correlation between dimensional ratings and organizational tenure was .19 (*ns*) with a range of .10 to .28. Indirect support for the validity of technology-enhanced open response formats similar to interviews comes from related research areas. For example, Funke and Schuler (1998) found that SJTs' criterion-related validity is higher when response fidelity is high in open formats. Another study found open-ended webcam tests to be related to placement success above and beyond job knowledge tests (Oostrom et al., 2010).

Incremental validity

Explanation and findings regarding FTF interviews: Combining multiple predictors is common in practice. Thus, the incremental validity of one predictor over another is a central issue in personnel selection research (Sackett, Dahlke, Shewach, & Kuncel, 2017). Structured interviews often provide incremental validity over other selection procedures. Meta-analytic evidence suggests that they are only moderately related to cognitive ability tests and

personality traits (Berry, Sackett, & Landers, 2007; Cortina, Goldstein, Payne, Davison, & Gilliland, 2000; Salgado & Moscoso, 2002). Most prominently, Schmidt and Hunter's (1998) meta-analysis concluded that combining GMA tests and structured employment interviews is most effective for predicting overall job performance.

One field study further found that predicters richer in response fidelity (e.g., open response formats) are incrementally valid and less cognitively saturated than less rich ones (Lievens et al., 2015). Although the majority of existing research concludes that combining cognitive testing and interviews is an effective selection strategy, Sackett, Dahlke, et al. (2017) showed that combining two predictors could result in lower validity compared to one predictor alone when cut-off scores are applied in so-called hurdle selection procedures with multiple stages.

Media impacts: A field study by Lievens et al. (2015) found that increased response fidelity positively affects predictive validity in the selection of entry-level police officers, by reducing cognitive saturation and increasing applicants' reactions. This might indicate that if fidelity and media attributes are related, changes in media attributes could impact incremental validity.

Construct-related validity

Explanation and findings regarding FTF interviews: Research into the construct validity of employment interviews is important for two reasons: First, it shows which specific constructs (e.g., motivational dispositions, personality traits) explain variation in interview scores. Second, by showing how construct variation measured in employment interviews is related to job performance, construct validity serves as the theoretical underpinning for understanding employment interview criterion relationships (Hamdani, Valcea, & Buckley, 2014).

Employment interviews measure multiple constructs, including cognitive ability, job knowledge and skills, applied social skills, and personality (Harris, 1999; Huffcutt, 2011;

Huffcutt et al., 2001). Huffcutt et al.'s (2001) meta-analysis summarized a sample of 47 primary studies with respect to the measured constructs, finding that employment interviews most frequently measured basic personality tendencies (35%), followed by applied social skills (28%), mental capabilities (16%), and knowledge and skills (10%). Applied social skills were addressed more frequently in highly structured interviews (34%) compared to weakly structured interviews (18%). With respect to construct-related validity, Huffcutt and colleagues reported mean average validity coefficients ranging from .24 to .58, with the highest estimates for creativity (.58), followed by agreeableness (.51), and organizational fit (.49). The lowest validity estimates emerged for interests and preferences (.24), general intelligence (.24), communication skills (.26), and applied mental skills (.28).

Interview ratings are further influenced by candidates' demographic characteristics and ability to use impression management tactics (Huffcutt, 2011). The relationship between interview ratings and cognitive ability has received the most attention in the literature on employment interviews' construct validity (Berry et al., 2007; Cortina et al., 2000; Huffcutt et al., 1996; Roth & Huffcutt, 2013). Meta-analytic evidence has found GMA to be among the best single predictors of job performance and training program performance (Schmidt & Hunter, 1998). The relationship between employment interviews and cognitive abilities could be a double-edged sword, as high cognitive saturation (i.e., a strong correlation) could increase construct validity when interviewers intend to predict job behavior dependent on 'g', but also could lower the potential to explain additional criterion variance when job behavior also greatly depends on other characteristics (Huffcutt et al., 1996).

Meta-analyses have reported a corrected average correlation between face-to-face interviews and GMA of a magnitude ranging between .27 and .42 (Berry et al., 2007; Cortina et al., 2000; Huffcutt, 2011; Huffcutt et al., 1996; Roth & Huffcutt, 2013; Salgado & Moscoso, 2002). Interview questions related to technical problems or problem-solving often assess cognitive ability directly (Huffcutt et al., 2001). Candidates with higher cognitive

ability understand interviewers' intentions better (König, Melchers, Kleinmann, Richter, & Klehe, 2007) and may also be better at applying impression management tactics (Barrick, Shaffer, & DeGrassi, 2009).

Media impacts: Blackman's (2002) study comparing telephone to face-to-face interviews found that valid judgments on social skills and some personality traits require richer media formats. Schuler and Funke (1998) found in a student sample that the response mode of a test is critical for its validity: Open-ended written responses to a SJT were revealed to be significantly better criterion predictors for a behavior-oriented roleplay than the same test with a (less rich) multiple-choice response format. Lievens and Sackett (2006) assessed differences in validity between written versus video-based SJTs and found that the richer video-based format exhibited higher construct validity. Most recently, Lievens et al. (2015) reported based on a field sample with police officers that behavioral responses were found to be slightly higher in validity, more highly correlated with personality measures, less cognitively saturated, and perceived as more favorable compared to less rich written responses.

Evidence on the construct-related validity of video interviews that do not involve direct interactions, such as asynchronous video interviews, comes from research on thin slices of personality, which has found that people are capable of quite accurately encoding complete strangers' personality traits, abilities, and motivational states on the basis of only short behavioral samples in the form of video segments (Ambady & Rosenthal, 1992; Borkenau, Mauer, Riemann, Spinath, & Angleitner, 2004; Hirschmuller, Egloff, Nestler, & Back, 2013). A study by N. A. Murphy (2007) reported that ratings of intelligence are more valid when media formats are richer, which was tested by comparing video material to the corresponding transcripts of interaction excerpts. The same study further found that people who applied impression management behaviors were successful in appearing more intelligent to their direct interaction partners, but the accuracy of intelligence ratings by judges who rated the

same interactions using video recordings did not change. This implies that interviewers who are not involved in richer face-to-face interactions can make more accurate judgments by gaining distance from potentially distorting cues conveyed via richer communication. A second study from the same article found that perceived intelligence was correlated with several nonverbal cues, including speech style, eye contact, expressive voice, and gesturing, while psychometrically-measured intelligence did not correlate with many of these cues. This indicates that individuals largely base their judgments of a target's intelligence on cues that exhibit no cue validity with respect to psychometric intelligence.

3.3. Applicant reactions

Explanation and findings regarding FTF interviews: Research on applicant reactions investigates attitudes, affects, and cognitions that individuals have or might have at any point during the recruitment and selection process (Ryan & Ployhart, 2000). Studying applicant reactions to selection procedures is important because they are significantly related to job offer acceptance intentions, organizational attractiveness, and the likelihood of buying a company's products (McCarthy et al., 2017). Overall, meta-analyses have consistently found that traditional face-to-face interviews are seen as the most preferable selection procedure compared to many others, including mental ability tests or work samples (Anderson, Salgado, & Hulsheger, 2010; Hausknecht, Day, & Thomas, 2004). A recent meta-analysis suggests that technology in employment interviews generally exhibits lower levels of applicant favorability (Blacksmith et al., 2016).

A higher level of structure has been suspected of having a negative impact on applicant reactions because it limits candidates' use of impression management tactics (Chapman & Zweig, 2005). However, the more recent review by Levashina et al. (2014) did not find any systematic differences in applicant reactions depending on level of structure. Beyond structure, several studies have investigated other possible moderator variables of applicants' reactions with respect to the fairness of structured interviews, including the level of

transparency (Day & Carroll, 2003), informational content (Kohn & Dipboye, 1998), and the medium itself (Chapman & Rowe, 2002).

Media impacts: The most popular model for explaining variations in applicant reactions is rooted in organizational justice theory and suggests that candidates' overall fairness perceptions and process favorability are based on the extent to which procedural and distributional justice rules are followed or violated (Gilliland, 1993). One study that applied the Gilliland model to new selection technologies in a student sample found that justice rules referring to interpersonal communication explained variation in fairness reactions between face-to-face interviews, telephone interviews, and interactive voice response formats with identical content (Bauer et al., 2004). Sears et al. (2013) found that candidates rate justice rules concerning opportunity to perform, job-relatedness and the availability of selection information significantly lower in videoconference interviews compared to face-to-face interviews with identical questions. The Gilliland model was also used by Langer et al. (2017), who showed that asynchronous video interviews were rated lower than synchronous video interviews on the dimensions two-way communication and interpersonal treatment and higher with regard to privacy concerns and creepiness. Brenner et al. (2016) assessed the antecedents of attitudes towards asynchronous video interviews by applying the technology acceptance model, showing that perceived usefulness and perceived ease of use account for a large proportion of the variance in candidates' favorability ratings of asynchronous video interviews.

Privacy concerns, specifically the unauthorized and abusive use of information or test results, have also been suspected to impact applicant reactions to technology-enhanced assessments (Bauer et al., 2006). Evidence from both lab studies and a field sample have found that information privacy concerns negatively affect applicants' procedural justice perceptions (Bauer et al., 2006). Similarly, computer experience was found to moderate the relationship between procedural justice perceptions and test-taking motivation (Bauer et al.,

2006). Langer et al. (2018) found in a student sample that the provision of detailed information to candidates, but not computer experience per se, affects candidates' reactions to new selection technologies and organizational attractiveness. This effect is likely moderated by the extent to which candidates perceive a given selection device as useful.

Applicant reactions are also influenced by major social and technological meta-trends (McCarthy et al., 2017). In the technology acceptance model as well as the theoretical tradition based on Ajzen's theory of planned behavior, this effect is explained by pointing to subjective norms as one antecedent of user acceptance (Ajzen, 1991). Anecdotal evidence in support of this comes from the growing acceptance of social media recruitment: a few years ago, there was extensive debate among I/O practitioners (Roth, Bobko, Van Iddekinge, & Thatcher, 2016) about whether and how to use data from social and professional networks due to potentially negative reactions from candidates. Today, most applicant tracking systems have already features that allow candidates to apply directly using their social media profiles, and CVs are becoming more and more outdated. The social norm has changed to such an extent that many candidates currently expect to apply for jobs using their social media profiles.

Previous research has also addressed interview anxiety, leading to the development of a multidimensional measurement, the Measure of Anxiety in Selection Interviews (MASI) (McCarthy & Goffin, 2004). These authors further found negative correlations between job interview anxiety and interview performance. It has also been proposed that interview anxiety might influence predictive validity (Macan, 2009), or that job interview anxiety might be a moderator for technology-enhanced employment interview format: if interview anxiety depends on direct contact with the interviewer, persons with high interview anxiety might be less anxious in interviews with a lower social bandwidth.

3.4.   Adverse impact (Subgroup differences)

Explanation and findings regarding FTF interviews: One major criticism of unstructured interviews is their susceptibility to various biases, including with respect to personal characteristics such as race, gender, disability, or overweight job candidates (Arvey & Campion, 1982; Kutcher & Bragger, 2004; Levashina et al., 2014). Historically, score differences between subgroups such as people with different ethnic backgrounds have been the subject of highly controversial discussions (Herrnstein & Murray, 1994). Legal regulations, such as the "Allgemeines Gleichstellungsgesetz – AGG" or "General Equal Treatment Act – GETA" in Germany, strengthen the rights of applicants from such disadvantaged subgroups and require employers to prove that their selection systems are fair if individual candidates file lawsuits alleging that they have been treated unfairly. Hence, selection systems should be designed to avoid any adverse impacts for specific subgroups in order to avoid lawsuits and public damage to employers' reputations.

For traditional employment interviews, past meta-analyses have found significant subgroup differences between black and white applicants of $d = .32$ for weakly structured interviews and $d = .23$ for highly structured interviews. Furthermore, according to Huffcutt and Roth (1998), the differences were more pronounced for interviews high in cognitive saturation than for interviews with lower saturation ($d = .45$ and $d = .26$, respectively). In contrast, the more recent review by Levashina et al. (2014) analyzed all studies on structured interviews after 1996 and did not find any meaningful mean differences due to gender, race or minority status. Thus, large-scale studies using a high level of structure in high-stakes solutions seem to be less susceptible to adverse impact. Kutcher and Bragger (2004) conducted the only laboratory studies on this topic, showing with the help of a fat suit that the level of structure reduced advance impact against overweight job candidates.

Media impacts: Employment interviews are generally vulnerable to potential adverse impacts, as visual cues that can activate stereotypes are very salient by default. One of the

best known stereotypes is the attractiveness stereotype, in which people tend to evaluate physically attractive people more favorably (Feingold, 1992). Thus, unsurprisingly, employment interview ratings correlate significantly with physical appearance (Barrick et al., 2009). As some technology-enhanced employment interview formats, e.g. telephone interviews, do not transmit such visual information, technology might be a relevant factor in addition to structural components. Straus et al. (2001) found a significant interaction between physical attractiveness and interview medium on interview ratings. Less physically attractive candidates received lower ratings in face-to-face and videoconference interviews compared to telephone interviews, which filtered out negative cues. Nevertheless, physically attractive candidates did not receive higher ratings in the face-to-face condition. Sears et al. (2013) reported that candidates' physical appearance was rated lower in synchronous video interviews compared to face-to-face interviews, but did not report any moderation effect of medium on the link between physical appearance and interview ratings. Kroll and Ziegler (2016) conducted a laboratory study investigating the susceptibility of highly structured asynchronous video interviews to discrimination due to gender and ethnic background. The results did not indicate any bias with respect to gender or minority status (German versus Turkish). In summary, there are arguments for both perspectives postulating that media intermediation can be either beneficial or harmful for psychometric properties.

3.5. Fakeability / Susceptibility to faking / Impression management

Explanation and findings regarding FTF interviews: Impression management (IM) refers to the process through which people seek to control the impressions others form of them, and plays an important role in many interactive behaviors (Leary & Kowalski, 1990). In employment interviews, applicants frequently engage in impression management tactics (Ellis, West, Ryan, & DeShon, 2002; Stevens & Kristof, 1995), and a meta-analysis by Barrick et al. (2009) found impression management tactics to be related to interview ratings. The relationship was more pronounced in weakly structured interviews compared to highly

structured interviews. However, impression management tactics were not correlated with actual job performance.

The latest meta-analytic evidence from Levashina et al. (2014) indicates that self-promoting IM tactics (e.g., claiming achievements as one's own success) are most strongly related to interview ratings, followed by non-verbal IM tactics (e.g., eye contact), other-focused IM tactics (e.g., blaming others for failures), and defensive IM tactics. The frequency of IM tactic usage also depends on the type of questions posed in the interview (situational questions versus past behavior questions).

As almost all candidates employ some IM tactics, the literature also differentiates between deceptive and non-deceptive IM tactics, with deceptive tactics referred to as interview faking behavior. Levashina and Campion (2007) identify four distinct types of interview faking behaviors: (1) slight image creation – exaggerating or embellishing one's qualifications or experiences, (2) extensive image creation – the intentional invention of experiences and qualifications, (3) image protection – intentionally omitting or masking undesirable experiences or qualifications, and (4) ingratiation - insincere praise of the interviewer or organization. One recent laboratory study found in a faking condition that applicants can intentionally fake structured interviews (Van Iddekinge, Raymark, & Roth, 2005). One open point in the discussion on the usage of impression management tactics refers to the distinction between impression management tactics that are "noise", or irrelevant variance that distorts ratings and decreases validity, and tactics reflecting job-relevant variance, which can be the case for restaurant servers as well as white-collar jobs such as consultants or sales representatives (Levashina et al., 2014).

Media impacts: In technology-enhanced interviews, the medium and the available social bandwidth can expand or limit applicants' opportunities to apply impression management tactics. In telephone interviews, all kinds of nonverbal impression management tactics such as nodding or maintaining eye contact are prevented by default. In synchronous video

interviews, the webcam's focus on the torso also limits the opportunity to engage in extensive gesturing or other similar behaviors. In asynchronous interviews, in which "the interviewer" remains completely unknown and unseen, impression management tactics such as ingratiation via opinion confirmation or the like are not possible. With respect to video resumes, one laboratory study found that the effectiveness of the use of impression management tactics could also be moderated by gender (Waung, Hymes, Beatty, & McAuslan, 2015), highlighting the potential for interaction effects between media formats and candidate characteristics. Thus, there are many reasons to take a closer look at the medium as an important factor or moderator.

3.6.   Economic efficiency of technology usage / Utility

Explanation and findings regarding FTF interviews. Cascio and Boudreau (2011) describe a method for conducting utility analyses of selection systems based on the Brogden-Cronbach-Gleser general utility equation. It contains quantity variables on applicants (number selected, number of total applicants), quality indicators (including validity, SD_y, average score of selectees on predictor, probationary period, multiple selection devices, departures form top-down hiring), and cost parameters (number selected, cost of selection, other economic factors). This model allows for the modeling of technology-related impacts on different parameters in the equation. The literature often refers to economic advantages, such as saving costs and time, as the main reason for technology usage (Anderson, 2003; Bauer et al., 2004). Applying the logic of Cascio and Boudreau (2011), technology usage will affect the number of applicants (if applicants can apply globally) and cost of selection (due to reduced travel costs), but still strongly depends on the validity and potential changes in the applicant pool.

Media impacts: Gorman et al. (2018) lists a range of benefits from using video technology in selection: (a) reduced travel expenses and related costs for candidates and interviewers, (b) the ease of flexible scheduling across time zones and borders, (c) a high level of consistency and structure, as all candidates are presented the same questions, (d)

interview questions tailored to the company's specific needs, (e) the ability to reprocess evaluations via replaying and reviewing, and (f) the ability to prescreen candidates before inviting them to face-to-face interviews (*p*. 130). The authors also refer to a study by the Aberdeen Group from 2010 that 73% of companies using video interviews saved an average of 9% on travel costs. Video-based interviews have also been suggested as an alternative to telephone interviews (Guchait, Ruetzler, Taylor, & Toldi, 2014). Various commercial providers have published ROI calculations on their websites[15], although these hardly be verified. One study presented by Tiller et al. (2013) found that conducting multiple mini-interviews via Skype in graduate recruiting in comparison to face-to-face interviews saved 84% of costs, including costs for airfare, accommodation, and other travel-related expenses, without causing significant score differences. In summary, cost and time savings as well as increased flexibility seem to be the major benefits of technology usage in interviews, although use cases vary strongly depending on cost structures, validity estimates, and selection ratios. Finally, it has been noted that utility estimates can be much too optimistic when they assume that all selected candidates will actually accept the job offer (K. R. Murphy, 1986).

3.7. Special issues

In addition, there are some other topics that have received attention in previous research on employment interviews that do not fit into the terminologies presented before but might nevertheless be relevant for technology-enhanced interviews.

When candidates are invited to complete asynchronous video interviews for the sake of selection, most software platforms provide a certain amount of time to prepare before the actual recording starts. This timeframe, hereinafter referred to as 'preparation time', can range from a few seconds in order to protect candidates from clicking on the record button by accident up to several minutes. The latter option gives candidates a proper chance to prepare

---

[15] E.g., https://cdn2.hubspot.net/hubfs/433227/ROI%20Workbook/Summary%20-%2010%20Metrics%20to%20Measure%20Video%20Interviewing%20ROI%20-%20Sonru.pdf; https://www.rivs.com/wp-content/uploads/ 2018/02/ROI-of-Video-Interviewing-RIVS.pdf

an adequate answer before recording starts. This inclusion of preparation time is one of the most crucial differences between asynchronous video interviewing and face-to-face employment interviews, (synchronous) videoconference interviews conducted via Skype, and telephone interviews. As candidates are highly motivated to create and maintain a favorable impression in selection situations (Leary & Kowalski, 1990; Levashina & Campion, 2007), they will tend to use this preparation time to engage in techniques that should inflate performance appraisals, such as impression management tactics or cheating. In terms of media synchronicity theory (Dennis et al., 2008; Dennis & Valacich, 1999), the availability of a certain time frame that can be used to fine-tune responses refers to the level of rehearsability. Given candidates' high motivation to produce a favorable impression in selection situations, they might find tactics of improving their interview performance or test scores, such as tutoring or coaching (T. J. Maurer, Solamon, & Troxtel, 1998).

One other special topic is the impact of technical disruptions during technology-mediated interviews, for example, when the Internet connection drops. One initial study that systematically manipulated technical distortions in a laboratory setting did not reveal large impacts on interview ratings (Gelléri & Kiefer, 2018). Nevertheless, the level and kind of disruptions might affect selection outcomes in high-stakes settings.

3.8. Summary

This chapter has summarized research on employment interviews with respect to the interview outcomes of reliability, validity, applicant reactions, adverse impact, fakeability and utility considerations. Previous research has consistently showed that structure is essential for achieving favorable results on most outcome dimensions. Unfortunately, the existing studies on the equivalence of employment interviews conducted using different media formats are limited by their lack of generalizability to new applications that appear on the market and due to their neglect of the fact that mean differences need not be directly linked to differences in validity. Empirical results on variables other than mean differences are still scare, and there is

almost a complete lack of studies investigating media effects on other outcomes such as reliability or validity. The few exceptions have identified significant differences when the interview medium is changed. For example, Van Iddekinge et al. (2006) found differences between face-to-face interviews and their corresponding video recordings in Cronbach's alpha as well as the correlations between the interview results and the Big Five personality dimensions. However, more research into the psychometric properties of other technology-enhanced employment interview formats is required, because differences resulting from the chosen media format could threaten the utility of a selection system as a whole.

4. Working Model & Research Questions

4.1. A modular approach to media attributes in employment interviews

Research on media impacts on employment interview outcomes has so far been conducted mainly from the equivalence perspective, which considers different forms of technology-enhanced interviews as holistic all-in-one packages. From this perspective, telephone interviews, synchronous video interviews, and asynchronous video interviews represent more or less homogeneous sub-categories within the higher-order selection procedure class of interviews. This point of view might have been justifiable historically, as telephones are quite homogeneous devices, but the traditional boundaries between distinct categories of devices have becoming increasingly porous since webbooks, tablets and hybrids have come to market and expanded the possible constellations of devices that can be used to conduct interviews (to speak nothing of differences in software). Thus, 'video interviewing' has experienced an impressive technological evolution since the first studies on this topic, which typically referred to videoconferencing systems that were long only affordable for larger organizations (Chapman & Rowe, 2001; Straus et al., 2001). Today, even the difference between phone and videoconference interviews is blurring as new telephones are capable of running a whole range of video chat applications, and the visual format can be supplemented by screen sharing, avatar usage, coding tests, and so on.

Rather than maintaining the holistic, all-in-one perspective as the supreme research paradigm in the field, Lievens and Sackett (2017) suggest adopting a modular approach. Drawing upon product design[16], they suggest breaking down assessments into smaller pieces, which they refer to as building blocks. Furthermore, they argue that a modular approach to selection systems would have three conceptual and practical benefits: First, a modular approach allows large, complex systems to be broken down into manageable parts while

---

[16] Baldwin & Clark's original idea behind the idea of modularity was threefold: (1) make complexity manageable, (2) enable parallel work, and (3) accommodate future uncertainty.

allowing the functionality of the system as a whole to remain a black box due to multiple possible interactions and interferences. For instance, analyzing technology-enhanced interview formats from this perspective makes it possible to consider mixed forms that are likely to occur in the future and analyze their individual components. Second, the modular approach makes it possible to identify and exploit similarities among selection procedures, which is beneficial for drawing hypotheses about the impacts of technologies that do not fit into a distinct class. Finally, the modular approach makes it possible to formulate *specific hypotheses that can be tested empirically*. Lievens and Sackett (2017) proposed that selection systems consist of the following components: (1) stimulus format, (2) contextualization, (3) stimulus presentation consistency, (4) response format, (5) response evaluation consistency, (5) information source, and (6) instructions. The value and practicability of this approach were demonstrated in a field study on entry-level police positions using the approach mentioned in previous sections of this dissertation (Lievens et al., 2015).

A modular approach is proposed to be beneficial for better understanding technology impact in employment interviews when the modules of interest refer to media attributes. Variation in modules along specific attributes allow for the formulation of very specific and testable hypotheses regarding assessment outcomes. As the existing typologies of media attributes contain both similarities and differences, it might be helpful to take the most advanced framework for the selection context as the foundation for a pragmatic working model and add attributes from other theories only when they are obviously meaningful and absolutely necessary for understanding a specific criterion or phenomenon, for the sake of parsimony.

4.2.   A modular framework for technology effects in employment interviews

As recommended by Morelli et al. (2017), research on technology in the assessment and selection domain should aim to understand *why* and *how* technology affects the measured

constructs, psychometric properties, or candidate reactions rather than simply comparing the outcomes of two different procedures from a holistic perspective.

With regard to structural components, frameworks already exist that are quite beneficial for explaining which aspects of structure influence reliability, validity, and applicant reactions. For instance, Chapman and Zweig (2005) proposed that reliability, validity and applicant reactions are influenced by the following factors: (a) questioning consistency, (b) evaluation standardization, (c) question sophistication, and (d) rapport building. They note that structure in employment interviews is not a continuum but rather a multifaced construct. Similarly, technology can hardly be considered along a unidimensional continuum from more to less technological (rich, synchronous); rather, it makes sense to apply a modular approach that identifies meaningful modules referring to specific media attributes.

As Potosky presented the most recent framework with the highest level of maturity for the selection domain, the basic ideas of this framework will form the foundation of this thesis. Moreover, three additional dimensions of media attributes relevant for selection outcomes were identified and added to the working model. To be included in the model, variation in the presence or absence of a given media attribute must be relevant for either the conveyance and convergence of selection-relevant information, and variation in or manipulation of the attribute should have an impact on either the construct measurement, applicants' or recruiters' reactions to the selection procedure, effectiveness, or other relevant outcomes.

As noted by Dennis et al. (2008) for MST, media characteristics cannot be described as physical phenomena. Instead, media attributes such as the immediacy of feedback, personalization, or social presence are social constructs influenced by prior experience and the usage context. Thus, media synchronicity theory introduces the term 'capabilities' to describe attributes or characteristics of media that are (1) specific enough for testing and (2) of potential relevance for communication *performance*. In this case, performance refers to evaluation performance in terms of reliability, validity and candidate reactions.

In addition to the four attributes of transparency, social bandwidth, interactivity, and surveillance from Potosky's framework, three capabilities from media synchronicity are considered to be important for understanding media impacts on selection outcomes: parallelism, rehearsability, and reprocessability. It is suggested that these seven attributes affect selection outcomes via broad underlying principles described in psychological theory (see Table 4.2).

In this chapter, I will first briefly discuss the relationship between media attributes and structural components before turning to a more detailed discussion of the impact of media attributes on interview outcomes. First, frameworks of interview structures (e.g., Chapman & Zweig, 2005) and frameworks of media attributes (e.g., Potosky, 2008) are not necessarily completely independent; the same phenomena can be described from both a structure perspective and a media attribute perspective. For instance, Levashina et al. (2014) proposed a structural component referring to whether or not an interview is recorded, which is in fact identical to the attribute of reprocessability from MST (Dennis et al., 2008). Second, media attributes can directly affect structural components or be at least strongly related to them. For example, when the level of interactivity is low, as is true for asynchronous interview formats, the structural component of prompting is limited by default due to a specific media attribute. Conversely, the levels of specific structural components might determine the levels of specific media attributes. Third, technology attributes can be completely independent of structural components. For instance, conducting a job analysis in order to construct interview questions is not affected by any media attributes and could occur before making the decision on whether to use any technological enhancements to conduct the interview. Lastly and most importantly for research design, there might be numerous interactions between media attributes and structural component. Thus, the development of a working model that makes specific hypotheses on the relationship between media attributes and interview outcomes must

acknowledge numerous possible interactions with structural components. *Figure 4.1* depicts the possible relationships between media attributes and structural components schematically.
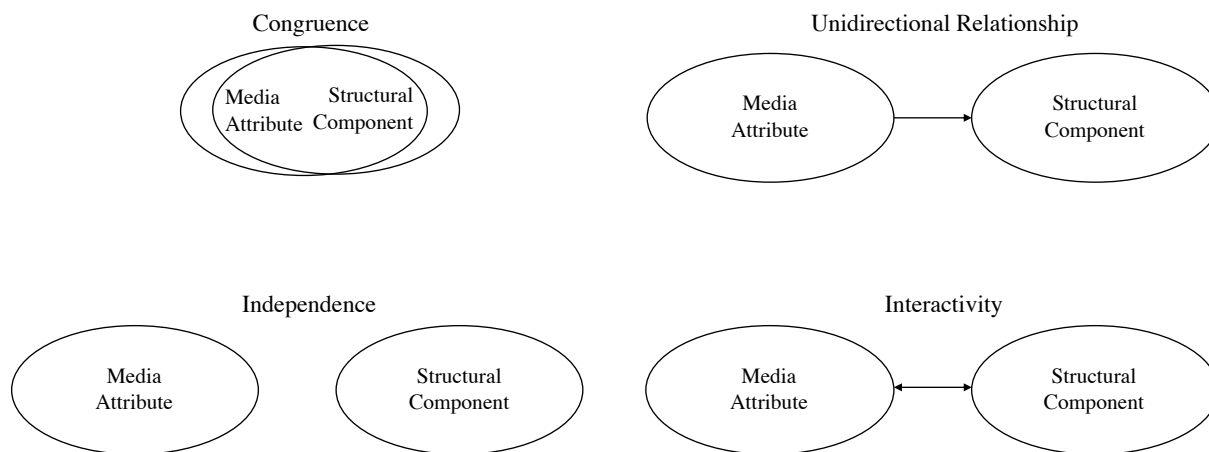
Figure 4.1. Possible relationships between media attributes and structural components

4.3.   How media attributes affect selection outcomes in employment interviews

Transparency

Definition. Transparency refers to the extent to which a medium facilitates a clear or unobstructed communication exchange (Potosky, 2008). In employment interviews, transparency acts as an important enabler of clear or unobstructed communication exchange. Face-to-face interviews possess the highest level of transparency by default.[17] Technology has low transparency when apps such as Skype suffer disturbances, as they draw attentional focus and become more salient. Consequently, interviewers and applicants must pay more attention to what is said, potentially leading to attention shifts.

Prior Research. There are at least two underlying theoretical frameworks for predicting differences in selection outcome. First, drawing upon an *information processing* perspective, such as cognitive load theory (Sweller, 1994) or multiple resource theory, it has been suggested that redundancy between visual and auditory information is an important predictor

---

[17] In the future, virtual reality applications might extend transparency on a horizontal level.

of an individual's information load and thus the amount of free resources (Basil, 2012). Transferred to the interview context, this means that high transparency should free up resources that can be used for other processes like judging a candidate's qualifications. Thus, reliability and accuracy should increase when the level of transparency is high.

Second, according to the *attribution literature* (Kelley, 1973), interviewers tend to confirm or disconfirm their expectations. Consistent information will further entrench the interviewer's initial opinion, while inconsistent information will be resisted (Albarracin & Shavitt, 2018). The greater a tool's transparency, the greater the opportunity to attribute behaviors to the medium. In a student sample, Chapman and Webster (2001) compared how differences in post-interview expressions in different media formats (face-to-face versus video interviews) were mediated by internal and external attributions and found that raters who perceived a medium to be less rich made more external attributions concerning applicants' performance and rated them more favorably. A first exploratory study by Gelléri and Kiefer (2018) found limited support for the notion that disturbances in sound, image or both affect interview performance ratings. It remains unclear whether a certain level of disturbance is needed for effects to emerge or whether effects depend on the timing of the disturbance.

Conclusion: In summary, both the information processing perspective and attribution literature see technology with low transparency mainly as a source of error in the communication exchange process that draws interviewers' focus away from judging a candidate's qualifications or makes interviewers' more lenient due to external attributions. Thus, when interviewers rate situations at random and technical disturbances randomly emerge, interviews will be both less reliable and less valid, as reliability sets the upper limit for validity.

Social bandwidth

Definition. Social bandwidth refers to the number of cues that a medium for employment interviews can transmit. Social bandwidth is an essential part of several media

theories (Daft & Lengel, 1986; Dennis et al., 2008; Potosky, 2008). In addition, social presence or social impact is an important driver of human behavior more generally (Latane, 1981; Short et al., 1976).

Prior Research. One foundation for predicting differences caused by an interview format's social bandwidth can be drawn from findings on *thin slices of personality*, which indicate that access to even small behavioral samples from complete strangers can produce accurate ratings on a wide range of traits (Ambady, Bernieri, & Richeson, 2000; Ambady, Krabbenhoft, & Hogan, 2006; Ambady & Rosenthal, 1992, 1993). Thus, it is not surprising that prior research has found vocal (e.g., pitch, pitch variability) and visual cues to be related to interviewers' judgements and even to predict job performance (e.g., DeGroot & Gooty, 2009; DeGroot & Kluemper, 2007; DeGroot & Motowidlo, 1999; Imada & Hakel, 1977; Parton, Siltanen, Hosman, & Langenderfer, 2002).

Nevertheless, evidence is scare on what amount of social bandwidth in interviews is required. A study by Motowidlo and Burnett (1995) in which 40 managers conducted simulated job interviews that were video-recorded found almost identical correlations between supervisor ratings and ratings by undergraduates who were provided the videotapes with sound, audio only, and videos only (.36 versus .33 versus .32), which indicates that all kinds of cues refer to some of the same variance between interviewees. A later study by Burnett and Motowidlo (1998) used video recordings of 60 managers providing answers to interview questions to compare student ratings on four dimensions of managerial effectiveness when using either the complete videotapes with video and sound, videos in which the audio file was erased, or transcripts of the answers (content only). A hierarchical regression analysis indicated that visual cues had a significant effect over and above content information, and that both visual and vocal cues were most strongly related to supervisor ratings. Indirect evidence of increased validity in response formats that can transmit more cues comes from a field study by Lievens et al. (2015), who found that audiovisual responses

were higher in validity and extraversion saturation but lower in cognitive saturation than written responses. Moreover, Lievens and Sackett (2006) discovered in a high-stakes selection process that video-based SJTs were higher in criterion-related validity compared to written SJTs. Video-based SJTs were also found to produce fewer subgroup differences due to the lower cognitive saturation of the stimulus format (Chan & Schmitt, 1997).

With respect to applicant reactions, as noted previously, prior research indicates that interview procedures that can transmit more social cues are preferred compared to ones that can transmit fewer social cues (Bauer et al., 2004; Sears et al., 2013).

From a social cognition perspective, it has been noted that high social bandwidth might be a potential source of bias because they make personal characteristics such as race, gender or physical attraction more salient, activating stereotypes (Blacksmith et al., 2016). Nevertheless, meta-analytic as well as experimental evidence suggests that structure is an effective instrument for avoiding these kinds of biases (Kutcher & Bragger, 2004; Levashina et al., 2014). Consequently, structure is suggested to be a moderator of the relationship between social bandwidth and reliability in technology-enhanced employment interviews.

Conclusion. According to the theory of small slices of behavior, social cues are important for making accurate judgements about strangers, which is also required in employment interviews. Prior research suggests that visual and vocal cues provide information above and beyond content information (Burnett & Motowidlo, 1998). Thus, reliability, validity and applicant reactions should benefit from technology that is richer in terms of social bandwidth as long as the interviews possess a sufficient level of structure.

Interactivity

Definition. Interactivity refers to the ability of interviewers and interviewees to give immediate feedback. A medium that is low on interactivity does not allow any form of immediate feedback, as is the case for asynchronous video interviews. Furthermore, communication can be impersonal, interpersonal, or hyperpersonal (Walther, 1996).

Prior Research.

From an information processing perspective, the theories of *cognitive load* (Sweller, 1994) and multiple resource theory (Basil, 2012) imply that face-to-face interviews are complex, cognitively challenges tasks for interviewers because they have to integrate and consider many input variables and cues in addition to simultaneously fulfilling their multiple roles of rater, stimuli provider, and organizational representative (Graves & Karren, 1992; Middendorf & Macan, 2002). Because cognitive resources, and especially working memory capacity, are limited in time and duration (Baddeley, 2012), resources expended on one role, e.g., representing the organization as an attractive employer, might mean that fewer resources are available for making accurate judgments. Thus, media formats that reduce cognitive load in employment interviews by eliminating the interactive component might increase inter-rater agreement and accuracy. Empirical evidence for the information processing perspective comes from Marr and Cable (2014), who found that interviewers who focused more on their role as organizational representative by adopting a selling orientation exhibited a decrease in rating accuracy.

A second perspective on the underlying mechanism behind interactivity comes from research on interview structure, and specifically on components such as probing that rely on direct interaction. A lack of interactivity does prevent probing or follow-up questions. The question of whether probing is a curse or blessing for interview outcomes is rather new in research on traditional employment interviews. Schwab and Heneman (1969) found in a sample of eighteen experienced interviewers that interrater agreement was higher when interviewers were not allowed to ask any follow-up questions, although these results may have been influenced by other structural elements, suggesting that including an interaction term might increase psychometric properties. Levashina and Campion (2007) found in a sample of undergraduate students that probing increases faking behavior in interviews regardless of the question type used. However, although these two studies found negative

71

effects of probing, Levashina et al. (2014) conclude that the impact of probing on employment interviews is still unclear due to a lack of studies investigating the level of probing and the purpose of follow-up questions. In theory, probing is suggested to provide interviewers with opportunities to go deeper into candidates' responses and seek additional explanations, clarifications, and justifications. Thus, Levashina et al. (2014) suggest that only planned and limited probing enhances validity and candidate reactions, while unplanned and unlimited probing and no probing do not.

Next, interactivity allows the interview to become a real dyadic process in which the interviewer and interviewee react to each other's actions. Like probing, this dyadic process might be beneficial or harmful for the interview. Interviews usually start with a rapport building phase (Chapman & Zweig, 2005). As most interviews are between two strangers who have never previously met, individuals may have feelings of ambiguity due to their lack of information about their social interaction partner. Initial interactions during the rapport building phase are important for adjusting behaviors in the direction of the desired outcomes (Swider et al., 2016). Swider and colleagues showed in an undergraduate sample that interviewees who believed they were seen as less suitable during the initial rapport building phase were more effective in using impression management tactics. Another study found that successful applicants adapt to the interviewer's communication style, and successful candidates react to the interviewer and exhibit dominant behaviors (Tullar, 1989). Thus, it is suspected that interactivity causes interviews to become dominated by the interviewer, leading to more inaccurate ratings. On the other hand, only in interactive settings are interviewers able to apply information-seeking strategies such as planned probing and thus revise their first impression bias[18].

---

[18] A recent study found that almost one-third of interviewers make their decision within the first five minutes (Frieder, Van Iddekinge, & Raymark, 2016)

Conclusion. In summary, research has not yet demonstrated whether more or less interactivity is beneficial or harmful for the reliability or validity of interviews. Some evidence suggests that less interactivity increases reliability due to less cognitive load. Potosky (2008) proposed that reliability measures and situational control are higher when interactivity can be dynamically adjusted. Reactions should definitely suffer because two-way interaction and interpersonal treatment are seen as valuable procedural justice rules in selection processes (Gilliland, 1993).

Table 4.1. *Overview of media attributes and prior research on selection outcomes*

Media Attribute	Definition	Relevant theories	Reliability	Validity	Applicant reactions	Adverse impact
Transparency	Extent to which a medium facilitates a clear or unobstructed communication exchange (Potosky, 2008, p. 634)	Attribution theory (Kelley, 1973); cognitive load (Sweller, 1988); media richness theory (Daft & Lengel, 1986); media synchronicity theory (Dennis et al., 2008); Potosky (2008); social presence (Short et al., 1976)	Proposition 2a from Potosky (2008) proposes a positive transparency reliability relationship - No empirical evidence found	Rater correction processes due to videotechnology usage (Chapman & Rowe, 2001)	Meta-analytic evidence: Lower applicant reaction for technology-mendiated vs. FtF interviews (Blackman et al., 2016)	Media low on transparecy filters out negative visual cues (Straus et al., 2001)
Social bandwidth	The number of cues that a medium can transmit (Potosky, 2008, p. 636)	Cognitive load (Sweller, 1988); dynamic theory of media adaption (Barry & Fulmer, 2004); Gilliland (1993); media richness theory (Daft & Lengel, 1986); media synchronicity theory (Dennis et al., 2008); Potosky (2008); social impact (Latane, 1981); social presence (Short et al., 1976); thin slices of personality (Ambaddy, 1991)	Proposition 4b from Potosky (2008) proposes a negative social bandwidth reliability relationship; social cues are important to make accurate judgements (Burnett & Motowidlo, 1998)	More accurate interview ratings for FtF vs. TI (Blackman, 2002); SJTs: higher validity for video versus text based assessments (Funke & Schuler, 1998; Lievens et al., 2015; Lievens & Sackett, 2016; Christian et al., 2010); visual and vocal cues provide information above content information only (Burnett & Motowidlo, 1998)	Lower process favorablities for less social rich selection formats (Bauer et al., 2004)	Lower subgroup differences for video-based vs. text-based SJTs (Chan & Schmitt, 1997); Media low on social bandwidth filters out negative visual cues (Straus et al., 2001)
Interactivity	Pace of mutual, reciprocal, and immediate feedback in a communication process of two or more parties (Potosky, 2008, p. 636)	Cognitive load (Sweller, 1988); dynamic theory of media adaption (Barry & Fulmer, 2004); hyperpersonal communication (Walther, 1996); Gilliland (1993); media richness theory (Daft & Lengel, 1986); media synchronicity theory (Dennis et al., 2008); Potosky (2008); social impact (Latane, 1981); social presence (Short et al., 1976)	Higher reliabilities without promting (Schwab & Heneman, 1969); Higher reliabilities for videotaped vs. FtF interviews (Van Iddekinge et al., 2006)	Higher construct validities for videotaped vs. FtF interviews (Van Iddekinge et al., 2006)	Lower process favorablities for less interactive selection devices (Bauer et al., 2004); Meta-analysis evidence (Anderson et al., 2010; Hausknecht et al., 2004)	higher process favorability for more interactive selection devices (e.g., Anderson et al., 2010)
Surveillance	Extent to which an outside party could monitor or intercept the message carried by a medium (Potosky, 2008, p. 637)	Dynamic theory of media adaption (Barry & Fulmer, 2004); Potosky (2008);	Proposition 3b from Potosky (2008) proposes a positive surveillance reliability relationship; bogus pipeline is increasing reliability and validity (Roese &	Differences in proctored vs. unproctored testing (Hough, 1998; Landers et al., 2011)	Information privacy concernce decrease procedural justice perception (Bauer et al., 2004)	Improper documentation as legal risk in selection (Williams , Schaffer & Ellis, 2013)

Parallelism	The extent to which signals from multiple senders can be transmitted over the medium simultaneously (Dennis et al., 2008, p. 585)	Media synchronicity theory (Dennis et al., 2008); multiple addressability (Rice, 1987); the mathematical model of communication (Shannon and Weaver, 1947)	Higher reliabilities for panel (multiple raters) versus separate interview (Huffcutt et al., 2013; Thorsteinson, 2018); multiple raters increase interrater reliability (Schmidt & Zimmerman, 2004; Zimmerman, Triana & Barrick, 2010)	multiple raters increase criterion-related validity (Schmidt & Zimmerman, 2004; Zimmerman, Triana & Barrick, 2010)	-	The use of multiple interviews is supposed to reduce adverse impact (e.g., Moscoso, 2000)
Rehearsability	Extent to which the media enables the sender to rehearse or fine tune a message during encoding, before sending (Dennis et al., 2008, p. 587)	Media synchronicity theory (Dennis et al., 2008); multiple addressability (Rice, 1987); Impression Management (Leary & Kowalski, 1990); Gilliland (1993)	Differences in proctored vs. unproctored tests (Tippins, 2015)	Faking decreases the validity of selection procedures (Donovan, Dwight, & Schneider, 2014); Small but consistent meta-analyses test score improvement for cognitive-ability coaching programs (Sackett, Burris &	Higher process favorabilities for procedures with higher chance to perform and information known (e.g., Anderson et al., 2010, Hausknecht et al., 2004)	-
Reprocessability	Extent to which the medium enables a message to be reexamined or processed again (Dennis et al.,	Media synchronicity theory (Dennis et al., 2008); multiple addressability (Rice, 1987)	Small improvments of observer ratings (Ryan et al., 1995)	-	Misuse of recording could raise information privacy concernces and decrease procedural justice perception (Bauer et al., 2004)	Improper documentation as legal risk in selection (Williams , Schaffer & Ellis, 2013)

Surveillance

Definition. Digital infrastructure readily facilitates new methods of control, coordination, and collaboration on activities. Potosky (2008) suggested that the perceived level of surveillance among candidates would lead to an increase in situational control and reliability by preventing cheating. In fact, this argument reminds us of the bogus pipeline paradigm (Jones & Sigall, 1971), in which candidates believe that cheating will be discovered by the technology. For example, they might believe that algorithms incorporated into a web-based platform for asynchronous video interviews can detect cheating based on micro-expressions (Chamorro-Premuzic et al., 2016).

Prior research. Candidates in recruitment and selection processes have good reasons to engage in impression management and faking behaviors as a means of achieving their desired outcomes (Leary & Kowalski, 1990). Thus, applicants frequently engage in both impression management and deceptive faking behavior in employment interviews (Bolino, Kacmar, Turnley, & Gilstrap, 2008; Ellis et al., 2002; Levashina & Campion, 2007). In line with meta-analytic research showing that the bogus pipeline increases reliability and validity (Roese & Jamieson, 1993), studies in the selection context have found that warning systems are able to reduce faking behaviors (Hough, 1998; Landers, Sackett, & Tuzinski, 2011). Based on these prior research results, it is assumed that perceived surveillance (e.g. the presence of an algorithm-based "lie detector") will reduce faking behaviors and thus will increase reliability, as deceptive impression management tactics are considered a source of bias in interview scores.[19]

With respect to applicant reactions, surveillance is expected to raise privacy concerns. As shown by Bauer et al. (2006), privacy concerns are negatively related to procedural justice perceptions and organizational attractiveness. Technologies that preserve personal data such

---

[19] In fact, whether and to what degree IM tactics should actually be considered bias is discussed very controversially by Levashina et al. (2014).

as recordings could particularly give rise to high privacy concerns. For example, candidates might be afraid that a segment of their asynchronous video interviews could be abusively made accessible online for unauthorized viewers. Thus, it is not surprising that asynchronous video interviews very high on surveillance received higher ratings on privacy concerns compared to interviews conducted via Skype (Langer et al., 2017). In addition, according to the socio-material perspective, multiple data leaks by large Internet companies and stronger regulations like the European General Data Protection Regulation (GDPR) might increase candidates' awareness of data security concerns and alter their behavior in technology-enhanced employment interview formats, e.g., they might disclose less job-critical information.

Conclusion. In summary, there is some evidence that an increased level of surveillance increases reliability and validity due to a potential reduction in faking behavior, which can lead to organizations' hiring of unqualified candidates. On the other hand, it is proposed that surveillance negatively affects applicants' reactions because it can be seen as an invasion of privacy.

Parallelism

Definition. Parallelism refers to the number of parallel existing conversations. It is more limited in face-to-face employment interviews compared to videoconferencing or digital interviewing in which multiple raters can be included in the evaluation process. The critical question is: Are these additional eyes better than interviews conducted by only one person?

Previous Research. With respect to reliability, more raters have been found to produce more reliable results. For example, Huffcutt et al.'s (2013) meta-analysis found that mean interrater reliability was considerably higher in panel interviews than in separate interviews conducted by different interviewers (.74 vs. .44). Controversially, an older meta-analysis by McDaniel et al. (1994) found that interviews conducted by only one interviewer were – independent of structure – higher in validity compared to panel interviews. Schmidt and

Zimmerman (2004) found a positive relationship between interview validity and the number of interviewers. For instance, parallel interviewers are considered better than serial one-on-one interviews within the same organization as candidates cannot adjust their answers to the characters of the individual interviewers and interviewers cannot rely on first impressions if they have to justify their judgments to the other interviewers (Camp, Schulz, Vielhaber, & Wagner-Marsh, 2004).

A curvilinear relationship has been proposed with respect to applicant reactions, as candidates might appreciate the involvement of more than one interviewer and have increased confidence in the interview's procedural justice. The involvement of multiple interviewers might send a message to the candidate about the importance of teamwork in the organization, signify that the organization values the candidate's application and participation in the interview by increasing the time investment, or represent a form of early socialization by providing candidate with the opportunity to get to know more members of the team (Camp et al., 2004). In a survey conducted among HR professionals, Camp et al. (2004) found a slight preference for board compared to individual interviews from the applicants' point of view, even though board interviews are considered more stressful due to the interactions with several interviewers. Nevertheless, involving too many people – e.g., all of a company's employees – in the interview or rating process, which is very easy to implement from technical point of view, might raise massive privacy concerns, diminishing a candidate's perception of the interview process.

Conclusion. It is first proposed that parallelism increases reliability and validity, as the involvement of multiple raters reduces random error. Next, it is suggested that applicant reactions will have a curvilinear relationship with parallelism because they perceive benefits from the involvement of more persons up to a certain point but feel threatened when the number of people involved becomes too high.

Rehearsability

Definition. *Rehearsability* describes the ability to fine-tune a message before sending it. In media synchronicity theory (Dennis et al., 2008; Dennis & Valacich, 1999), the level of rehearsability refers the availability of a certain time frame that can be used to fine-tune responses. An interview is rehearsable when applicants are given enough time to prepare their answers to questions or repeatedly listen to recorded questions before submitting their answers to the interviewer.

Prior Research. As this issue only arose with the advent of asynchronous video interview platforms, direct research is non-existing. Nevertheless, fine-tuning a video recording occurs whenever someone gets help in generating responses (e.g., self-coaching, test-beating literature) or uses impression management techniques. Thus, impression management and coaching offer two underlying frameworks for explaining differences in selection outcomes due to the level of rehearsability. Previous research on practice, tutoring, and coaching has classified techniques that can be used to improve interview performance or test scores (T. J. Maurer et al., 1998). Practice involves taking alternative versions of a test under standardized conditions and learning from that experience without explicit supervision. Tutoring refers to a technique for improving performance that involves receiving direct instruction, while coaching is a broader category that encompasses many tactics (T. J. Maurer et al., 1998).

There are several tactics that can be applied during preparation time in asynchronous video interviews. First, almost every interview tips booklet highlights the importance of 'thinking before you speak', which means that applicants should practice, review, and modify their answers mentally before providing their response to interviewers. Second, as interviews and other assessment situations are perceived as stressful, the preparation time could be used to reduce anxiety (affect-focused coping) or enhance motivation. Third, Cuddy, Wilmuth, Yap, and Carney (2015) found that even very simple manipulations of body position ("power

posing") significantly increased interview performance in a mock interview setting. These or related tactics might also be used during preparation time. Fourth, as applicants are highly motivated to create and maintain a desirable impression, impression management (IM) tactics are frequently used in employment interviews (Levashina & Campion, 2007). The effectiveness of such tactics might be related to the amount of time available to fine-tune a specific impression management strategy. For example, candidates frequently invent, construct or borrow answers (e.g., describe behavior of others) to behavioral interview questions targeting previous behavior in order to create a desirable image. A higher amount of preparation time can be assumed to be advantageous for activating relevant content within episodic memory and making "borrowed" answers sound more credible. Finally, preparation time might be used for faking or cheating behavior like searching the web for an optimal answer or having someone at one's side for assistance who is not visible on the webcam.

Whether and to what extent specific response distortions or faking behaviors harm the utility of a selection system is still discussed controversially. Recently, one study found that faking behavior in selection tests was related to lower job performance, indicating that faking actually decreases the validity of selection procedures (Donovan, Dwight, & Schneider, 2014). Furthermore, a meta-analysis on impression management in employment interviews revealed that self-presentation tactics in interviews are related to interview scores but not to actual job performance (Barrick et al., 2009). However, other researchers state that faking or response distortion does not affect the psychometric properties of a selection procedure (Smith & Ellingson, 2002; Smith, Hanges, & Dickson, 2001). Thus, it is not clear whether there is a direct link to media attributes that provide applicants with greater opportunities to optimize their selection procedure performance.

Conclusion. In summary, rehearsability should increase irrelevant variance in interview scores; when candidates apply IM tactics or even faking behavior, ratings will be inflated and reliability and validity will decrease. Applicants' reactions should be more positive because

they ought to experience more confidence and self-efficacy due to a higher perceived opportunity to perform.

Reprocessability

Definition. Reprocessability refers to the ability to reprocess something and is therefore equivalent to the structural component of recording an interview identified by Levashina et al. (2014), who suggest that recordings increase interviewers' accountability with respect to following the prescribed structure and hence increase validity. Levashina et al. (2014) further proposed three levels of recordings with various media formats (video with audio, audio only, and no recording at all). However, as this proposition exhibits some overlap with social bandwidth and is also incomplete, as a detailed written protocol would also be a sort of recording, it is of limited usefulness for the proposed working model.

Prior Research. With respect to interviews, the key question is whether recordings have been found to be helpful in affecting selection outcomes. Ryan et al. (1995) found only small differences in the accuracy of observer ratings in a sample psychology students when the observers could rewind tapes from an assessment center exercise, indicating that greater reprocessability though rewinding has only limited benefits in comparison to direct observations and indirect observations without rewinding (less reprocessable). Nevertheless, no direct research on the impact of video-recording on validity and applicant reactions exists. The mechanism described by Levashina et al. (2014) should also theoretically be able to reduce adverse impact, as organizers and interviewers would be held more accountable in the presence of video evidence. Previous research has found that improper human resource documentation is a major legal risk for employers (Williams, Schaffer, & Ellis, 2013). From the applicant's point of view, recording interviews might raise privacy concerns and thus have a negative impact on reactions, as shown in the study by Bauer et al. (2006). In a very extreme case, candidates might be worried that the recordings would be leaked and made available on the Internet to unauthorized third parties. Applicants might also worry that claims

made when applying deceptive impression management tactics might be validated and corrected afterwards, which would be contrary to their goals. Thus, it is more likely that recordings will decrease applicant reactions. Finally, in accordance with socio-material theory (Orlikowski, 2007), the direction of applicants' reactions might change if usage and social norms surrounding recordings change, e.g. if video resumes, which often are sent proactively to potential employers (Behrend et al., 2012), begin to be published to draw the interest of potential employers.

Conclusion. Due to higher potential accountability for interviewers, reliability, validity, and adverse impact are likely to increase when interviews are reprocessable. On the other hand, applicant reactions might be negatively affected, as there are more potential harms than benefits for applicants.

Table 4.2. *Hypothezided effects of media attributes on reliability, validity, and applicant reactions*

	Reliability	Validity	Applicant reactions	Adverse impact
Transparency	+	+	+	-
Social Bandwidth		+	+	+
Interactivity	-		+	
Surveillance			-	-
Parallelism	+	+		-
Rehearsability	-	-	+	
Reprocessabilty	+	+		-

Note. "+" means positive effect and "-" means negative effect (for advers impact, negative effects mean reduced adverse impact).

In summary, examining media attributes offers a fruitful foundation for developing testable hypotheses concerning the impact of specific technological features in assessment devices on psychometric properties. Table 4.1summarizes the existing evidence with respect to the media attributes included in the framework and Table 4.2 summarizes the expected relationships between each media attribute and selection outcomes. Nevertheless, the media

attributes perspective also has its blind spots. While Potosky noted that users can change and control some of these attributes, beyond this meso-level user environment, the media attributes perspective fails to consider influences of the macro-environment or interactions between technologies and users, as is described in socio-material theory. The theory of technology acceptance could also offer valuable insights into the process of technology adaption beyond media attributes, e.g., by helping us understand reactions when new technology is introduced and implemented in organizations.

4.4. Research questions

This section develops testable research questions that are used to inform five empirical studies on the selection technology of asynchronous video interviews for personnel selection, the results of which are reported in the next chapter. In asynchronous video interviews, candidates record their answers to tailored interview questions on web-based platforms. Afterwards, one or more representatives of the hiring organization evaluate the answer segments with a temporal delay in order to make a selection decision (Brenner et al., 2016; Gorman et al., 2018).

Although the majority of new software applications in the recruiting and selection domain are designed in a modular manner, prior research approaches have almost exclusively been holistically driven (e.g., Blackman, 2002; Langer et al., 2017; Sears et al., 2013; Straus et al., 2001). Existing conceptual frameworks criticize the limitations of the holistic, all-in-one perspective without considering variation in media attributes within a single class of assessment procedures:

> *"In assessment exchanges, some telephone interviews may be more similar to face-to-face interviews than they are to other telephone interviews, some computerized tests may be more similar to paper-and-pencil tests than they are to other computerized tests, and so forth"* (Potosky, 2008, p. 643).

Consequently, the following studies assess *specific aspects* of asynchronous video interviewing taking the modular media attribute perspective as their underlying framework, which enables them to develop specific testable hypotheses and gain more knowledge into how and why selection outcomes change depending on variations in different media attributes. As the working model of media attributes presented above provides dozens of testable hypotheses, the following five studies focus on a selection of research questions of great practical and theoretical relevance for the relatively new field of asynchronous video interviews, without claiming to be exhaustive. Specifically, the five studies presented below investigated (1) the interrater agreement and accuracy of asynchronous video interviews and its relationship to the level of evaluation standardization, (2) the impact of different levels of social bandwidth in the response format on rating accuracy and its relationship to specific constructs typically assessed in employment interviews, (3) the congruent and discriminant validity of dimensional ratings between asynchronous video interviews, face-to-face interviews and behavioral samples from assessment centers, including their predictive validity, (4) the impact of personalization as one important aspect of social bandwidth in recorded non-interactive response formats on applicant reactions, and (5) the impact of rehearsability on the inflation of interview ratings based on asynchronous video interviews.

4.4.1. Interrater agreement and accuracy of asynchronous video interviews

Reliability sets the upper limit on the validity of any selection procedure. Relying on the modular framework of media attributes presented above, and specifically information processing theories, e.g., cognitive load (Sweller, 1994), it is proposed that relatively low levels of interactivity lead to a one-task setting in which raters can focus exclusively on judging a candidate's interview performance. Based on the perspectives presented above, it is far from trivial to assume that asynchronous video interviews will either exceed or lag behind traditional face-to-face interviews in terms of their psychometric properties. Thus, the first goal of Study 1 is to investigate the level of interrater agreement and accuracy in a typical

asynchronous video interview setting. Asynchronous video interview formats are typically

used in earlier stages of multiple-hurdle selection processes, where the usage of cut-off scores

is quite common (Sackett, Dahlke, et al., 2017). Thus, Study 1 incorporates a design

optimized for investigating the extent to which different raters achieve the same rating scores

rather than rank stability.

Research Question 1.1. Do ratings based on asynchronous video interviews achieve a

similar level of interrater agreement and accuracy as usually reported for traditional

interviews?

In addition, research on traditional employment interviews has found the level of

evaluation standardization to be an important source of rating quality (Campion et al., 1997;

Chapman & Zweig, 2005). Behaviorally anchored rating scales provide observable

illustrations of behavioral scale points to reduce ambiguity, which typically leads to better

interrater reliability, but research on performance appraisals is not unambiguous about their

value compared to simpler scales (Campion et al., 1997). Moreover, a recent study found that

frame-of-reference training and the provision of behaviorally anchored rating scales, which

each have a positive effect on rating accuracy in interviews individually, offer no additional

benefits when combined (Melchers, Lienhardt, Von Aarburg, & Kleinmann, 2011). While

Levashina et al. (2014) introduced recordings as a potential new component of structure,

which, as previously mentioned, is supposed to reduce cognitive load, there is no evidence yet

on the potential additional benefit of behaviorally anchored rating scales when (recorded)

asynchronous video interviews are used. As Melchers et al. (2011) found that the combination

of two structural elements for improving rater quality does not have necessarily a multiplying

effect, the structural complexity was varied in the study design. Specifically, given the

extensive effort involved in creating behaviorally anchored rating scales, the second goal of

Study 1 was to explore whether they have an additional benefit on interrater agreement in

AVI. Since the use of behaviorally anchored rating scales varies in applied settings, two forms

of standardized asynchronous video interviews were used (with and without behaviorally anchored rating scales).

Research Question 1.2. Can the finding that interrater agreement and accuracy are contingent on interview structure (in our case, the availability of behaviorally anchored rating scales) be replicated for asynchronous video interviews?

4.4.2. Rating accuracy of AVI and media format

The underlying assumption behind using recorded asynchronous video interview formats rather than asynchronous audio interviews is that a video-based format must have significant benefits that are independent of personal interaction compared to audio-only formats or even written text. Otherwise, it would be more convenient for both applicants and probably recruiters as well to rely on audio-only recordings, as candidates would not have to pay attention to factors such as proper dress code or lighting.

Based on the framework presented above, the justification for asking candidates to record both audio and video of their responses to interview questions is the proposed relationship between the level of social bandwidth and interview outcomes. As reviewed above, the verbal and vocal cues displayed by interviewees have been found to have important effects on interviewers' judgments and to be correlated to interview outcomes (e.g., DeGroot & Gooty, 2009; DeGroot & Kluemper, 2007; DeGroot & Motowidlo, 1999; Imada & Hakel, 1977; Parton et al., 2002). Two studies using video-recorded interviews with managers found that a variety of vocal cues (e.g., pitch, pitch variability, speech rate, pauses) and visual cues (e.g., physical attractiveness, smiling, gaze) with correlated with performance ratings provided by supervisors, supporting the importance of such cues for interview validity (Burnett & Motowidlo, 1998; Motowidlo & Burnett, 1995).

Nevertheless, in the context of technology-enhanced employment interviews, it has also been hypothesized that the presence of more cues is a double-edged sword. Information transmitted via visual cues are often not job-relevant and thus harmful with respect to adverse

impact or bias due to attractiveness stereotypes (Blacksmith et al., 2016; Potosky, 2008). Support for this perspective was provided by Straus et al. (2001), who found in a hypothetical setting with MBA students that less physically attractive applicants benefitted from lower social bandwidth in that they achieved higher scores in telephone compared to face-to-face or synchronous video interviews. Nevertheless, these results have never been replicated and a more nuanced investigation is needed to enhance our understanding of the impact of social bandwidth on interview ratings. Specifically, a potential moderator of the positive vs. harmful effects of social bandwidth is the construct domain assessed in the interview.

The employment interview construct literature suggests that interview ratings reflect various sources of variance (Harris, 1999; Huffcutt, 2011; Huffcutt et al., 2001). Huffcutt (2011) proposed that interview ratings represent variation in general traits, experience, core job elements (largely declarative knowledge), personal characteristics, and social effectiveness as well as personal and contextual factors. While general traits, experimental factors, and core job elements are considered *job-related* interview content, personal characteristics like demographic variables and impression management tactics are considered to be largely *job-irrelevant* sources of variance (Huffcutt, 2011). Previous research has shown that both personal characteristics, e.g., candidates' weight, and impression management tactics significantly affect interview ratings (Bartels & Nordstrom, 2013; C. C. Chen, Wen-Fen Yang, & Lin, 2010; Y.-Y. Chen & Fang, 2008; Roulin, Bangerter, & Levashina, 2015; Tsai, Huang, Wu, & Lo, 2010). For example, attractiveness and interview ratings were found to be meta-analytically correlated with an average correlation coefficient of .23, while impression management and interview ratings had a correlation coefficient of up to .26 (Huffcutt, 2011).

Investigating the effects of social bandwidth on interview ratings while simultaneously considering the construct domain suggests that the effect of the absence of specific cues differs according to the construct being assessed. In fact, research on thin slices of behavior reveals that rater accuracy and thus also criterion validity is very sensitive to the construct

domain that is being assessed (Ambady et al., 2006; Ambady & Rosenthal, 1993; Borkenau et al., 2004; John & Robins, 1993; Kenny, Horner, Kashy, & Chu, 1992). For instance, observable variables such as personality traits are typically judged more accurately than non-observable and task-related variables, including cognitive abilities (Ambady et al., 2000).

The literature provides at least two perspectives to explain differences in rater accuracy with respect to different construct domains. First, construct domains differ from one another in their level of *subjective visibility*. For instance, Funder and Dobroth (1987) reported the highest judgment accuracy for extraversion, which manifests in directly observable behavior, compared to neuroticism as a less observable trait, providing support for the notion that significant differences in perceived visibility exist between traits. In summary, meta-analyses and review articles have likewise concluded that traits associated with more observable behavior exhibit higher interrater agreement and accuracy than those associated with less observable behavior (Ambady et al., 2000; Funder & Colvin, 1988; John & Robins, 1993). Second, construct domains differ from one another in the *availability of implicit theories*. Borkenau et al. (2004) argued that shared stereotypes and implicit theories among raters contribute to interrater agreement. For example, a friendly smile might help two different raters make a judgment on a candidate's communication skills if they share the same implicit theory in which smiling is considered a valid cue for the aforementioned construct dimension. On the other hand, Frauendorfer and Mast (2015) found that inferred characteristics do not necessarily match applicants' real characteristics, which indicates that shared stereotypes do not necessarily possess functional validity. In summary, prior research provides strong evidence in favor of considering the construct domain as an important factor in analyzing media effects from the media attribute perspective, specifically with respect to the attribute of social bandwidth, which either transmits or filters out specific cues.

Taking up Huffcutt et al.'s (2001) construct framework, Study 2 focuses on knowledge, skills, applied social skills, and personality.[20] *Knowledge and skills* refers to a candidate's accumulated knowledge from prior experience, education, training, declarative knowledge, procedural skills and abilities (Huffcutt, 2011). Imada and Hakel (1977) reported in a laboratory study that the impact of nonverbal communication on interview ratings is lower for formal qualifications compared to personality traits, indicating that knowledge assessments marginally depend on media rich in social bandwidth. A meta-analysis including 26 primary studies found that people do not share implicit theories about the relationship between visual and vocal cues and the construct domains of knowledge and analytical thinking (Ambady et al., 2000). Thus, assessments of knowledge and skills in employment interviews should be less sensitive to lower social bandwidth compared to social skills and personality. Transferring previous research on thin slices of personality to the employment interview field (Ambady et al., 2000; Borkenau et al., 2004; Frauendorfer & Mast, 2015), it is predicted that a larger number of nonverbal and vocal cues, e.g., the availability of nonverbal communication and gestures, does not provide any additional useful information in construct domains for which no shared implicit theory is available. This is true for job-relevant knowledge and skills. Additional cues might even lead to rating biases, as the lack of a shared implicit theory could lead multiple interpretations of these cues, reducing rating accuracy.

Research Question 2.1. Will an increasing level of social bandwidth decrease rater accuracy for construct dimensions of knowledge and skills?

Applied social skills have been defined as a candidate's ability to apply effective behavior in social situations such as acting appropriately in a team (Huffcutt et al., 2001). Burnett and Motowidlo (1998) compared ratings of video-recorded interviews in different media conditions and found different correlations with nonverbal cues (e.g., gaze, smile, hand

---

[20] Christian et al. (2010) identified the three categories of knowledge and skills, applied social skills and basic personality tendencies as the constructs typically assessed in situational judgment tests.

gestures, body posture) depending on the construct dimension of managerial effectiveness being assessed. For example, significant correlations were reported between gaze and the dimensions of leadership and teamwork, but not the dimensions of drive and planning/organization (Burnett & Motowidlo, 1998). Previous research found that dimensions that are important for social communication are rated particularly less favorably in media low in social bandwidth (Blackman, 2002). Thus, in line with previous research, it is suggested that accurate judgments of applied social skills benefit from the availability of nonverbal and vocal cues and therefore from the use of technology high in social bandwidth.

Fundamental personality traits are frequently assessed in employment interviews, with conscientiousness the most frequently assessed among the Big Five (Huffcutt et al., 2001). Conscientiousness measured using various methods has been found to a valid predictor of job performance (Barrick, Mount, & Judge, 2001). Assessments of personality traits and applied social skills involve paying some degree of attention to nonverbal behavior, such as eye contact or smiling (Blackman, 2002; Motowidlo et al., 1992). When additional cues are meaningful for a given construct domain due to the availability of implicit theories, both visual and vocal cues can provide additional valid information. This is the case for the constructs of applied social skills and personality traits. In other words, the observability of these kinds of traits is rather high, making it easier to detect confirming and disconfirming behaviors (Funder & Dobroth, 1987; John & Robins, 1993). Overall, additional cues *enrich* the evaluation process and should therefore support the generation of shared interpretations among independent raters and therefore increase the quality of ratings.

Research Question 2.2. Will an increasing level of social bandwidth increase rater accuracy for construct dimensions associated with applied social skills and conscientiousness?

In summary, Study 2 was designed as an initial investigation of the interplay between social bandwidth and various construct domains on interviewer accuracy in a randomized

experimental study in a simulated highly-structured employment interview setting. In accordance with our research questions, social bandwidth was proposed to increase rating accuracy for the construct domains of applied social skills and personality traits but decrease rating accuracy for knowledge and technical skills.

This research is important as it contributes to the interview literature (1) by being (to our knowledge) the first systematic empirical test of the interplay between media format and construct domain in highly-structured digital interviews. (2) It provides an additional primary study on interview outcome equivalence between different media formats in a field in which existing research is somewhat contradictory, thus contributing to the further clarification of the issue (Blacksmith et al., 2016).

Moreover, the study is deeply relevant for selection and assessment practitioners' understanding of differences related to media usage and construct domains for the following reason: In selection practice, multistage selection systems often focus on different construct domains in different stages. Early screening interviews often focus on knowledge and skills, while the final selection stage focuses more on personality and applied social skills, potentially leading to different impacts of media usage in different stages. Therefore, it is important to understand potential interactions between media usage and the construct domain assessed in a job interview.

4.4.3. Validity of AVI in preselection

Personnel assessment methods are designed to predict future behavior, most importantly job and training performance (Schmidt & Hunter, 1998). Validity refers to the extent to which test scores are related to criteria such as job performance. In multi-hurdle selection, the additional proportion of variance explained by a given predictor method over and above other assessment instruments is referred to as incremental validity. While some studies on the incremental validity of low-fidelity simulations have been published in major journals in the field (Lievens et al., 2015; Lievens & Patterson, 2011), research on the incremental validity of

asynchronous video interviews as an alternative predictor method does not yet exist and therefore represents a notable gap in the existing literature.

As previously noted, the validity of traditional employment interviews is well established, and multiple meta-analytic works have identified relevant boundary conditions and moderator variables (Huffcutt & Arthur, 1994; Huffcutt et al., 2014; McDaniel et al., 1994; Thorsteinson, 2018). Typically, meta-analyses report average correlations between structured interviews and job performance corrected for range restriction and reliability that are greater than .50 (Huffcutt et al., 2014; Schmidt & Hunter, 1998; Thorsteinson, 2018). The most recent meta-analysis by Thorsteinson (2018) also reports uncorrected average correlations for a range of moderators, including separate vs. panel interviews (.24 vs. .31) and type of interview, specifically situational interviews (.24), behavioral interviews (.32), and other forms (.24). With respect to the criterion domain, correlations with job performance (.25) and training performance (.22) as well as five levels of question standardization (.20 - .31), three levels of response standardization (.19 - .30), and three levels of overall structure (.21 - .30) were reported. In addition, Thorsteinson's (2018) meta-analytic study was the first to report on media format as a moderator, with uncorrected average correlation coefficients of .24 for face-to-face interviews, .32 for video and audio, .22 for audio only and .57 (k =2) for mixed-media formats. The media format referred to how the interview was presented to judges; for example, face-to-face interviews that were video-recorded and presented in this format were coded as video interviews (Thorsteinson, 2018). Moreover, interviews can measure different constructs; meta-analytic construct validity has been reported for general intelligence (.24), knowledge and skills (.42), basic personality tendencies (.33 for extraversion, .33 for conscientiousness, .51 for agreeableness, and .47 for emotional stability), applied social skills (.26 for communication skills, .39 for interpersonal skills, .47 for leadership, and .24 for persuading and negotiation), occupational interest (.24), and finally .49 for organizational fit (Huffcutt et al., 2001, p. 906).

Structured interviews also exhibit incremental validity over a range of other classes of selection procedures, as they are correlated with such procedures at only weak to medium levels (e.g., Berry et al., 2007; Cortina et al., 2000; Huffcutt et al., 1996; Roth & Huffcutt, 2013; Salgado & Moscoso, 2002). Research has mainly paid attention to the relationship between cognitive ability and interview ratings, which Schmidt and Hunter's (1998) meta-analysis suggests is one of the best combinations of two selection procedures for predicting job performance. Understanding the cognitive saturation of a given selection method is important for modeling the effectiveness of selection systems with multiple predictors in the sense of incremental validity for the quality of hiring decisions. Furthermore, the cognitive saturation of different selection methods is considered to be the main driver for subgroup differences and hence affects whether a given alternative predictor method can be applied (Lievens et al., 2015; Ployhart & Holtz, 2008). Transferring these considerations to the use of technology for employment interviews, the incremental validity of technology-enhanced interviews would be low if the use of technology leads to high cognitive saturation. It is not trivial to assume that technology usage affects the cognitive saturation of selection procedures. For example, previous research has reported significantly lower cognitive saturation in video-based compared to text-based SJTs (Lievens & Sackett, 2006). Hence, increasing the cognitive load of technology-enhanced interviews could also be a potential source of adverse impact.

Recently, a first study on the validity of asynchronous video interviews by Gorman et al. (2018) delivered promising results with respect to the predictive validity of asynchronous video interview formats. The authors reported that dimensional ratings on mental capabilities, knowledge and skills, applied social skills, and basic personality tendencies were related to criterion measures of job performance (self-ratings) and organizational tenure. However, while this study provides encouraging first support for the general validity of asynchronous

video interviews in selection, it does not permit any statements to be made about their relationships with other assessment procedures.

With respect to the media attribute of transparency, all forms of technology are less transparent compared to face-to-face interviews. Transparency is defined as the extent to which a medium allows for the unobtrusive transmission of information; hence, any medium is a potential source of unsystematic error variance due to potential attention shifts away from the content and towards the medium as well as potential attribution effects, as demonstrated in a laboratory study on rater attribution processes by Chapman and Webster (2001). In addition, difficulties understanding what interviewees wish to say can frequently occur, as the quality of the recording depends on the candidate's hardware and the physical bandwidth or Internet connection at the time of recording. Next, social bandwidth is lower in asynchronous video interviews, which affects the transmission of a variety of nonverbal cues. Specifically, limited visibility due to the narrow focus of webcam cameras means that nonverbal cues such as extensive bodily gestures appear less pronounced, forcing observers/interviewers to rely more on the content of answers to make judgments about candidates. As previous research has found that nonverbal and vocal cues provide valuable information above and beyond content (Burnett & Motowidlo, 1998), only medium correlations between face-to-face and asynchronous video interviews can be expected. It is further suggested that interviews higher on social bandwidth possess a higher level of validity due to greater symmetry between the predictor and criterion variable in most workplaces (Jansen et al., 2013). Additional evidence for a positive relationship between social bandwidth and validity comes from research finding that open-ended written responses to an SJT were a significantly better criterion predictor for a behavior-oriented role play than the same test with a multiple-choice response format (less rich on social bandwidth; Funke & Schuler, 1998). In a meta-analysis, Lievens and Sackett (2006) assessed validity differences between written vs. video-based SJTs and found that the richer video-based format exhibited higher construct validity. Most recently, Lievens et al.

(2015) reported in a field sample that behavioral responses were found to be slightly higher in validity, more personality-saturated and less cognitively saturated, and perceived more favorably compared to less rich written responses. Therefore, the less rich format of asynchronous video interviews is expected be lower in criterion validity with respect to workplace outcomes compared to richer face-to-face interviews. As asynchronous video interviews are commonly used as a screening instrument rather than as a final selection hurdle, the next research question asks how asynchronous video interviews are related to face-to-face interviews and behavioral outcomes when the asynchronous video format is used to preselect candidates at an early stage of the overall selection process.

Next, in asynchronous interview formats, any behavior that relies on interpersonal interaction such as prompting is suppressed by default. As previous research is somewhat contradictory with respect to the benefits or drawbacks of prompting for interview outcomes (Levashina et al., 2014), no empirical findings exist on potential differences between face-to-face and asynchronous video interviews with respect to their degree of cognitive saturation. As interpersonal interactions are considered cognitive challenging tasks, which should therefore produce more cognitive load[21] (Sweller, 1994), it is suggested that the cognitive saturation of non-interactive video interviews is lower than that of face-to-face interviews. This research question can be depicted with the hypothesized model of relationships between predictor and criterion variables at different points in the multi-hurdle selection process considered in Study 3 (see *Figure 4.2*). In the study's field-based setting, cognitive ability tests and asynchronous video interviews were considered as initial predictor variables and assessment center performance and face-to-face interviews as subsequent criterion variables. Consistent with prior research on the relationship between cognitive ability and interview performance (e.g., Berry et al., 2007; Roth & Huffcutt, 2013), a significant medium-sized

---

[21] See also the theory of ego depletion as an alternative explanation (Baumeister, Bratslavsky, Muraven, & Tice, 1998)

relationship between asynchronous video interview ratings and tests of GMA is proposed.

Next, and in line with first results on the validity of AVI in a laboratory setting (Gorman et al., 2018), it is expected that this study will deliver the first field-based evidence on the criterion validity of AVI for behavioral outcomes. Moreover, this relationship should be incrementally valid over cognitive abilities. Finally, a significant positive relationship between short asynchronous screening interviews and full-length face-to-face selection interviews is predicted, while the relationship between cognitive testing and assessment center performance is expected to be in the low to medium range, in line with previous research (e.g., Sackett, Shewach, & Keiser, 2017).

Research Question 3.1. Are asynchronous video interviews predictive for behavioral outcomes in assessment centers and performance in traditional face-to-face interviews?

Research Question 3.2. Are asynchronous video interviews incrementally valid over cognitive ability testing?
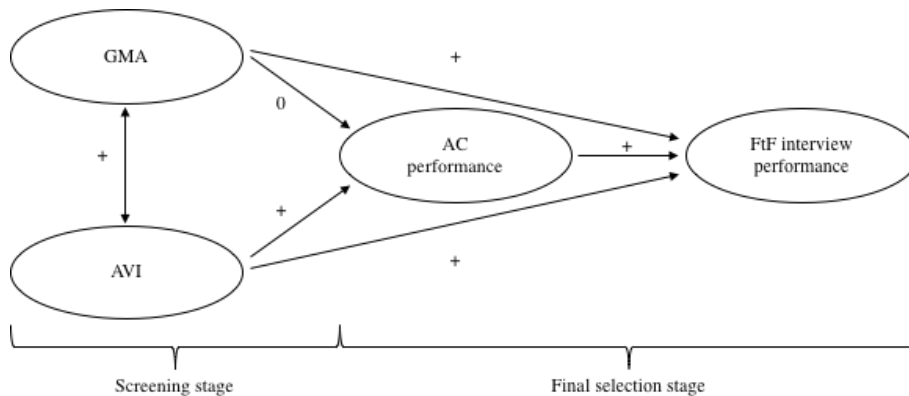
Figure 4.2. Hypothesized model of relationships between predictor and criterion variables. GMA = general mental ability, AVI = asynchronous video interviews, AC = assessment center, FtF = face-to-face.

4.4.4. Applicant reactions to AVI and the role of personalization

Research on applicant reactions to selection procedures has found a replicable hierarchy among popular selection methodologies with respect to their process favorability: (traditional

face-to-face) interviews and work sample tests are most preferred, cognitive ability tests and personality inventories are located in the middle, and honesty tests or hiring due to personal contacts represent the lower bound of the preference scale (Anderson et al., 2010; Hausknecht et al., 2004).[22] Gilliland (1993) presented one of the most popular models of applicant reactions, which suggests that procedural justice rules affect overall fairness reactions and selection outcomes. Previous research has shown that fairness perceptions are lower for asynchronous video interviews than for synchronous video interviews via Skype, are broadly comparable to those of other non-interactive procedures such as cognitive tests (Brenner et al., 2016; Langer et al., 2017). Because Gilliland's (1993) model identifies two-way communication and interpersonal warmth as two important procedural justice rules, it is not surprising that asynchronous and non-interactive assessment procedures have lower favorability ratings than interactive ones. Consequently, as elaborated upon in the review section, interactivity is an important attribute for explaining differences in candidate reactions.

Another media attribute that might be important for explaining variation in applicant reaction is social bandwidth, which is defined the extent to which a medium can transfer multiple cues simultaneously (Potosky, 2008). One aspect of social bandwidth is the extent to which a message can be made personal, which is referred to in media richness theory (Daft & Lengel, 1986) as the level of personalization. While media richness theory proposes that the richest communication is both personal and involves face-to-face interaction (Daft & Lengel, 1986), the framework used in this thesis allows a distinction to be made between "personal" in the sense of interactivity and "personalization" achieved by manipulating social bandwidth, as was done in this study. The next section elaborates in greater detail how a higher level of personalization is proposed to affect applicant reactions even in non-interactive procedures.

---

[22] Some studies still include graphology in this typology. However, a study from a decade ago showed that this technique is more myth than reality, and Internet-based recruiting system should have brought an end to this phenomenon.

It has been suggested that selection processes ought to be considered from a signaling game perspective, as candidates face incomplete information during recruiting processes (Bangerter et al., 2012). From this perspective, candidates seek out information in order to determine whether or not an organization fits their interests. One possible result of this evaluation process is continuation or withdrawal from the recruitment process. As hiring organizations have a strong interest in maintaining applicants during the selection process (Allen, Mahto, & Otondo, 2007), they should choose communication patterns that satisfy candidates' need for information even at early stages of the recruitment process, when candidate communication is not yet interactive, but still rich in social bandwidth. Research on realistic job previews[23] (RJPs) suggests that the most critical underlying element or key mechanism of the attraction process is not social bandwidth, but candidates' perception of the organization's honesty (Earnest, Allen, & Landis, 2011), and previous research on interpersonal communication has shown that honesty perceptions depend on the level of social bandwidth (Rockmann & Northcraft, 2008; Swaab, Galinsky, Medvec, & Diermeier, 2012). Applying this principle suggests that providing applicants with highly personal and honest information via rich media should be an effective lever for moving applicant reactions in a positive direction. In the Gilliland model, selection information and honesty are both considered procedural justice rules. With these concepts, Gilliland primarily refers to providing a justification for the selection decisions made, but also considers them to include a priori information provision (Gilliland, 1993, pp. 706-707). He further highlights honesty and truthfulness in communication as important procedural justice rules (Gilliland, 1993, p. 707), and the meta-analysis by Hausknecht et al. (2004) found a positive relationship between explanations/accounts and candidates' procedural justice perception. Consequently, the present study adds to our knowledge about whether or not information should be presented by

---

[23] RJPs are defined as materials, programs, presentations or videos that provide applicants with realistic and balanced (positive and negative) information about a job (Earnest, Allen, & Landis, 2011).

additionally considering the question of *how* information should be presented. Individual differences and personality have been discussed as another source of variance in explaining differences in applicant reactions. Although Hausknecht et al. (2004) found only low or near-zero correlations between personality traits and procedural justice perceptions, personality variables have been found to play an important moderating role. For example, neuroticism, or the tendency to adjust poorly to stress and be anxious and insecure, has been proposed to be related to privacy concerns (Chan & Schmitt, 2004). As honesty perception in a given situation is not independent of personality traits, a candidate's personality dispositions ought to considered as a potential moderator; previous research in the I/O domain reveals that justice perceptions and personality interact, specifically with respect to the traits of agreeableness, conscientiousness, and neuroticism (O'Neill, Lewis, & Carswell, 2011).

From a modular perspective, a lack of knowledge exists on whether specific aspects of social bandwidth, like level of personalization, can compensate for the negative effects of the lack of interactivity in asynchronous video formats. Therefore, the fourth study tested whether manipulating cues in the messaging process to make them more social and personal could lead to improvements in procedural justice perceptions in a randomized experimental study with $N = 98$ non-applicants, controlling for individual differences.

Research Question 4. Will a manipulation of personalization, as one aspect of social bandwidth, in asynchronous video interviews increase candidates' procedural justice perceptions?

4.4.5. Rating inflation in AVI

When candidates are invited to complete an asynchronous video interview, popular software platforms make it possible to customize various elements of the workflow design, including the amount of time provided to candidates before the actual recording starts. Preparation times can range from a few seconds in order to protect candidates who clicked on the record button by accident up to several minutes. It is also quite common to present

candidates with selected interview question up front, which allows for very extensive preparation and is comparable to the third level of interview transparency described by Levashina et al. (2014). However, practitioners might express concerns that providing too much preparation time or candidate foreknowledge of the interview content from leaked Internet sources might artificially inflate interview scores. Preparation times are also a clear difference between asynchronous video interviewing and face-to-face employment interviews, (synchronous) videoconference interviews conducted via platforms such as Skype, and telephone interviews. Preparation time is related to the media attribute known as rehearsability, which is defined as the ability to fine-tune a message before sending it. On the basis of research on faking in unproctored Internet test settings (e.g., Arthur & Glaze, 2011), impression management (e.g., Bolino et al., 2008), and the coaching/training literature (e.g., T. J. Maurer et al., 1998), it is suggested that preparation time will alter interview ratings. As previously reviewed, applicants might apply several strategies during preparation time to increase their performance ratings, including cheating[24], practicing, mentally reviewing and modifying their answers, borrowing answers, or engaging in emotion-based coping tactics.

In line with the assumption that response distortion and faking behaviors are negatively associated with the utility of a selection system, a previous study found that faking behavior in selection tests was related to lower job performance, meaning that faking actually decreases the validity of a selection procedure (Donovan et al., 2014). However, other researchers state that faking or response distortion does not affect the psychometric properties of a selection procedure (Smith & Ellingson, 2002; Smith et al., 2001). A study by Day and Carroll (2003) compared the validity of traditional employment interviews among candidates who did and did not have prior knowledge of the interview questions, and found that both groups' scores

---

[24]A quick Internet search reveals thousands of tips on specific favorable answers to interview questions (e.g., https://www.themuse.com/advice/how-to-answer-the-31-most-common-interview-questions, https://www.getsmarter.com/career-advice/planning/powerful-answers-to-15-common-interview-questions).

were similarly correlated with a criterion measurement; no significant differences emerged. However, no studies assessing this issue with regard to asynchronous video interviews exist at this point. For this reason, the fifth study's goal was to examine whether individuals who were given preparation time before each interview question performed better in the interview than individuals who were not granted preparation time. Hence, this study is the first to test the effect of rehearsability in a controlled randomized experimental design with $N = 51$ non-applicants. Based on the underlying frameworks on rehearsability presented above, the last dataset will test the significance of the absence or presence of preparation time on interview ratings according to the following research hypothesis:

Research Question 5. Will interview ratings in AVI be higher when candidates have more time to prepare their answers?

5.   Results

5.1.   Study 1: Interrater agreements in AVI

The purpose of Study 1 was to provide an initial investigation into the level of raters' agreement and accuracy tested in a non-applicant sample and with the manipulation of two levels of standardization of evaluation using either dimension ratings or global ratings.

5.1.1. Method

Participants

A total of 111 adults (74 women) were recruited from different university classes (in psychology and business) during regular class hours as well as through private networks. The average age was 25.4 years ($SD = 5.4$). The final sample consisted predominantly of students (91%). On average, participants reported 1.2 years ($SD = 3.8$) of work experience. Participants did not receive any financial compensation for their participation.

Design and Procedure

A one-factorial (behaviorally anchored vs. global rating scales) between-subjects experimental design was used to explore the study's research questions. To keep the stimulus material constant, a semi-professional actress acted out the role of a fictitious candidate for an entry-level white-collar position. An entry-level position with non-technical skill requirements was chosen so that no in-depth knowledge of the job was necessary on the part of the raters. Two subject-matter experts derived six competency dimensions from a sample of approx. 30 similar job postings for junior consultants – a popular entry-level position for university graduates – found on public online job boards. Four interview questions and corresponding scripted answers were developed[25], which were then pretested to ensure that the answers reflected a medium level of qualification and finally recorded with an ordinary webcam. To provide one example, the first question was *"Please introduce yourself briefly and your previous career path. Please elaborate why you are interested in the junior consultant position"*, which was rated on the dimensions *Drive* and *Communication Skills,* mirroring typical procedures in applied settings.

The resulting video material was then included in an online survey in which participants first received a short introduction to the technology and read a job description to gain familiarity with the job requirements. The participants also received some basic instructions on how to rate interviews as a low-level interview training. To examine interrater agreement across two different levels of interview structure, participants were randomly assigned to either the "behaviorally anchored score" condition or the "global rating scale" condition.

---

[25] Following a procedure that is typical in applied settings, we did not conduct a 1:1 mapping of interview questions to competency dimensions. Rather, two interview questions were scored on two dimensions each. More than one criterion is typically used when the construct domains do not overlap. For example, *teamwork* and *foreign language* are both coded on the basis of a question about previous achievements in teams when candidates provide their answers in a foreign language. In assessment centers, multiple criteria are commonly scored during or after a single exercise (Monahan, Hoffman, Lance, Jackson, & Foster, 2013).

Dimensional ratings consisted of three behavioral description anchors and one overall rating for each dimension. All anchors were developed and pretested with subject-matter experts. To provide examples for one dimension, the anchors for *Drive* were (1) "sets challenging goals that fit the job requirements", "demonstrates that he/she believes in his/her ability to achieve his/her goals", and "shows willingness to succeed and take on responsibility". Each of the three anchors as well as the overall dimensional appraisal ("How do you rate the candidates *Drive* overall?") were separately rated on a 5-point scale. These four ratings were then into averaged into one score to form the overall dimensional score. *Global ratings* required the raters to rate the overall quality of each video answer segment on a 5-point scale from 1 = *very poor* to 5 = *very good* after being shown each sequence without any instructions concerning dimensionality or detailed anchors. Finally, all participants filled out likeability, attractiveness and demographic control variables.

Analytical approach

The level of agreement between different raters was assessed using the r_{wg}(j) indicator from James, Demaree, and Wolf (1984), and the *AD* index from Burke, Finkelstein, and Dusig (1999). The within-group correlation r_{wg}(j) assesses interrater agreement over various items – in this study, j refers to interview questions – by comparing the variance obtained from multiple raters to the expected variance that would be obtained if ratings were completely random. r_{wg} can vary from 0 to 1, with greater values indicating better interrater agreement. The *AD* ("average deviation") index involves determining the extent to which each item rating differs from the mean or median item rating and summing up the absolute values of these deviations.

Rater accuracy was assessed using the *E* index, which refers to the overall discrepancy between observed and "true" scores (Sulsky & Balzer, 1988). In our design, the "true" scores were predetermined by our use of pretested answer scripts that were iteratively fitted to reflect

a medium level of qualification. All agreement indices were calculated using R for Macintosh (R Core Team, 2014).

5.1.2. Results

Table 5.1 presents the $r_{WG}(j)$ and AD for interrater agreement and E for rater accuracy. As can be seen, $r_{WG}(j)$ is higher than the common cutoff score of .70 for strong agreement (LeBreton & Senter, 2008) in both conditions (with and without behaviorally anchored rating scales). Therefore, the overall level of interrater agreement in asynchronous video interviews is sufficient. In addition, interrater agreement was higher (.90 vs. .81, $p < .05$ one-tailed, $p = .07$ two-tailed) for behaviorally anchored rating scales compared to ratings on global scales according to a simple z-test testing for significant differences between the two correlations following Eid, Gollwitzer, and Schmitt (2011, p. 542). This is consistent with findings reported for traditional interviews. Also, the AD values for both groups were below the cutoff of .80 suggested by Burke and Dunlap (2002). Please note that smaller AD values signify higher agreement. Hence, the smaller values found for behaviorally anchored rating scales compared to global scales indicate stronger agreement for the former measure.

Next, rating accuracy E was found to be 0.81. Unfortunately, there is no absolute cutoff for E. The data revealed significantly higher E scores for behaviorally anchored ratings compared to ratings on a global scale, with $E = .92$ vs. $E = .69$ [$F(1/102) = 11.57, p < .001$].

Table 5.1. *Interrater Agreement and Rater Accuracy (Study 1)*

	Total	Global Ratings	Anchored Rating Scales
		Level of Standardization	
r_{wg}(j)	.86	.81	.90
AD (median)	.69	.75	.59
E	.81	.92	.69

Note. N = 111. r_{WG} is James et al.'s (1984) within-group interrater agreement index, AD is Burke et al.'s (1999) average deviation index (with values for both reflecting average agreement over all 4 video segments). *E* refers to Cronbach (1955, in Sulsky et al., 1988).

5.1.3. Discussion

Study 1 presented first results concerning the interrater agreement and accuracy of an asynchronous video interview. Our data indicative that sufficient levels of agreement and accuracy can be achieved, even under unproctored rating conditions with untrained raters. Importantly, agreement was quite similar to values obtained for traditional interviews using the standard Pearson correlation coefficient to determine the degree of rater consensus (Huffcutt et al., 2013). It was also found that the use of behaviorally anchored rating scales further increased the level of interrater agreement and accuracy, which is consistent with previous research conducted in the interview domain (Maas, 1965; S. D. Maurer, 2002; Melchers et al., 2011).

The first conclusion to be drawn based on Study 1's findings is that there is no genuine disadvantage resulting from asynchronous video interviews. The initial evidence presented here suggests that asynchronous video interviews – especially when used in combination with

behaviorally anchored rating scales – enables HR practitioners to make decisions that are comparable to traditional face-to-face interviews.

Nevertheless, several limitations of this study should be noted. First, this study referred to a hypothetical setting rather than high-stakes selection situation, which might have influenced participants' motivation to make accurate judgments. Second, the use of scripted answers set performance variation to zero. Hence, it was not possible to calculate rank-order correlations. On the other hand, this research design has several advantages. First, iteratively anchoring the candidate's answers to a medium level of qualification allows "true" scores to be used to estimate rating accuracy (even if leniency biases might have caused our observed means to be somewhat higher than intended). Second, the use of a single rating target giving standardized holds constant all variation due to physical appearance, gender, age, or any other visible characteristic. Nevertheless, future research should aim to replicate our results in a field setting with a broader range of applicant behavior.

5.2. Study 2: Social bandwidth and rating accuracy

This section reports the results of Study 2 exploring *Research Question 2.1* and *2.2* Study 2 was an experimental study testing the impact of systematic variation in social bandwidth on rating accuracy in different construct dimensions in highly structured but non-interactive asynchronous interviews.

5.2.1. Method

Participants

This study involves 4 x 4 x 2 x 2 between-subjects design assessing which four factors: the social bandwidth manipulation (video and audio vs. audio only vs. muted video with text vs. text only), construct dimension (knowledge vs. technical skills vs. applied social skills vs. personality), answer quality (high vs. low), and question type (situational vs. behavioral description). Participants were recruited mainly from a medium-sized German university in the Berlin area, with students receiving course credits for their participation in the study, as

well as different social media channels like LinkedIn as well as and private and professional networks. A total of 281 participants completed the study. One participant had to be excluded because he was younger than 16 and an additional participant had to be excluded because they did not watch the video sequences completely, resulting in a final sample of $N = 279$. The average age was 29.51 years ($SD = 10.23$), with $M = 8.15$ years of work experience ($SD = 9.26$). 74 members (26.52%) of the sample were males, 202 (72.40%) were females, and 3 (1.08%) refused to indicate their gender. The sample mainly consisted of students (66%). 40% already had experience with technologically-supported interviews (46% as an applicant, 15% as an interviewer, and 39% as both).

The total of 279 participants were randomly distributed into the four interview conditions as follows: video/audio (original) $N = 68$, audio only $N = 66$, video/text (mute) $N = 72$, text only $N = 73$. No significant differences between the groups occurred for demographic variables, knowledge of commerce, or motivation. However, a significant difference emerged for participants' experience in rating candidates based on interviews, with $F(3, 275) = 4.092$, $p < .01$. A post-hoc pairwise comparison with Bonferroni adjustments revealed that participants in the original video/audio condition reported significantly more experience compared to participants in the audio condition, with $t(124) = 3.537$, $p < .001$. Due to this significant effect, participants' experience in rating candidates based on interviews was included as a covariate in the subsequent analysis.

Development of Stimulus Material

Job Description. The job profile of an office clerk position was used based on the assumption that the requirements for this job profile are understandable even for participants without extensive HR management experience or industry-specific knowledge. Furthermore, it was assumed that this profession is neutral with respect to job-specific gender stereotypes. The job description also included a short description of the fictitious company doing the hiring, the job tasks, and the expected qualifications.

Interview Development. Four distinct construct dimensions drawn from the interview and selection construct literature were used (Christian, Edwards, & Bradley, 2010; Huffcutt et al., 2001). Two dimensions fell under the category of knowledge and technical skills, namely *job knowledge* and *organization and planning* skills. The *teamwork* dimension represented applied social skills and *conscientiousness* represented one of the Big Five personality traits. Based on the job description, two experienced HR experts developed one situational and one biographical interview question including behavioral anchors for each of the four construct dimensions, resulting in a total of eight interview questions. Behavioral anchors were provided for the two extremes and the middle category of a 7-point rating scale. Next, two answer scripts for each question were developed, with one script representing a *true score of 2.5* (poor answer) and the other representing a *true score of 5.5* (good answer). The answer scripts were tested iteratively and optimized by three independent subject-matter experts with several years of experience in HR management. *Figure 5.1* provides an overview of the development of the materials.
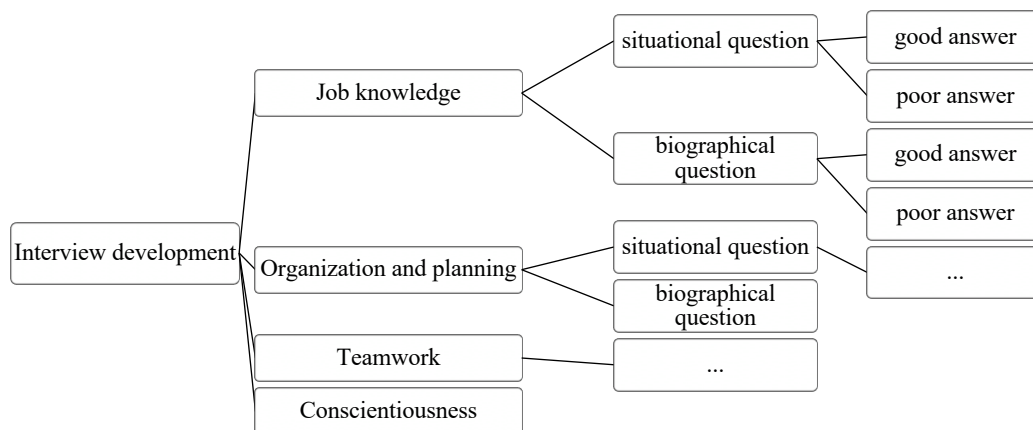
Figure 5.1. Schematic showing the development of the interview materials (Study 2)

Stimulus Material (Videos Sequences). All 16 answer scripts were video-recorded separately using the built-in camera of a Macintosh laptop and a professional desk microphone. A semi-professional actor was hired to play the role of the job candidate. The

actor was male, 37 years old, and Caucasian. In the recordings, he wore a white button-down shirt with a dark tie. A conference room was chosen as the scene for recording the scenarios. No interviewer could be heard or seen in the recordings. Afterwards, the same subject-matter experts involved in question development checked the recordings again to confirm the accuracy of the intended true score.

Manipulation of Social Bandwidth: After the recordings were completed, the final 16 video sequences were manipulated with respect to the available cues for each condition. For the video/text version, the sound was erased and subtitles were added. The subtitles were placed in the upper left corner of the screen to avoid blocking body movements or other visual cues by the actor. For the audio only version, the image track was set to black and only the actor's voice remained. For the text only version, all answers were transcribed. No modifications were made in the original video/audio version.

Procedure

This study was conducted as an online survey with a widely used software package for social research[26]. First, the job description was presented to all participants. After they had read through the description, each participant had to complete a short interview training encompassing a description of all four construct dimensions assessed in the interview, an explanation of the behavioral anchors as well as an example and explanation of an appropriate interview rating.

In a between-subjects design, participants were randomly assigned to one of the four media conditions (video and audio, audio only, muted video with text, text only) with a balanced presentation of good vs. poor answers (2x good, 2x poor, 2x good, 2x poor OR 2x poor, 2x good, 2x poor, 2x good). To avoid any order effects, the sequence of construct dimensions was randomized. However, the situational and biographical questions referring to

---

[26] https://www.soscisurvey.de/

the same construct dimension were presented in blocks. The presentation of video sequences and behavioral anchors was structured as follows: Participants saw first the question and the corresponding behavioral anchors and next the answer segment corresponding to their experimental condition. Afterwards, the behavioral anchors were presented again and they were asked to rate the participant. After the participants had completed all ratings, they were asked to rate the overall fit of the fictitious candidate for the position. Several control variables (i.e., motivation, experience in conducting and rating interviews, demographic variables) were completed last.

Variables

Interview performance. Each of the eight interview questions was rated separately on a 7-point rating scale. The overall hireability of the candidate was measured using a three-item measure with a 7-point Likert scale (Cronbach's alpha = .86). An example item for overall fit was "*I think the candidate deserves this job*."

Interrater agreement. Interrater agreement was assessed for each group with r_{wg} from James et al. (1984). The interrater within-group correlation $r_{wg}(j)$, with j in this study referring to interview questions, assesses interrater agreement over various items by comparing the variance obtained from multiple raters to the expected variance that would be obtained if ratings were completely random. r_{wg} can vary from 0 to 1, with greater values indicating better agreement. All interrater agreements are depicted in Table 5.1. *ICC* as presented by Shrout and Fleiss (1979) is not meaningful for our study design because at least two ratees are required to calculate the *ICC,* which represents a rank correlation. Interrater agreement $r_{wg}(j)$ was calculated at .86 for the original video/audio group, .85 for the audio group, and .80 for both the video/text and the text group. As shown in Table 5.1, r_{wg} was found to be at a sufficient level over according to Wagner, Rau, and Lindemann's (2010) recommendations. Hence, no significant differences in the level of r_{wg} were found among the four media groups when simple *z*-tests were calculated.

Table 5.2. *Interrater Agreement for all Interview Questions*

Dimension	Type	Quality	Video/Audio	Audio	Video/Text	Text	Overall
Job Knowledge	situational	poor	0.60	0.60	0.25	0.68	0.53
		good	0.67	0.76	0.74	0.66	0.71
	biographical	poor	0.73	0.79	0.81	0.73	0.77
		good	0.85	0.81	0.68	0.84	0.80
	overall	poor	0.80	0.82	0.69	0.83	0.79
		good	0.86	0.88	0.83	0.86	0.86
		Overall	**0.83**	**0.85**	**0.76**	**0.85**	**0.82**
Organization & Planning	situational	poor	0.54	0.75	0.62	0.55	0.62
		good	0.76	0.84	0.6	0.78	0.75
	biographical	poor	0.72	0.73	0.6	0.6	0.66
		good	0.61	0.66	0.6	0.59	0.62
	overall	poor	0.78	0.85	0.76	0.73	0.78
		good	0.81	0.86	0.75	0.81	0.81
		Overall	**0.80**	**0.86**	**0.76**	**0.77**	**0.79**
Teamwork	situational	poor	0.58	0.46	0.53	0.55	0.53
		good	0.85	0.76	0.89	0.86	0.84
	biographical	poor	0.72	0.7	0.74	0.42	0.65
		good	0.68	0.44	0.62	0.71	0.61
	overall	poor	0.79	0.73	0.78	0.65	0.74
		good	0.87	0.75	0.86	0.88	0.84
		Overall	**0.83**	**0.74**	**0.82**	**0.77**	**0.79**
Conscientiousness	situational	poor	0.62	0.72	0.59	0.41	0.59
		good	0.69	0.78	0.71	0.83	0.75
	biographical	poor	0.6	0.79	0.76	0.64	0.70
		good	0.52	0.72	0.63	0.58	0.61
	overall	poor	0.76	0.86	0.81	0.69	0.78
		good	0.76	0.86	0.8	0.83	0.81
		Overall	**0.76**	**0.86**	**0.81**	**0.76**	**0.80**
Overall	situational	poor	0.56	0.63	0.50	0.55	0.57
		good	0.74	0.79	0.74	0.78	0.76
		Overall	0.71	0.69	0.61	0.59	0.65
	biographical	poor	0.69	0.75	0.73	0.60	0.69
		good	0.67	0.66	0.63	0.68	0.66
		Overall	0.79	0.78	0.71	0.74	0.76
	Overall	Overall	**0.86**	**0.85**	**0.80**	**0.80**	**0.83**

Note. *N* = 279

Rater accuracy refers to the discrepancy between an observed score and a true score, which is often defined as ratings from an expert sample (Sulsky & Balzer, 1988). In our design, the true scores were fixed with the help of an expert sample by using pretested answer

scripts that were fitted to either a poor or good level of qualifications, as experts might be biased by the use of any media format.

Controls. To rate the candidate's perceived vocal and physical attractiveness, a 7-point Likert scale based on Zuckerman, Hodgins, and Miyake (1990) was used. To measure cognitive load, the German version of the NASA-TLX scale (Hart & Staveland, 1988) was used (Cronbach's alpha = .61). Participants' motivation to provide accurate judgements were assessed with three items on 7-point Likert scale (Cronbach's alpha = .70), also based on past research (Motowidlo & Burnett, 1995; Salvemini, Reilly, & Smither, 1993). Furthermore, participants were asked to rate their level of knowledge about designing interviews, conducting interviews, rating candidates based on interviews, the job of an office clerk, and the commerce industry. They also were asked whether and what kind of experience they had with technologically supported interviews (like telephone and video conference interviews). Demographic variables included gender, age, and occupation. Potential gender effects were also checked prior to further analyses; no significant effects emerged. Perceived *likeability* and *closeness* were assessed with two-items measures, which reached internal consistencies of Cronbach's alpha α = .66 for likeability and α = .93 for closeness. Voice attractiveness was assessed as a one-item measure.

Analytic Strategy

To test the research questions regarding rating accuracy, I calculated the discrepancy between the observed score and the true score. In contrast to other studies on interview accuracy, such as the study by Melchers et al. (2011), the four accuracy components identified by Sulsky and Balzer (1988), namely *elevation* (E) as average accuracy across all ratings, *differential elevation* (DE) as accuracy in distinguishing between ratees, *stereotype accuracy* (SA) as accuracy with respect to the evaluation of different dimensions across ratees, and *differential accuracy* (DA) as accuracy to detect ratees' specific patterns of strength and weaknesses across dimensions, did not apply to our study goal and design for the following

reasons: The study's focus was to investigate media format and construct dimension effects after *eliminating all within-candidate variance*. Thus, DE is not meaningful when only one ratee is considered and detecting within-ratee variance is not considered relevant. SA is not meaningful when the true scores on each dimension are all fixed to a constant level. Thus, DA is not meaningful either because it represents the combination of the aforementioned two scores.[27] Nevertheless, the same strict analysis of variance (ANOVA) logic was used to extract each component expressing a different portion of the discrepancy between rater ratings and true scores with respect to our four factors. Specifically, linear mixed-effect modeling was used to calculate a 4 (media format: video/audio vs. audio vs. video/text vs. text) x 4 (construct dimension: job knowledge vs. organization and planning skills vs. teamwork vs. conscientiousness) x 2 (question type: situational vs. biographical) x 2 (answer quality: high, low) nested ANCOVA using restricted maximum-likelihood (REML) estimation on discrepancies between evaluated ratings and true scores with experience in rating interviews as a covariate and participants' IDs as random effects. The restricted maximum-likelihood method is unbiased to possible variance heterogeneity (Vallejo, Fernández, Cuesta, & Livacic-Rojas, 2015). Prior to hypothesis testing, possible effects on the overall rating due to gender, age, randomization order, scores on individual competences, and cognitive load were explored in several similar ANCOVAs, which did not reveal any significant effects for these control variables. Hence, they were not included in the further analyses for the sake of parsimony.

---

[27] As we expect questions about this approach, we also calculated E and DE by considering the different answer quality levels rather than different ratees, as these accuracy components refer to the detection of differences between good and poor candidates. Using E and DE in a simple ANOVA design did not change the results. E and DE were correlated with $r(277) = .28$, p < .001, which is in line with previous research (e.g., Melchers et al., 2011)

5.2.2. Results

Descriptive Statistics

Table 5.3 presents descriptive statistics and correlations for all study variables. Intercorrelations for questions falling under the same construct dimension ranged from *r(278)* = .39, *p* < .001 for consciousness to *r(278)* = .60, *p* < .001 for teamwork. A confirmatory factor analysis was performed to verify the intended four-factor model for the eight questions used in this study. Common fit indices indicated good fit for the four-factor model with $\lambda^2(14)$ = 19.73, *p* = .14, *CFI* = .987, *RMSEA* = .037 (Hu & Bentler, 1999). Thus, the assumption that each construct dimensions was represented by two questions each was supported.

Table 5.3. *Means, standard deviations and intercorrelations for all study variables across all media groups (Study 2)*

Nr.	Variable	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	Age	29.52	10.23	-																
2	Gender (1 = female, 2 = male)	1.27	0.44	0.22***	-															
3	Working Experience (Years)	8.15	9.26	0.93***	0.18**	-														
4	Experience in Rating Interviews	2.47	1.79	0.28***	0.12*	0.30***	-													
5	Cognitive Load	7.28	2.45	-0.11	-0.06	-0.10	-0.13*	-												
6	Likeability	3.85	1.15	-0.01	0.05	0.00	0.01	0.05	-											
7	Perceived Closeness	2.49	1.21	-0.04	0.08	-0.01	-0.04	0.16**	0.57***	-										
8	Voice Attractiveness	3.89	1.53	0.01	-0.04	0.01	0.03	0.12	0.46***	0.26***	-									
9	Motivation	5.94	0.87	0.01	-0.05	0.03	-0.08	-0.19**	0.14*	0.04	0.11	-								
10	Job Knowledge (SI)	4.66	1.59	-0.10	-0.11	-0.06	-0.05	0.04	0.02	0.03	-0.10	-0.02	-							
11	Job Knowledge (BD)	4.48	1.52	0.04	-0.08	0.08	0.12*	0.07	0.01	-0.01	-0.09	-0.12*	0.56***	-						
12	Organization & Planning Skills (SI)	3.86	1.79	-0.02	0.01	-0.02	-0.08	-0.08	0.10	0.12*	0.03	0.06	-0.19**	-0.27***	-					
13	Organization & Planning Skills (BD)	4.99	1.46	-0.07	0.05	-0.04	-0.04	-0.02	0.11	0.10	0.00	0.05	-0.04	-0.19**	0.55***	-				
14	Teamwork (SI)	4.83	1.81	-0.07	-0.08	-0.07	-0.05	0.02	0.09	0.10	0.01	-0.04	-0.09	-0.15*	-0.12*	-0.15*	-			
15	Teamwork (BD)	3.83	1.69	0.00	-0.02	-0.01	-0.02	-0.04	0.19**	0.15*	0.03	-0.05	-0.06	-0.13*	-0.12*	-0.06	0.60***	-		
16	Conscientiousness (SI)	4.91	1.40	0.00	-0.02	0.03	-0.02	-0.02	0.01	-0.04	0.12	0.09	-0.10	-0.18**	-0.05	0.01	-0.14*	-0.07	-	
17	Conscientiousness (BD)	3.99	1.33	0.02	-0.04	0.06	0.05	0.14*	0.01	0.02	0.07	-0.09	0.04	0.01	-0.16**	-0.03	-0.08	-0.02	0.39***	-
18	Overall Hireability	4.57	1.11	-0.21***	-0.06	-0.18**	-0.13*	0.10	0.35***	0.31***	0.11	0.09	0.23***	0.14*	0.18**	0.19**	0.18**	0.23***	0.10	0.25***

Notes. N = 279. SI = situational question, BD = behavior description question, * *p* < .05, ** *p* < .01, ***p* < .001
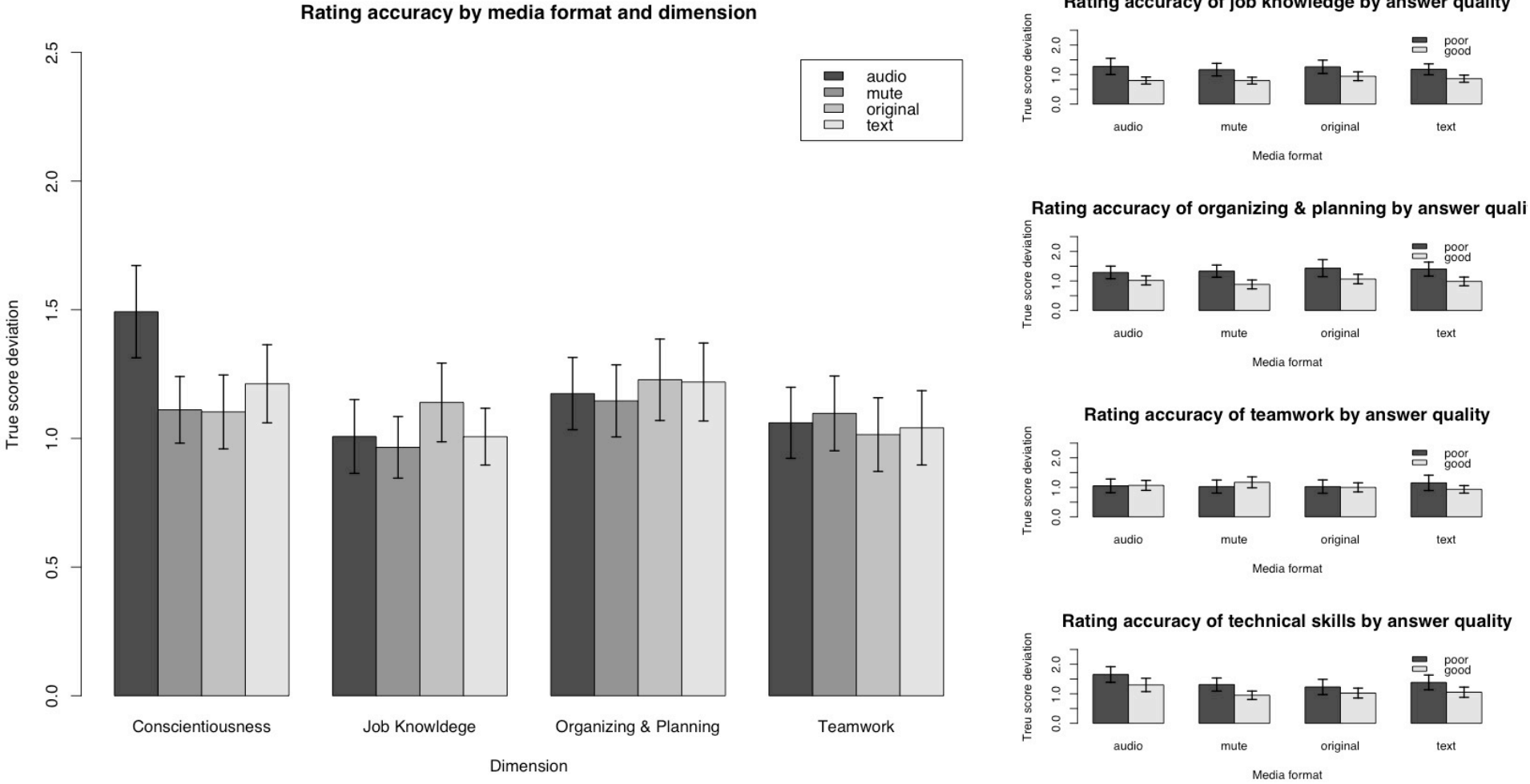
Figure 5.2. Rating accuracy by media format and construct dimensions

Testing Research Questions

To test *Research Questions 2.1* and *2.2* regarding rating accuracy, a nested ANCOVA (Table 5.4) was calculated to test for significant effects of media format, construct dimension, question type, and question quality on rating accuracy (= true score deviation) with experience in rating interviews as a covariate, as described above. The analyses revealed significant main effects for construct dimension, with $F(3/1939) = 7.875, p < .001$, and answer quality, with $F(1/1939) = 63.004, p < .001$. Model effects indicated that true score deviations for the dimensions job knowledge ($t = -4.476, p < .001$), organization and planning ($t = -3.205, p < .01$), and teamwork ($t = -3.955, p < .001$) were significantly lower than for conscientiousness, with lower deviations indicating a higher level of accuracy. True score deviations were significantly lower for good answers compared to poor ones ($t = -7.686, p < .001$). No significant main effect emerged for media format and question type. The interaction between media format and construct dimension failed to reach significance; with $p = .06$, it was instead on a marginally significant level. Therefore, the answer to *Research Questions 2.1* and *2.2* must both be negative. *Figure 5.2* depicts the mean values of all true score deviations.

Table 5.4. *Nested ANCOVA to test for significant effects of media format, construct dimension, question type, and question quality on rating accuracy (= true score deviation) in Study 2*

Variable	df	F-statistic	*p*
Experince in rating interviews	1	0.082	0.775
Media format (Social bandwidth)	3	1'289	0.279
Construct dimension	3	7'875	**0.001***
Answer quality	1	6'300	**0.001***
Question type	1	1'839	0.175
Media format:construct dimension	9	1'814	0.061

Note. $N = 279$, *** $p < .001$

In ancillary analyses, a hierarchical dummy regression analysis was calculated to predict overall hireability ratings using the control variables, dimension construct ratings, and media formats. To code the media format, three dummy variables (n factors – 1) referring to video/audio, audio, and video/text were computed, with participants in the relevant group coded as 1 and participants in the other two groups coded as 0. Participants in the text only group were coded as -1 in all three dummy variables (Miller & Haden, 2006). First, the control variables referring to raters' age, gender, experience in rating interviews, cognitive load, likeability perceptions, and closeness perceptions were entered. Next, the construct dimension ratings averaged across the two questions for each dimension were entered. Finally, the dummy-coded media variables for predicting media format effects were entered. As shown in Table 5.5, the hierarchical dummy regression analysis regarding the overall hireability rating yielded highly significant results [$R^2 = 0.43$, $F(13,262) = 15.11$, $p < 0.001$]. Control variables alone accounted for 20% ($p < .001$) of variance, and dimensional ratings for an additional $\Delta R^2 = .22$ ($p < .001$) of variance. The last block encompassing the dummy media format variables did not explain significantly more variance compared to the previous model ($\Delta R^2 = .01$, $p = .23$). The same was true for the fully mediated model, with $\Delta R^2 = .05$, $p = .08$ (not depicted in Table 5.5 due to the large number of interaction terms). Thus, I will only report Step 2: Participants' age ($\beta = -0.163$, $p < .01$) and likability perceptions ($\beta = 0.243$, $p < .001$) significantly predicted overall hireability rating as well as all four dimensional ratings, namely job knowledge ($\beta = -0.375$, $p < .001$), teamwork ($\beta = 0.321$, $p < .001$), organization and planning ($\beta = 0.333$, $p < .001$), and conscientiousness ($\beta = 0.319$, $p < .001$).

Table 5.5. *Hierarchical regression analysis predicting overall hireability rating (Study 2)*

	Step 1		Step 2		Step 3	
	beta	*t*	*beta*	*t*	*beta*	*t*
Age (rater)	-0.185	-2.907**	-0.163	-2.961**	-0.153	-2.759**
Gender (rater, 1 = female, 2 = male)	-0.046	-0.736	0.014	0.266	0.016	0.302
Experience in rating interviews	-0.074	-1.167	-0.079	-1.428	-0.084	-1.499
Cognitive load	0.020	0.328	0.015	0.281	0.019	0.348
Likability	0.321	4.361***	0.243	3.804***	0.235	3.652***
Closeness	0.168	2.258**	0.117	1.816	0.133	2.043*
Job knowledge (overall)			0.375	6.815***	0.372	6.736***
Teamwork (overall)			0.321	5.821***	0.326	5.904***
Organization & planning (overall)			0.333	5.951***	0.338	6.013***
Conscientiousness (overall)			0.319	5.996***	0.316	5.942***
Video/audio (dummy coded)					-0.042	-0.627
Audio (dummy coded)					-0.061	-0.917
Video/text (dummy coded)					0.134	2.064*
Delta R^2				0.22***		0.01
R^2		0.20***		0.42***		0.43***
F		11.47		19.11		15.11

Note. $N = 279$. *$p < .05$, **$p < .01$, ***$p < .001$

5.2.3. Discussion

Summary

Study 2 was designed to test the interaction effects between social bandwidth and construct dimensions on rating accuracy in employment interviews in a randomized, balanced experimental research design. Contrary to our hypotheses, the data did not provide evidence of significant main effects of a social bandwidth manipulation on rater accuracy or meaningful interactions between social bandwidth and the construct domain being assessed. Nevertheless, these non-significant results are not trivial, as they encourage as revisiting of both the theory and the methodology used in the present study to treat multiple sources of variance.

First of all, the results are at first glance inconsistent with past research on interview and performance ratings, which has repeatedly found significant main effects of the media format

on several outcomes (Motowidlo & Burnett, 1995; Sears et al., 2013; Vanhove, Gibbons, & Kedharnath, 2016). Upon closer examination of these initially counter-intuitive results, three possible explanations emerge that might inform further research on personnel selection in the age of digitalized recruitment processes.

(1) This study incorporated the highest level of interview structure in Chapman and Zweig's (2005) classification system, involving the development of behaviorally anchored rating scales for each question, the usage of a framework that ensures maximum evaluation standardization and maximum administration consistency, and a highly sophisticated question development process. The study design further eliminated any potential effects due to different levels of rapport building, which has been discussed as another possible source of variance and bias (Barrick, Swider, & Stewart, 2010; Chapman & Zweig, 2005; Dipboye, 1994). Meta-analyses have repeatedly shown that the level of interview structure is positively related to interview reliability and validity (Huffcutt et al., 2013, 2014; Levashina et al., 2014). Thus, the first interpretation of these results contributes to the literature on interview structure by suggesting that the level of structure might even suppress media effects to almost zero variance due to bias.

(2) As the results of this study do not indicate either a beneficial or a harmful influence of social bandwidth, the underlying conceptual frameworks and their relation to interview structure should be revisited. Media theories suggest that media rich in social bandwidth perform better in situations with a large amount of uncertainty, which is often caused by the lack of a framework or available process for interpreting a message (Daft & Lengel, 1986; Dennis et al., 2008). Behaviorally anchored rating scales like the ones used in this study do provide such a framework for interpretation. Thus, the usage of highly structured interviews can no longer be used as an explanation for media selection, which might instead be driven by social motives or economic rationales. Hence, approaches such as the theory of reasoned

action (Albarracin, Johnson, Fishbein, & Muellerleile, 2001) might provide more adequate explanations as to when cost and time pressure might justify trade-offs in accuracy.

(3) Following the modular approach to selection procedures proposed by Lievens and Sackett (2017), this study did not explicitly consider the stimulus, which was not as interactive as the response mode. It is possible that interactional dyads account for most of the variance transmitted by socially richer media. Hence, more research on this subject is needed in the future.

In three out of four dimensions, the rating accuracy was lower for poor-quality answers compared to high-quality answers. One explanation for this surprising effect might be that behavioral description anchors provide better guidance for good answers, in which the described expected behaviors are present, than for poor answers, which reflect the absence of the described expected behaviors. This might not have appeared had the participants received some kind of frame-of-reference training, which is effective for increasing rating quality (Melchers et al., 2011). Future research should examine this phenomenon in more detail.

Limitation & suggestions for future research

A first limitation of this study is that the study design precludes any interpretation of face-to-face interviews and is limited to between-media comparisons. Past research has found that interviewers' behavior and conflicting goals and motivations have a significant effect on interview outcomes. For example, interviewers might oversell the position by overemphasizing the recruiting function of the employment interview (Marr & Cable, 2014), producing a self-fulfilling prophecy (Dipboye, 1982), or engage in varying levels of rapport building (Barrick et al., 2010). The null results of this study can be seen as an inspiration to take a closer look at how interaction is affected by different media formats and how media-dependent dyads might affect interview outcomes.

The laboratory design of this study allowed us to completely control for the level of candidate performance and availability of different cues, which would have not been possible

in a field setting. Furthermore, our sample mainly consisted of students in a hypothetical setting, which might have decreased the motivation to provide accurate ratings. Nevertheless, past research has repeatedly suggested that experts and novices achieve comparable results in rating interviews when descriptive behavioral anchors are provided (S. D. Maurer, 2002). Future research should aim to replicate these findings with more advanced methods in field-based studies.

Furthermore, agreement and accuracy are two important criteria for the quality of performance ratings, but past research has shown that different measures differ in direction and magnitude (Sulsky & Balzer, 1988). Therefore, future research may wish to use multiple measures of psychometric quality.

In addition, for reasons of standardization, only one male actor served as the rating target. However, while the findings are somewhat conflicting, past research indicates that a candidate's gender can influence interview outcomes with d greater than .20 (Cable & Judge, 1997; Levashina & Campion, 2007; McCarthy, Van Iddekinge, & Campion, 2010). Huffcutt (2011) summarizes this research as follows: "The evidence does not appear to support a gender similarity effect" (p. 73). As gender might also influence interaction style or the usage of different media formats, future research should attempt replicate these results while varying the target's gender.

Moreover, despite our elaborate procedure for developing true scores, we cannot be certain that the expert ratings were truly accurate due to the lack of an appropriate criterion for evaluating such expert ratings (Sulsky & Balzer, 1988).

The theory section of this thesis referred to research finding that the non-observable construct domains of knowledge and analytical thinking exhibit lower interrater agreement than more observable characteristics (Ambady et al., 2000). One explanation for why this effect did not occur in our study might be that the chosen operationalization of job knowledge

and skills was not sufficiently cognitively loaded. Future research should attempt to replicate these results using a parallel measure of cognitive saturation with different targets.

Conclusion

This study contributes to the growing body of research on technology-enhanced selection formats (Blackman, 2002; Chapman & Rowe, 2002; Chapman & Webster, 2001; Levashina et al., 2014; Sears et al., 2013; Silvester & Anderson, 2003; Silvester et al., 2000; Straus et al., 2001).

To the best of my knowledge, this was the first primary study to systematically compare the effects of different media formats and construct domains. The results presented once again demonstrate that the importance of the level of structure for the psychometric quality of employment interviews is independent of the media format used to conduct the interview. Furthermore, the findings might further encourage practitioners to develop good behavioral description anchors not only in order to decrease idiosyncratic interpretations (S. D. Maurer, 2002; Melchers et al., 2011), but also to reduce the effects caused by the cue multiplicity of a given media format. Therefore, a high level of structure remains the most effective way to prevent rating distortions in technology-mediated interviews.

5.3. Study 3: Predictive and incremental validity of AVI

Study 3 presents an initial investigation into the validity of asynchronous video interviews in the field in high-stakes selection testing. In high-stakes selection, test results play a critical role in whether or not a candidate receives an employment offer, is accepted into a training program, or other credentials (Sackett, Borneman, & Connelly, 2008; Sackett, Schmitt, Ellingson, & Kabin, 2001). In high-stakes selection, the use of cognitively loaded test batteries is very common, and the validity of such procedures is widely accepted as the meta-analytic evidence is unambiguous (Sackett et al., 2008).

5.3.1. Method

Sample and procedure

This study was conducted in the context of a selection process for future managers of a large multinational corporation in the energy industry based in Southern Europe. This is a high-stakes entry-level setting, as the hired candidates receive manager positions after a two-year training program. The sample for Study 3 consisted of a total of $N = 899$ applicants for 25 future leadership positions for young professionals in 2012. These 899 individuals (578 women, age $M = 24.51$, $SD = 1.74$, range 21 - 29) had already been preselected based on their résumés out of a total applicant pool of about 3,300 applicants. After the second screening stage, which consisted of a battery of unproctored online assessments (on applicants' cognitive abilities, personality, integrity and foreign language proficiency), 436 applicants (293 women) aged $M = 24.43$ ($SD = 1.85$) remained. At the third stage, asynchronous video interviews were used to conduct a further screening. After this stage, data for a final pool of 74 applicants (26 men, aged $M = 24.16$, $SD = 1.29$) who participated in an assessment day and a behavioral interview were available. Out of this final pool, 29 candidates were selected. Therefore, the overall selection ratio was less than 1%. Out of the 899 individuals in the initial sample, only 18 candidates reported that they had a minority status. Due to this very small ratio, minority status was included in all subsequent analyses.

Measures

General cognitive abilities (GMA): GMA was assessed using a Rasch-scaled adaptive test on logic reasoning *(cut-e, 2018a)*. Deductive logical reasoning refers to perceptional speed in applying a number of rules to new situations. The test developers report excellent internal consistencies in two samples ranging from .92 - .93 as well as a test-retest reliability of r (141) = .70 (cut-e, 2018a). In addition, Rasch-scaled tests of numerical and verbal intelligence that were previously found to have split-half reliabilities ranging from .71 - .86 (cut-e, 2018b) were used.

Asynchronous video interviews: The asynchronous screening interviews consisted of four interview questions that were rated on the two dimensional criteria *Innovation* and *Goal*

Orientation on 5-point anchored rating scales. For example, one question addressing goal orientation was *"What was the most challenging goal that you have already set for yourself and why was it so challenging? What did you do specifically to achieve this goal? What was the result in the end? Would you do anything different if you had the opportunity?"* These two dimensional ratings were averaged to create a total score ($\alpha = .97$).

Assessment Center: The following screening step consisted of a two-day on-site assessment center that was also partially designed as a recruiting event and included several exercises (e.g., role plays). Candidates' performance was scored on post-exercise dimensional ratings for the four dimensions of *Goal Orientation*, *Client Orientation*, *Teamwork*, *Vision & Strategy*. 5-point Likert scales were used. These dimensions were averaged to create a total score ($\alpha = .65$)

Behavioral interviews: Behavioral interviews were conducted with respect to the dimensions of *Client Orientation*, *Teamwork*, *Innovation*, and *Vision & Strategy*. The interviews were evaluated on 5-point Likert scales. A total interview score was computed by averaging the scores on these four dimensions ($\alpha = .70$).

Analytical approach & treatment of range restriction

Careful treatment of range restriction is important for comparing selection predictors across different steps of the selection process; according to Sackett et al. (2008), failure to take range restriction into account can dramatically distort research findings. When some predictors exhibit more range restriction than others, the validity of the more restricted predictors are underestimated and the incremental validity of the other predictors are overestimated (Lievens & Patterson, 2011). Due to the two-stage top-down selection procedure, with some candidates eliminated from consideration after the online tests and after the asynchronous video interviews, the appropriate correction for range restriction is Thorndike's Case 3 (Thorndike, 1949), for which the appropriate correction methodology is the multivariate range restriction formula by Ree, Carretta, Earles, and Albert (1994). Thus, in

the first step, the AC and FtF interview groups ($N = 74$) were treated as the restricted group

and the video interview group ($N = 436$) as the unrestricted group to correct for range

restriction. Next, the video interview group was treated as the restricted group and the online

assessment group as the unrestricted group ($N = 899$), and the correction formulas were

calculated using the already corrected correlations from the previous correction. Following

Sackett and Yang's (2000) recommendations, statistical significance was determined prior to

range correction.

To investigate our research question, I first examined the zero-level intercorrelations

between the overall total averaged scores for the different selection procedures in order to

examine convergent and discriminant validity between the different classes of assessment

procedures in the tradition of Campbell and Fiske's (1959) multitrait multimethod approach.

Next, I applied path analyses based on previous research in the field with comparable settings

and samples (Lievens & Sackett, 2007).

5.3.2. Results

Intercorrelations. Table 5.6 summarizes the means, standard deviations, and zero-order

correlations for all study variables. With respect to the research question concerning the

relationship between asynchronous video interviews and much richer face-to-face

employment interviews with behavioral outcomes, the data indicate a corrected correlation

between asynchronous video interviews and assessment center outcomes of $r(74) = .32$, p

$< .01$. The correlation between face-to-face interviews and assessment center scores was also

significant, with $r(74) = .30$, $p < .05$. Hence, both correlations were at a similar level. The

correlation between asynchronous and face-to-face interview outcomes was found to be $r(74)$

$= .31$, $p < .05$.

The corrected zero-order correlation between cognitive abilities and asynchronous video

interviews did not reach significance [$r(434) = .08$, ns)]. The correlation between mental

abilities and face-to-face interviews did reach significance, with corrected $r(74) = .39$, p

< .001. These results indicate that asynchronous video interviews and face-to-face interviews differ with respect to their cognitive saturation, which seems to be higher for face-to-face interviews compared to asynchronous video interviews.

Table 5.6. *Means, standard deviations, and intercorrelations for all study variables (Study 3)*

	Variable	M	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
	After CV screening (N = 899)																			
1	Gender (females = 1)	1.36	0.48																	
2	Numerical intelligence	15.45	5.73	-0.07*																
3	Verbal intelligence	12.42	5.04	-0.02	0.37***															
4	Logical reasoning	16.36	6.63	-0.01	0.15***	0.09**														
	After online screening (N = 436)																			
5	AVI: Innovation	3.06	0.97	0.03	0.18***	0.14**	0.06													
6	AVI: Goal orientation	3.32	0.87	0.04	0.19***	0.17***	0.07	0.82***												
7	AVI: Total	3.19	0.88	0.04	0.20***	0.16***	0.07	0.96***	0.95***											
	After video interview (N = 74)																			
8	AC: Goal orientation	3.78	0.60	0.10	-0.21	0.07	0.08	0.15	0.15	0.18										
9	AC: Client orientation	3.50	0.59	-0.03	0.01	0.07	-0.09	0.17	0.09	0.16	0.40***									
10	AC: Teamwork	3.81	0.70	0.13	-0.15	0.08	0.04	0.03	-0.02	0.01	0.40***	0.29*								
11	AC: Inovation	3.33	0.87	0.02	-0.19	0.01	0.11	0.08	0.12	0.12	0.69***	0.40***	0.08							
12	AC: Vision & strategy	3.66	0.67	-0.24*	0.15	0.09	-0.09	0.12	0.24*	0.21	0.03	0.48***	0.31*	-0.14						
13	AC: Total	3.60	0.50	0.08	-0.19	0.07	0.06	0.14	0.12	0.16	0.84***	0.69***	0.60***	0.78***	0.20					
14	FtF interview: Client orientation	4.00	0.60	-0.02	-0.15	-0.05	0.10	0.10	0.14	0.14	0.27*	0.22	0.13	0.20	0.09	0.28*				
15	FtF interview: Teamwork	3.92	0.64	-0.16	0.02	-0.08	0.26*	0.13	0.01	0.09	0.21	0.09	0.26*	0.16	0.03	0.25*	0.35**			
16	FtF interview: Inovation	3.85	0.68	-0.07	-0.21	-0.19	0.23*	0.13	0.06	0.12	0.25*	0.07	0.23	0.23	-0.06	0.27*	0.45***	0.46***		
17	FtF interview: Vision & strategy	3.79	0.62	-0.32**	0.18	0.12	0.35**	-0.02	0.22	0.10	0.12	0.08	-0.12	0.24*	0.13	0.11	0.27*	0.30**	0.40***	
18	FtF interview: Total	3.89	0.46	-0.20	-0.05	-0.07	0.33**	0.12	0.14	0.16	0.28*	0.14	0.16	0.28*	0.05	0.30*	0.70***	0.73***	0.81***	0.68***

Note. N = 74 - 899 depending on stage of the process. *p < .05, ** p < .01, ***p < .001. AVI = asynchronous video interviews, AC = assessment center, FtF = face-to-face.

Table 5.7. *Intercorrelations between selection methods without correcting for range restriction (Study 3)*

	Variable	1	2	3	4	5	6	7
1	Integrity							
2	Numerical	-0.05						
3	Verbal	-0.03	0.37***					
4	Logic reasoning	0.00	0.15***	0.09**				
5	Foreign language skills	-0.07*	0.40***	0.41***	0.07*			
6	Asynchronous video interviewing	0.09	0.20***	0.16***	0.07	0.16***		
7	Assessment center	0.14	-0.19	0.07	0.06	0.06	0.16	
8	Face-to-face interview	0.22	-0.05	-0.07	0.33**	-0.03	0.16	0.30*

Note. N = 74 - 899 depending on stage of the process. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 5.8. *Intercorrelations between selection methods with two-step correction for range restriction (Study 3)*

	Variable	1	2	3	4	5	6	7
1	Integrity							
2	Numerical	-0.02						
3	Verbal	0.03	0.34					
4	Logic reasoning	-0.02	0.00	0.02				
5	Foreign language skills	-0.01	0.37	0.39	0.02			
6	Asynchronous video interviewing	0.09	0.22	0.17	0.08	0.17		
7	Assessment center	-0.01	-0.05	0.09	-0.03	0.02	0.39	
8	Face-to-face interview	0.22	-0.03	-0.08	0.40	-0.05	0.31	0.18

Note. N = 74 - 899 depending on stage of the process. Correlations equal to or above .23 are significant at $p < .05$; correlations equal to or above .30 are significant at $p < .01$.

Path Analyses. Next, I used the corrected intercorrelations from Table 5.8 (Table 5.7 shows the corresponding uncorrected correlations) to test whether the shared variance between asynchronous video interviews and face-to-face interviews both explain the same proportion of variance in behavioral outcomes. Therefore, I modeled the relationship as path analyses using the *lavaan* package for R for Macintosh (R Core team, 2014). The two regression paths from asynchronous video interviews to assessment center performance (β = .40, p = .001) and from asynchronous video interviews to face-to-face interview performance (β = .24, p = .05) were significant. The correlation between AVI and GMA remained below the level of significance. Therefore, our data indicate that the information gained from asynchronous video interviews is incrementally valid in preselection above and

beyond GMA testing for predicting behavioral outcomes in the assessment center and

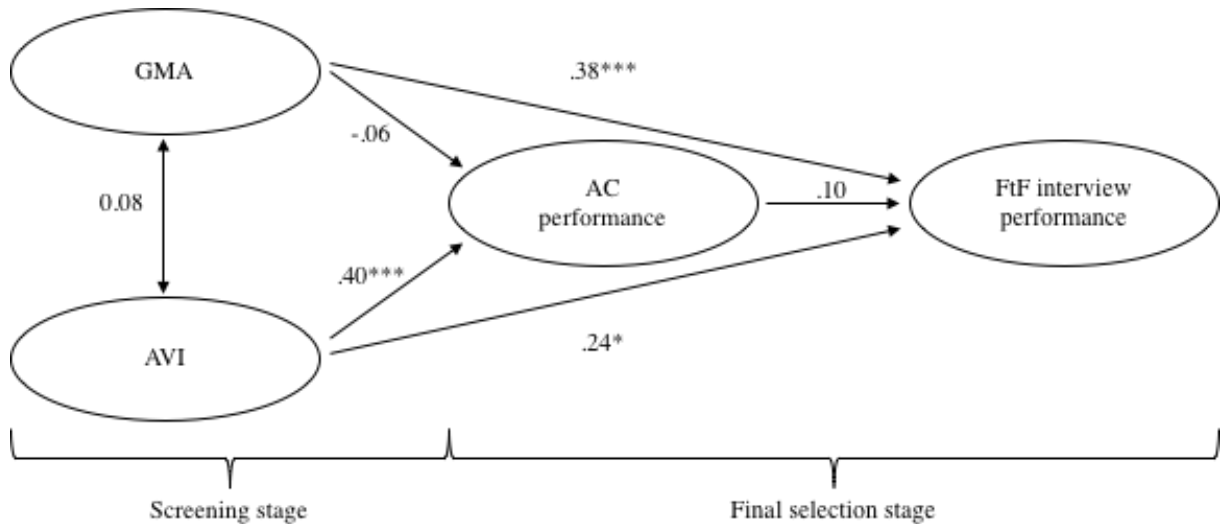performance in FtF interviews. Cognitive saturation was low in this data set.

Figure 5.3. Path Analysis (Study 3). Parameter estimates of hypothesized model. * *p* < .01;

*** *p* < .001. GMA = general mental ability, AVI = asynchronous video interviews, AC =

assessment center, FtF = face-to-face.

5.3.3. Discussion

Examining incremental validity in multi-hurdle selection processes is important because

the use of multiple predictors is quite common in selection practice (Sackett, Dahlke, et al.,

2017). The purpose of Study 3 was to provide an initial exploration of the criterion and

construct-related validity of asynchronous video interviews in high-stakes personnel selection,

as the only existing source of evidence for the validity of AVI comes from a MTurk sample

(Gorman et al., 2018). The data revealed that AVI significantly predicted assessment center

outcomes and was incrementally valid over cognitive testing.

As a first important result, this study suggests that ratings from asynchronous video

interviews are incrementally valid over GMA tests in a sample of future managers from a

large corporation in the industrial sector. The asynchronous video interview ratings were

significantly related to performance outcomes in an assessment center and a behavioral

interview. Interestingly, the correlation between asynchronous video interviews and assessment center outcomes was stronger compared to the correlation between AVI and a face-to-face interview. Hence, these results raise the question of to what extent the two procedures measure the same constructs vs. what proportion of variance is due to the procedure itself, implying a need to look much more deeply into the construct–method–media relationship in the sense of extending Campbell and Fiske (1959) by including a latent media factor in addition to the method (interview vs. cognitive test) factor.

Second, the results indicate that the cognitive saturation of asynchronous video interview rather low in magnitude, and lower than the cognitive saturation of face-to-face interviews. The magnitude of the correlation was found to be in line with previous research in the field (Berry et al., 2007). Next, the study found that asynchronous video interviews explained an additional proportion of the variance in behavioral outcomes compared to face-to-face interviews. This result indicates that the non-verbal cues facilitating accurate behavioral predictions might differ significantly from those used in face-to-face interactions. More research is needed here on the specific cues involved and their level. Finally, the study found a moderate relationship between asynchronous video interviews and face-to-face interviews, which might also be based on distinct cues that depend on interaction.

One of the limitations of this study is the lack of criterion data on job performance. Nevertheless, the results remain meaningful because job simulations are based on the behavioral consistency paradigm between simulated behavior and later performance on the job (Jansen et al., 2013; Motowidlo, Dunnette, & Carter, 1990) and thus commonly used for research purposes (Funke & Schuler, 1998; Ingold, Kleinmann, Konig, Melchers, & Van Iddekinge, 2015; Oostrom, Melchers, Ingold, & Kleinmann, 2016).

From a practical point of view, this research is important because a substantial overlap between asynchronous video interviews and other assessment procedures in early selection stages would negatively impact the utility of the former due to higher effort involved in rating

all candidates' AVI responses as opposed to computer-based psychometric tests with automated scoring. However, our data indicate that asynchronous video interviews measure different domains than cognitive tests or interactive video formats and explain an incremental proportion of behavioral outcomes.

Moreover, Sackett and Roth (1996) noted that changes in the correlations between individual predictors of job performance also lead to changes in the predictive power of the combination of these procedures. In line with this, this study provides first evidence on the relationship between asynchronous video interviews and tests of cognitive ability, face-to-face interviews and assessment centers. Sackett and Roth (1996) also noted that subgroup differences should be considered when evaluating a selection system with multiple predictors. However, this study's sample was very homogenous with respect to personal characteristics; for this reason, the issue of adverse impact was not assessed more deeply. While studies of video-recorded interviews have not found sufficient evidence in favor of high vulnerability to adverse subgroup impact for video interviews in general (Kroll & Ziegler, 2016), future research should aim to investigate this issue in a more diverse sample.

Finally, one might also criticize the relatively small sample size at the end of this very strict selection process, although it is similar to other research in the field, e.g., Lievens et al. (2015). Hence, although this study's findings should be interpreted with caution, it provides first evidence for the criterion-related validity of asynchronous video interview formats in a field-based setting and suggests that this procedure has low cognitive saturation and divergent validity compared to face-to-face interviews. Thus, the results are promising for the use of AVIs but also raise many questions for future research on the construct validity of interviews and combinations of different media formats.

5.4. Study 4: The impact of personalization on applicant reactions

Study 4 was designed to assess the influence of more personal messages on applicant reactions to asynchronous video interviews in a randomized experimental design with non-applicants.

5.4.1. Method

Participants

Ninety-eight adults (64 women, 34 men) were recruited from a medium-sized German university or via the private social network of one of the authors. Participants' age ranged from 19 to 71 years ($M_{Age} = 30.77$, $SD = 11.26$). 47% of all participants were students (mainly psychology majors). Participants did not receive any financial compensation for completing the study.

Procedure

Participants completed the study online. After registering for the study, participants received an invitation email with instructions and a link to the online platform on which they were asked to complete a pre-recorded video interview. Next, participants were forwarded to the online platform and completed an asynchronous video interview consisting of three typical interview questions selected by experienced human resource professionals. Before the interview began, the participants were randomly presented a video message from a fictitious company representative with a high level of personalization, one with low level of personalization, or no video message at all, which served as control group. After completing the video interview, the participants completed several questionnaires about their perception of procedural fairness, Big Five personality traits, computer self-efficacy, recommendation intentions, and demographic characteristics.

Materials

Level of Personalization: The manipulation involved three conditions. In the control group, participants did not see any video before they conducted the asynchronous video interview. In the second group, a video was presented at the beginning of the process in which a semi-professional actor posing as a fictitious HR manager provides general information about the company and video interview process. In the third condition, the same semi-professional actor introduces himself personally as the HR manager responsible for the selection process before providing the same information on the process. A pre-test ensured that the videos were perceived as different in their level of perceived personalization. In addition, one question was added to the design asking whether respondents perceived the person in the video as personal in his nature. The correlation between the two video groups (dummy coded) and the manipulation check was $r(66) = .58$, $p < .001$.

Big Five: The Big Five personality dimensions were assessed using self-ratings with the 42-item German Big Five inventory BFI (Lang, Ludtke, & Asendorpf, 2001). All items were rated on 5-point Likert-type scale. Internal consistencies were sufficient, with $\alpha = 0.76$ for openness, $\alpha = 0.79$ for conscientiousness, $\alpha = 0.87$ for extraversion, $\alpha = 0.68$ for agreeableness, and $\alpha = 0.84$ for neuroticism.

Procedural fairness: Procedural fairness was assessed using Steiner and Gilliland's (1996) the seven-item questionnaire of applicants' perception of the fairness of selection procedures. Two out of seven items were reverse-coded. Cronbach's alpha for the scale was sufficient, with $\alpha = 0.81$ (raw alpha).

Recommendation intention: Recommendation intention was assessed by using the 3-item scale by Bauer, Maertz, Dolen, and Campion (1998). Cronbach's alpha was calculated to be $\alpha = .90$.

Controls: As a control variable*, computer self-efficacy* was assessed using the Computer Usage Self-Efficacy Scale *($\alpha = 0.87$)* by Spannagel and Bescherer (2009).

5.4.2. Results

Table 5.9 provides a summary of the means, standard deviations, intercorrelations, and internal reliability correlations among all study variables. To explore the study's driving research question, I tested first for a main effect of the personal message on applicants' reactions toward AVI. Two separate one-factorial ANOVAs were conducted to determine whether there was a statistically significant effect of the video message's level of personalization on candidates' procedural justice perceptions and recommendation intentions, controlling for the Big Five traits as well as computer self-efficacy. The level of personalization did not have a significant effect on either procedural justice perceptions or recommendation intentions after controlling for personality and computer self-efficacy, $F(2/90) = 0.41$, ns, for procedural justice perceptions, and $F(2/90) = 0.73$, ns for recommendation intentions. Subsequent univariate ANCOVAs were calculated to test for significant effects of the video messages' level of personalization on single procedural justice dimensions, controlling for personality and computer self-efficacy. Again, no effect emerged for any of the scales. Table 5.10 presents the means and standard deviations for the dimensions of Steiner and Gilliland's (1996) questionnaire in each condition.

Finally, a moderated regression analysis to predict recommendation intentions with the level of personalization and procedural justice perceptions as predictors was calculated and accounted for 22% of the variance in the dependent variable, $F(3/94) = 8.70$, $p < .001$. A significant positive interaction between the level of personalization and procedural justice perceptions on recommendation intentions emerged ($\beta = .41$, $p < .05$). The simple slopes are presented in *Figure 5.4*.

Table 5.9. *Means, standard deviations and all intercorrelations for Study 4*

Nr.	Variables	M	SD	1	2	3	4	5	6	7	8	9	10	11	12	13
1	Gender (1 = females, 2 = males)	1.35	0.48	-												
2	Age	30.77	11.26	0.31**	-											
3	Control group (dummy-coded)	0.33	0.47	0.00	-0.13	-										
4	Low personalization (dummy-coded)	0.36	0.48	-0.01	0.16	-	-									
5	High personalization (dummy-coded)	0.32	0.47	0.01	-0.04	-	-	-								
6	Computer self-efficacy	4.61	0.73	0.24*	-0.04	0.06	0.02	-0.08	.87							
7	Openness	2.72	0.38	-0.08	0.04	0.02	-0.01	-0.01	0.01	.76						
8	Conscientiousness	2.13	0.53	-0.08	-0.27**	-0.08	0.03	0.06	-0.10	-0.06	.79					
9	Extraversion	2.14	0.64	-0.06	-0.10	0.02	0.05	-0.07	-0.10	0.08	0.11	.87				
10	Agreeableness	2.19	0.46	-0.06	-0.11	-0.08	0.06	0.02	-0.11	0.02	0.19	0.27**	.68			
11	Neuroticism	3.40	0.69	0.34***	0.26*	0.06	0.14	-0.20*	0.23*	-0.19	-0.27**	-0.47***	-0.22*	.84		
12	Process fairness	4.18	1.00	0.13	-0.14	0.00	0.08	-0.08	0.13	0.03	0.08	-0.07	-0.06	-0.01	.81	
13	Recommendation intention	3.11	0.99	-0.01	-0.12	0.06	-0.11	0.05	0.20*	0.07	-0.25*	-0.35***	-0.18	0.02	0.42***	.90

Note. N = 98. Cronbach's alpha values are depicted initalics on the diagonal. * $p < .05$. ** $p < .01$. *** $p < .001$.

Table 5.10. *Means and standard deviations of justice dimensions from Steiner and Gilliland (1996) by the level of personalization*

| Justice rule	Control group		Low personalization		High personalization	
	M	SD	M	SD	M	SD
Scientific evidence	4.03	1.12	3.94	1.33	4.03	1.33
Face validity	4.22	1.54	4.69	1.53	4.52	1.48
Chance to perform	3.94	1.56	4.00	1.55	3.97	1.33
Interpersonal treatment	3.62	1.76	3.46	1.75	3.71	1.9
Employer's right	4.69	1.15	5.06	1.51	4.48	1.29
Invasiveness of privacy	5.31	1.64	5.34	1.51	4.71	1.51
Widely used	3.38	1.36	3.51	1.34	3.00	1.26
Scale	4.17	0.94	4.29	1.05	4.06	1.02

Note. $N = 98$.
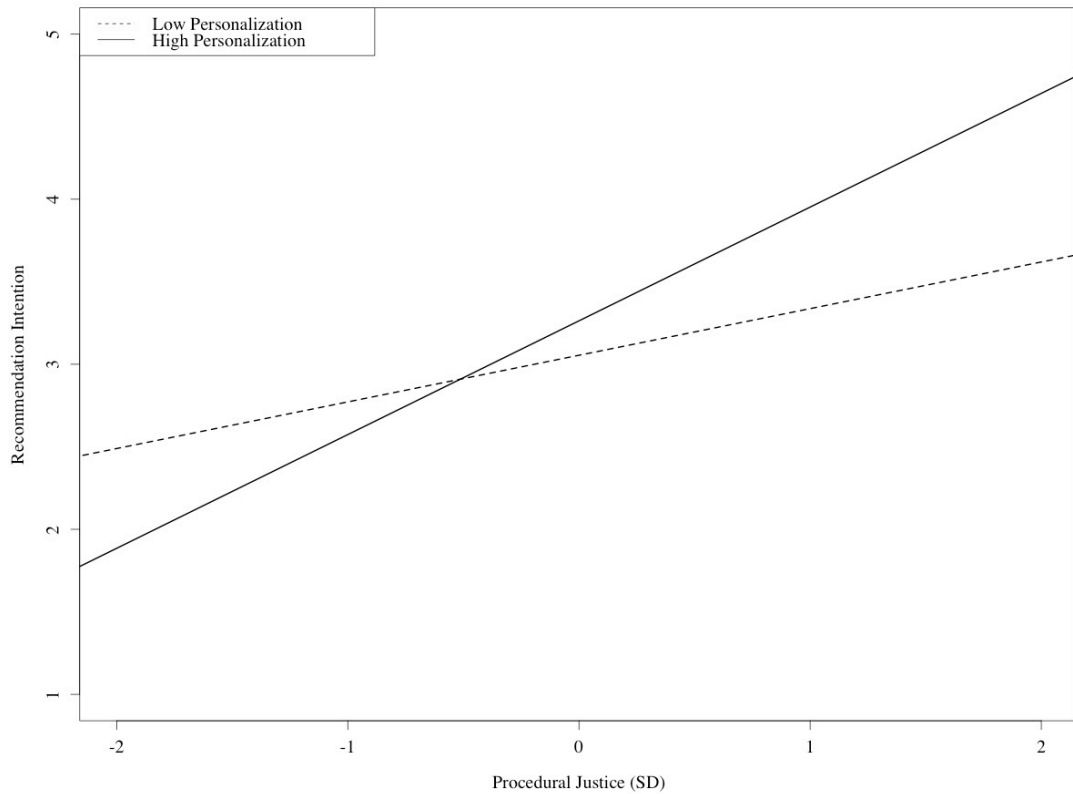
Figure 5.4. Moderating effect of level of personalization on the relationship between procedural justice perceptions and recommendation intentions.

5.4.3. Discussion

Applicants' attitudinal, cognitive, and affective reactions to selection procedures are important when organizations compete for a limited number of qualified candidates (McCarthy et al., 2017). The purpose of this study was to assess whether a social bandwidth manipulation, specifically the manipulation of the level of personalization of an introductory video sequence, impacts applicant reactions to asynchronous video interviewing. The results showed that fairness reactions toward the asynchronous video interviews were in the middle range of Anderson et al.'s classification (2010). In line with previous research (Brenner et al., 2016; Langer et al., 2017), the ratings on fairness rules that rely on interpersonal communication were rated at the lower end

of the scale, and the manipulation of the level of personalization had no effect on the ratings of procedural justice perceptions. Nevertheless, a moderating effect of the level of personalization on recommendation intentions was found.

As a means of influencing candidates' attitudes during the selection process, organizations have begun to use rich media like video blogs that do not include interpersonal communication between candidates and organizational representatives in early selection stages (Banerjee, 2016). This study is the first to explore the effectiveness of personal, media-rich messages for altering reactions toward asynchronous video interviews used in screening processes while controlling for personality and computer self-efficacy.

First, the lack of a main effect could have several explanations: First, as noted before, the term 'interview' has long been defined as face-to-face interaction only (Huffcutt & Youngcourt, 2007), or at least as a form of direct interaction. Thus, the term 'asynchronous video interview' might create an expectation of interpersonal communication. Given that prior expectations can be used as anchors (Tversky & Kahneman, 1974), applicants might expect direct interaction and experience dissatisfaction when this expectation is violated. Moreover, expectations should differ for other assessment procedures not termed 'interviews', such as cognitive tests or personality inventories. Hence, false expectations could explain why asynchronous video interviews are rated even lower than test formats with less socially rich cues.

However, these results are surprising in light of the fact that a manipulation check revealed the manipulation to be successful. First, it must be acknowledged that the introductory video is only one small part of the selection procedure as a whole. Overall, selection procedures are viewed as impersonal and cold, and there might have even been adverse effects due to reactance against the intervention (although this argument is purely speculative). Nevertheless, as noted by

Chan and Schmitt (2004), the entire context of a selection system must be taken into account and not just minor parts of it.

Another important contribution of this study concerns the findings on the role of individual differences for fairness outcomes. The zero-level correlation revealed significant effects of the personality traits of conscientiousness ($r = -.25. p < .05$) and extraversion ($r = -.35, p = .05$) on recommendation intentions. As conscientiousness is related to job performance (Barrick et al., 2001), organizations might risk losing candidates if they use AVI in their selection process, which will have negative effects on the effectiveness of the selection system as a whole (K. R. Murphy, 1986). However, due to the small effect size, future research should aim to replicate this effect with other important predicters of high job performance such as cognitive ability. A negative effect of extraversion on recommendation intentions was also found. According to Costa and McCrae (1992), people high in extraversion are described as sociable, vivacious, spirited and having positive affect. Individuals who score high on this trait may perceive the lack of interaction as an obstacle to effective social behavior and thus are less likely to recommend an employer who applies such a device.

Finally, it has been noted in the literature that applicant reactions are also influenced by major social and technological meta-trends as new forms of assessment procedures and communication media are adopted and accepted over time, as has been the case for social media recruitment via professional networks like LinkedIn (McCarthy et al., 2017). Hence, Gilliland's (1993) perspective on procedural justice is limited in that it does not consider the broader social context, in contrast to other frameworks such as the TAM in its later refinements.

Therefore, these results should be considered in light of the limitation that they were collected at a specific point in time under specific social norms in one cohort. Due to ongoing

technological progress and changing social norms, continued research into applicant reactions and how they relate to new technology is desirable. This might encourage further research to look beyond non-linear relationships to outcome variables that will gain in importance due to the emergence of platforms like Glassdoor, where candidates can rate the recruitment processes of potential employers and therefore attract or dissuade other potential applicants.

Like any study, this study has several limitations, which will be discussed in the following paragraph. However, these limitations do not indicate that the results are not meaningful, but rather that further research on the topic is required. First, a hypothetical setting was used in order to manipulate the level of social bandwidth in the introduction video. Hausknecht et al. (2004) reported substantial differences in the magnitude of effects, but not in the direction of differences, between hypothetical settings and field studies. Second, the study sample included a large share of students, who will be an important target group for selection procedures once they enter the labor market in the near future. Hence, there have been calls to consider students as the relevant population in research on personnel selection methods (Rynes & Boudreau, 1986). However, the sample also consisted of volunteers, which might have led to self-selection bias and restricted variance with regard to the distribution of some traits. Finally, the study is cross-sectional and only addressed a range of short-term reactions. Thus, it would be desirable for further research to also assess long-term effects on variables such as actual job acceptance intention or job performance (Chan & Schmitt, 2004).

5.5. Study 5: Rating inflation due to preparation times

The last study was designed to explore the impact of rehearsability on interview ratings in asynchronous video interview settings with an randomized experimental study design in the lab.

5.5.1. Method

Participants

A total of fifty-one adults (30 women, 21 men) volunteered to participate in the study. Ages ranged from 20 to 48 (M_{Age} = 28.73, SD_{Age} = 5.48). On average, the participants reported $M_{Experience}$ = 4.49 years of work experience ($SD_{Experience}$ = 5.86). No financial or other incentives were provided for study participation.

Design

A one-factorial design (preparation time vs. no preparation time) was applied to test the hypotheses, with three different dimensional ratings as dependent variables. Control variables were collected via a short web survey.

Materials

A mock interview was designed with questions selected by experienced HR professionals. The position of an office assistant was used based on the assumption that most people are in some way familiar with the requirements for this position. Four screening questions were used that addressed three dimensions: communication skills, conscientiousness, and planning & organization. Two trained raters with experience in the HR and selection domain independently rated all answers on these three dimensions on 5-point anchored Likert scales. Overall, interrater agreement between the two raters was very strong, with r = .94 for *Communication*, r = .92 for *Conscientiousness*, and r = .96 for *Planning & organization*. After participants completed the interview, data on demographic variables such as age, sex, and experience with employment interviews were collected using a separate questionnaire.

5.5.2. Results

Table 5.11 presents the means, standard deviations, and intercorrelations among the study variables. A multivariate analysis of variance (MANOVA) revealed that preparation time was a significant factor for explaining variation in interview performance on a global level over all competency dimensions: $F(1/49) = 2.95$, $p < .05$, $\eta^2 = .16$. The MANOVA statistics are presented in Table 5.12.

Subsequent ANOVAs for individual competency scores showed that the effects were reproducible across all three competency dimensions, with $F(1/49) = 5.24$, $p < .05$, $\eta^2 = .09$ for communication skills, $F(1/49) = 8.37$, $p < .01$, $\eta^2 = .15$ for conscientiousness, and $F(1/49) = 5.24$, $p < .05$, $\eta^2 = .10$ for planning & organization. The mean score differences are depicted in Table 5.13. Therefore, our data lend support to research question suggesting that preparation time significantly increases candidates' performance ratings in asynchronous video interviews.

Table 5.11 . *Means, standard deviations and intercorrelations for all study variables (Study 5)*

Nr.	Variable	M	SD	1	2	3	4	5	6
1	Preparation Time (1 = Not available, 2 = available)	1.51	0.50	-					
2	Communication skills	4.12	0.67	-0.31*	-				
3	Conscientiousness	3.29	0.90	-0.38**	0.57***	-			
4	Planning & organizing	3.14	0.93	-0.31*	0.61***	0.69***	-		
5	Gender (1 = females, 2 = males)	1.41	0.50	0.26	-0.15	-0.21	-0.23	-	
6	Age	28.73	5.48	0.05	0.04	-0.12	-0.12	-0.02	-
7	Work experince (years)	4.49	5.86	0.10	0.06	-0.11	-0.13	-0.05	0.92***

Note. N = 51 , * p < .05, ** p < .01, p < .001.

Table 5.12. *Summary of multivariate analysis of variance (MANCOVA) statistics (Study 5)*

Variable	Wilks lambda	df	F	p	η^2
Preparation time	0.84	1	2.95	.05	0.16

Table 5.13. *Means and standard deviations for all dimensional ratings (Study 5)*

Dimension	Preparation time			
	not available		available	
	M	SD	M	SD
Communication	3.92	0.71	4.32	0.57
Conscientiousness	2.95	0.92	3.64	0.75
Planning & organizing	2.86	1.00	3.43	0.76

Note. $N = 51$

5.5.3. Discussion

The goal of Study 5 was to examine the impact of preparation time on rating inflation in asynchronous video interviewing. As predicted, the data revealed a significant main effect of the availability of a time frame for preparation on three dimensional ratings for communication skills, conscientiousness, and planning and organization. The theory section of this thesis discussed different potential explanations for rating inflation, including cheating, impression management tactics, and self-coaching. Recalling the literature on interview training (Maurer 2008), there are three mechanisms that could increase interview ratings: (1) reduced error, (2) improved performance consistent with actual performance, (3) improved performance that is not consistent with actual performance. In this study's setting, no job performance criteria existed, which is one of its limitations. Due to the laboratory setting, the main effect found refers to improved interview performance that might not be related to on the job performance. This issue should be addressed in future research.

Another point that should be addressed by future research is whether individual differences serve as a moderator in the relationship between the availability of preparation time and its impact on interview ratings. There are several individual differences that might have a potential moderating effect. First, neuroticism is the tendency to adapt poorly to stressful situations (Costa

& McCrae, 1992). Individuals who score high on this trait might adapt more poorly to the interview situation, which might prevent them from applying effective tactics that would improve their ratings. Secondly, the ability to identify criteria (ATIC) describes an individual's capability to look behind the intentions of interviewers (Kleinmann, 1993). People with a high level of this ability might use preparation time more effectively because they perfectly understand the scope and intention of the questions. Third, GMA might be a moderator, as might several facets of intelligence like detecting relationships, recognizing analogies or just being cleverer than others. GMA describes the tendency for people who are good at one cognitive task to also tend to be good at other cognitive tasks (Deary, 2012). Thus, individuals with higher general intelligence will tend to find more effective ways of improving their answers or more quickly identify sources to help them improve their answers before recording.

This study employed a hypothetical setting to provide a controlled environment in which the presence vs. absence of preparation time could be objectively compared. However, it is possible that the study participants were not as motivated as they would be in a real high-stakes situation. Meta-analytic results suggest that lab and field results in I/O psychology are substantially correlated and higher than in other fields (Mitchell, 2012). Nevertheless, future research should investigate this research question in a field sample.

Finally, this study did not examine differences in validity. As previous research on the effect of fully disclosing interview questions to applicants before the interview did not find significant differences in criterion-related validity due to question knowledge (Day & Carroll, 2003), significant differences in validity are not very likely to arise when the preparation time available is much lower. Nevertheless, future research should examine effects on validity on top of rating inflation.

6. General Discussion

The goal of this dissertation was to gain a better understanding of the impacts of technology usage on different selection outcomes in employment interviews used for personnel selection. This research issue is highly important, as new technology has a massive impact on organizations' selection practices (Chamorro-Premuzic et al., 2016), but knowledge about these new technologies and their impact on selection outcomes is still scarce in terms of theory development (e.g., Morelli et al., 2017) and primary studies, with a few exceptions (Gorman et al., 2018; Langer et al., 2017).

To fill this gap, first an updated literature review on technology usage in employment interviews was conducted to integrate different perspectives on the issue, including the phenomenological appearance of different kinds of technology-enhanced variations on the interview format, theoretical directions on the conceptualization of technology usage, and previous research on specific outcomes including reliability, validity, applicant reactions, adverse impacts, susceptibility for faking, and overall utility considerations. Based on the assumption of an ongoing modularization process in selection practice, a conceptional framework was proposed that applies media attributes based on the work of Potosky (2008) and media synchronicity theory (Dennis et al., 2008; Dennis & Valacich, 1999) to make predictions about how variation in seven specific media attributes explains changes in reliability, validity, and applicant reactions when conducting employment interviews with technology.

In the second part, five empirical studies were presented as initial investigations into the phenomenological appearance of one new technology in employment interviews that differs most drastically from traditional face-to-face interviews: asynchronous video interviews. Based on the conceptional framework presented above, specific testable hypotheses were formulated

concerning how structural differences as described by media attributes can predict the level of reliability (interrater agreement and rater accuracy), validity, applicant reactions, and systematic mean differences in interview scores.

This general discussion begins with a summary and discussion of the major conclusions of the review section, making several propositions that might guide potentially fruitful avenues for future research. Next, the main findings of the five empirical studies will be summarized briefly, including major directions for further research. Subsequently, the results will be integrated into the previously presented body of theory in order to discuss implications for further theory development and implications for practitioners. Finally, an overall conclusion will be drawn to sum up the present dissertation.

6.1.   Summary

The review section of this thesis addressed several perspectives on technology usage in the interview. First, the phenomenological appearance of the interview has been deeply altered in recent decades and is strongly driven by technological advances such as the broad availability of consumer video applications or smart devices. Thus, it makes sense that current reviews extend the definition of what should be considered an interview, most importantly by including the notion of temporally delayed communication due to the use of new technology (Levashina et al., 2014). Nevertheless, several important criteria remain that distinguish interviews from other selection instruments like psychometric tests, namely (1) interpersonal interaction (even when it is temporally delayed), and (2) at least partially tailored interview questions as stimulus material (which can be presented as text, audio files or in written format), and (3) the evaluation of open-ended answers by one or more interviewers (even if they use information like algorithm-generated scores as *additional* sources of information. These three basic components distinguish

interviews from other formats like video resumes or video-based SJTs (no interview questions) or the use of an avatar as "interviewer", which would be more of a virtual simulation.

> *Proposition 1. All kinds of technology-enhanced interviews are still characterized by the presence of (1) interpersonal interaction, (3) tailored interview questions as response format, and (3) the evaluation of open-ended questions by at least one interviewer.*

Secondly, existing research on technology usage in employment interviews is largely dominated by comparative investigations into the measurement equivalence of two or more specific phenomenological appearances, such as telephone vs. face-to-face vs. videoconference interviews (e.g., Blackman, 2002; Langer et al., 2017; Sears et al., 2013; Silvester et al., 2000; Straus et al., 2001). While fully acknowledging the importance of these cross-technology comparisons, their findings have often led to conflicting implications, with a meta-analytic tendency in the direction of less favorable outcomes when technology is involved (Blacksmith et al., 2016). Nevertheless, even this first meta-analytic evidence does not allow predictions to be made about differences between different technologies, e.g., synchronous vs. asynchronous video interviews. Thus, the overall conclusion from these cross-technology comparisons is limited to the point that (1) technology does actually impact selection outcomes, but (2) there is still a clear lack of theory development that would explain the differential effects of different kinds of technology. This lack of theory development is especially critical when new applications come on the market that have no traditional counterpart, or even worse, exist in many modular constellations that can vary widely across applications and contexts (e.g., selection vs. recruitment). Thus, theory development that goes beyond the holistic all-in-one perspective on

technology-enhanced employment interviews is essential for gaining a better understanding of this research field.

> *Proposition 2. A holistic all-in-one perspective on distinct classes of interviews (i.e., telephone vs. video interviews) can no longer account for the modular appearance of new technologies used in conducting interviews and should be replaced by modular approaches in theory and research design.*

The next step after calling for increased theory development to better understand technology usage in employment interviews is to evaluate the large body of literature on existing conceptualizations of technology and media. As part of this process, several leading theoretical directions were discussed, including media richness theory (Daft & Lengel, 1984), media synchronicity theory (Dennis et al., 2008), two selection domain-specific frameworks (Arthur et al., 2018; Potosky, 2008), the technology acceptance model (Davis, 1989) and finally Orlikowski's (2007) socio-material theory. Out of these theories, Potosky's (2008) framework is the most applicable to assessing potential impacts of technology in employment interviews. Nevertheless, three additional dimensions from media synchronicity theory were identified as meaningful for describing potential structural differences between technology-enhanced interview formats with respect to allowing asynchronous and temporally delayed interactions as well as modification of the extent to which candidates are allowed to prepare their answers. It is important to note that this framework makes it possible to describe structural differences within a given class of technology-enhanced interviews, such as video interviews, in addition to between different classes, but is also limited by excluding other antecedents of specific outcomes like

applicant reactions or effects resulting from the interaction between the environment and a given technology, which is why the socio-material and the technology acceptance perspectives are also important.

> *Proposition 3. Describing structural differences in media attributes even within a given class of technology-enhanced interviews providing a useful framework for the development of testable hypotheses regarding technology impacts on selection outcomes.*

The next perspective offered in the review section was the integration of findings from the large body of research on interviews in general and on technology impacts specifically to the extent they exist with respect to several important selection outcomes including reliability, validity, applicant reactions, adverse impact, susceptibility to faking, and overall utility analysis. From this perspective, the most prominent finding is the critical importance of structural components used to construct and evaluate employment interviews when it comes to increasing psychometric properties (Levashina et al., 2014). In contrast, research on direct media effects is relatively slow and poses a conceptual challenge for theory and research practice when it comes to conceptualizing and distinguishing between structural and technological components of interviews. Overall, the review discussed four different ways in which structure and medium could relate to each other: (1) Structural components and media attributes describe the same phenomenon, e.g., whether the interview is recorded, (2) media attributes directly influence a structural component, i.e., question consistency, (3) media attributes and structural components are independent, i.e., conducting job analyses and question construction, or (4) there are interactions between structural components and media attributes, i.e., when structure moderates

the emergence of bias due to media attributes (e.g., Straus et al., 2001). Consequently, the relationship between structure and media attributes could be quite complex. With respect to research practice, it is proposed that studies on the impact of media attributes on selection outcomes should use a methodology that incorporates a constantly high level of structure in order to reduce this source of variance to a minimum and therefore try to maximize the variance in selection outcomes that can be explained by variations in the medium.

Proposition 4. Research practice on the impact of media attributes on selection outcomes should use methodologies that involve a high level of structure to avoid overlap between structural components and maximize observable variance due to media attributes.

Finally, the review section integrates research on employment interviews into the framework of seven media attributes taken from Potosky's framework and media synchronicity theory (Dennis et al., 2008). For each media attribute, evidence and relevant theories from the literature are used to derive expected effects on reliability, validity, candidate reactions and adverse impact. This terminology does not include overall utility considerations, which are most strongly related to various economic indices rather than single media attributes. Susceptibility to faking was considered a subset of validity, as it is often conceptualized as a source of bias in the predictor-criterion relationship (Barrick et al., 2009). Based on this framework, specific hypotheses were developed making predictions about the level of psychometric properties and impacts of specific attributes on selection outcomes for the specific case of asynchronous video interviews in selection.

Proposition 5. The research framework of media attributes provides a useful framework for studying the level of psychometric properties and impacts of specific attributes on selection outcomes in asynchronous video interviews.

Building on the research framework of media attributes in employment interviews, the first study was designed as an initial investigation of the level of interrater agreement in asynchronous video interviews, which are characterized by a lack of interactivity high reprocessability, and medium social bandwidth. In addition, to validate the proposition that research on media attributes should be conducted with a high level of standardization, a manipulation was added representing Levels 3 and 4 in Huffcutt et al.'s (2014) classification. In a non-applicant sample, the study found interrater agreement and accuracy to be within the range common in traditional structured face-to-face interviews. Moreover, altering the level of structure from Level 4 to Level 5 further increased interrater agreement, supporting the notion that media effects should be studied within highly structured interviews to avoid contamination between media and structural components.

The purpose of Study 2 was to assess the impact of social bandwidth on rater accuracy in asynchronous video interviews. In addition, the construct domain was added as a second independent variable, as previous research suggests that differences exist in how important specific cues are for making accurate judgments (Ambady et al., 2000; Blackman, 2002). Using Huffcutt et al.'s (2001) classification, the first part of the research question predicted that social bandwidth will decrease rater accuracy for the construct dimension of knowledge and skills. The second part of the research question predicted that higher levels of social bandwidth will increase rater accuracy for the construct domains of applied social skills and basic personality tendencies.

In data collected from a sample of 297 non-applicants, non-significant effects were found for both parts of the research question. Implications for the relationship between structure and media attributes as well as for further theory-building were discussed in detail.

Study 3 addressed the validity of asynchronous video interviews in a field setting in high-stakes selection. The data revealed asynchronous video interviews to be predictive of behavioral outcomes in an assessment center above and beyond cognitive testing. Moreover, asynchronous video interviews and behaviorally-oriented face-to-face interviews were correlated at an intermediate level. The results of this study are important with regard to the ongoing discussion on the role of cognitive saturation in media formats differing in social bandwidth (Lievens et al., 2015). Furthermore, the intermediate correlation between face-to-face and asynchronous video interviews shows an important need for further clarifying of the relationship between method and medium in regard to explain variation in interview performance.

Study 4 tested the effects of a social bandwidth manipulation with personal video messages as a means of increasing applicant perceptions of asynchronous video interviews, but failed to produce significant results on all procedural justice dimensions. Important implications of this study are the strong relationship between interpersonal communication and the formation of applicant reactions but also the critical discussion about the range of outcomes that can be explained by media attributes alone without considering important contextual factors of selection processes.

Finally, Study 5 tested the impact of rehearsability by either giving candidates the chance to fine-tune their video interviews in a condition with preparation time vs. a condition with no preparation time. This feature is unique to asynchronous video interviews, and process design with respect to this variable might highly impact selection outcomes for candidates and

organizations. In line with expectations, a significant main effect of preparation time on interview ratings emerged across three dimensions in a hypothetical setting. One major limitation of Study 5 is the failure to link rating inflation and impacts on criterion validity due to a lack of data availability. Day and Carroll (2003) found that knowledge of interview questions prior to a situational or behavioral description interview leads to higher interview scores, but not to lower predictive validity for academic performance. Nevertheless, the relationship with availability of preparation time might be different in high-stakes selection in the field for jobs with high complexity and when controlling for cognitive ability and the ability to identify criteria (ATIC). Thus, it would be desirable for future work to examine impacts on job performance when applying a 2 (preparation time vs. no preparation time) by 2 (high complexity vs. low complexity) design and controlling for cognitive ability.

6.2.   Implications for theory

In this section, I will integrate the points made in the review section with the findings from the five empirical studies and discuss their implications for theory and future research. First of all, the disappearance of the traditional boundaries between distinct media classes, such as telephone vs. video interviews, poses a challenge for conducting research on technology in employment interviews and designing empirical studies that go beyond investigating score equivalence. While the holistic perspective does consider some important characteristics, it is insufficient as it does not represent the continuum of media described in media richness or media synchronicity theory. For example, in a single interview conducted via instant messenger, it is possible to switch between text only, recorded sound, and live video chats depending on the requirements of the situation. Similarly, interviewers and interviewees can use the phone (verbal information) combined with additional media files. Thus, applying a modular perspective seems

to be a more adequate approach for hypothesis development than studying the equivalence between distinct classes of selection procedures categorized according to specific medium channels (Lievens et al., 2015). For example, Study 5 found that one procedural element, the opportunity to prepare for approximately two minutes before recording an answer segment, already significantly inflates candidate ratings. This means that even within a single class, (asynchronous video interviews) a procedural design characteristic (preparation time) can significantly influence measure equivalence. Similarly, there should be many other features and process design modifications with significant impacts on selection outcomes.

The results of Study 2 regarding social bandwidth are particularly challenging for theory, as no significant differences in rating accuracy emerged for the formats of video with sound, sound only, video with text, and text only, which differ in their level of social bandwidth. Research on structure in employment interviews has clearly revealed that higher structure increases reliability and validity (Campion et al., 1997); hence, media effects might disappear in highly structured interview situations. Consequently, future research should address potential moderating relationships between media attributes, interview structure and selection outcomes.

With respect to the media attribute of transparency, Study 3 found in a field sample that asynchronous video interviews accounted for a different proportion of variance in behavioral outcomes than face-to-face interviews, which are much more transparent. One major implication for theory is that these results remain unclear as to whether the differences in predicting assessment center scores were really due to differences in transparency, social bandwidth or interactivity because these three functionalities were confounded within the specific constellation of media attributes employed in the field setting. Future research should identify methodologies to assess different portions of variance in more controlled settings.

In a reply to Morelli et al. (2017), practitioners have criticized that "peer-reviewed theory development often serves the important role of supporting decisions that have already been made in practice" (Gibson et al., 2017, p. 676). Theory can be used for revisions or modifications, but rarely leads to advancements. Hence, highlighting new potential sources of error in areas which cannot be controlled in practice is not useful. Consequently, Gibson and colleagues suggest taking more holistic and multifaceted approaches that can be used for product development, such as "basic user design principles". Considering the theories presented in the review section, a modular framework of media attributes is the closest thing to such design principles that exists today, even if many of the relationships remain unexplored.

It is hard to maintain the classical test theory perspective adopted by Morelli et al. (2017), which considers technology as a potential source of random error, in light of new technologies such as algorithm-based assessments of "new talent signals" like voice profiling. In contract, using technologies that recognize complex patterns of importance for predicting actual performance, such as advanced statistical procedures or machine learning, might be useful for explaining additional variance and go beyond traditional methods like observer ratings. Such new technologies will also be a challenge for the media attributes perspective, as algorithms per se are not a meaningful characteristic upon which testable hypotheses can be derived. In addition, as noted in the theory of socio-material practice (Orlikowski, 2007), the use of algorithms could result in very complex interactions between technology and human behavior. Theoretical clarification and further theory development in this field would provide a fruitful basis for further empirical investigations.

6.3.   Avenues for future research

In addition to the limitations and future research direction discussed above, the following section describes the need for future research and potential research designs that could be employed to answer important questions concerning the impact of technology in technology-mediated employment interviews. Specifically, this section will briefly review major shortcomings and open questions resulting from the presented studies before outlining potential study designs that address these unanswered issues.

Study 1 and Study 2 both investigated the issue of how specific constellations of media attributes effect interview ratings. While Study 1 found that ratings based on non-interactive video interviews achieve a sufficient level of interrater agreement, the unexpected finding from Study 2 was that no significant main effect occurred due to variations in social bandwidth. Social bandwidth was manipulated by providing raters with only a written transcript, voice recording, transcript with video, or the original recording including both sound and video. This finding was not in line with previous research reporting significant main effects of different levels of social bandwidth on the psychometric properties of interview ratings (e.g., Van Iddekinge et al., 2006). However, Van Iddekinge and colleagues compared ratings of FtF interviews vs. video recordings of the same interviews, meaning that the found differences can be attributed to several sources (the medium, interaction patterns, etc.). On the other hand, Study 2's inconsistent results could be due to the fact that the study design did not cover the full range of social bandwidth. If social bandwidth is considered as a continuum defined by the number of cues that can be transmitted at any given time, as proposed by Potosky (2008) and media richness theory (Daft & Lengel, 1984), face-to-face communication is the richest type of communication, followed by synchronous videoconferencing, followed by the configurations used in Study 2. Thus, future

research should include synchronous video-mediated interviews as well as face-to-face interviews, which were not included in the present study design. Another limitation of Study 1 and Study 2 was the optimization of the design with respect to social bandwidth manipulation by fixing the target variation to zero (by using a single fictitious interviewee as the rating target and manipulating answer quality within the same interview). One drawback of this design is the neglect of potential effects on the rank-order stability of multiple targets, referred to as interrater reliability (LeBreton & Senter, 2008). These open questions could be answered with a laboratory study that covers the full range of social bandwidth as a factor with at least 6 levels (FtF, SVI, AVI, TI, IVR, written transcript only) while holding content constant. It might be challenging to develop comparable content for interactive and non-interactive settings, but previous research has shown that it is possible to find ways to compare interactive and non-interactive selection procedures (e.g., Bauer et al., 2004). Moreover, data on candidates' computer self-efficiency, general mental ability, conscientiousness and ability to identify criteria should be collected as control variables and to test for moderation effects.

Study 4 addressed the impact of personal messaging on applicant reactions to asynchronous video interviews, but the results did not indicate a significant main effect. For many years, a common criticism of research on applicant reactions is that it primarily focuses on short-term reactions (Chan & Schmitt, 2004). However, this criticism refers to short-term reactions in the sense of candidates experiencing multiple rejections, rather than applicant reactions to new selection technologies as a function of acceptability perceptions, the time that has lapsed since the application came to market, and rate of distribution, based on the assumption that an assessment instrument's market share refers whether it has developed into a social norm in the sense of the theory of planned behavior (Ajzen, 1991). Consequently, a longitudinal

sample would be helpful to better understand the dynamics of how applicant reactions change as a function of time and the quality or usefulness of technological innovations.

Study 5 addressed the impact of preparation time on rating inflation. Participants who had the opportunity to prepare for up to two minutes prior to recording a video sequence received significantly higher ratings compared to participants who were not granted preparation time. However, an important open question concerns the impact of preparation time on validity, because the study does not indicate whether all candidates benefit from preparation time to the same extent or whether a specific population (e.g., top candidates) benefit more or less from preparation time, which would have an effect on criterion-related validity. To take a deeper look at this issue, a laboratory study could be conducted with a similar design as in Study 5. However, Study 5's dichotomous manipulation (preparation time vs. no preparation time) could be extended by introducing a second factor with two levels that refers to the procedural feature of disclosing all questions prior to the interview: full disclosure of all interview questions upfront (e.g., with the invitation to the study) vs. no disclosure. In addition, there are at least two more variables that will probably moderate the relationship between the preparation times and rating inflation in a sufficiently large sample size: conscientiousness and computer self-efficacy. In the literature, conscientiousness is associated with achievement striving, dependability, planning, and job performance (Barrick et al., 2001). With respect to preparation time in asynchronous video interviews, individuals who score high on conscientiousness will probably prepare for the recording more accurately and effectively or spend more time developing optimal answers when they know the questions ahead of the interview. Computer self-efficacy refers to an individual's judgment of their capabilities to use computers and is positively associated with actual computer usage (Compeau & Higgins, 1995). In this context, individuals who score high on computer self-

efficacy will probably find it easier to adjust to the situation and prepare more effectively to record their answers in a technical sense. For example, such individuals might know how to adjust their video or audio equipment to show them in a better light or how to browse quickly through potentially useful sources. The emergence of a positive moderator relationship would affect fairness concerns with regard to such assessment procedures, as specific subgroups (e.g., older applicants) score lower on such constructs on average. Another issue that should be addressed in this context is the question of which tactics candidates actually apply that lead to rating inflation. The development of a questionnaire similar to Levashina and Campion (2007) for asynchronous video interviews could be a fruitful approach here. Finally, a measure of actual workplace performance (e.g., supervisor ratings) should be used to take a deeper look at whether the validity of the procedure actually changes or the correlations remain the same.

6.4.　　Implication for practice

The importance of this research field for selection practitioners has already been highlighted at several points in this dissertation. Nevertheless, in this section, implications for practice will be discussed and structured with respect to different selection outcomes and underlying questions practitioners may have. First of all, the review section reported considerable evidence of significant differences between face-to-face and technology-enhanced employment interview formats with respect to reliability, validity, applicant reactions, and potential adverse impact. Thus, organizations should be very careful when shifting from face-to-face interviews to any form of technology-enhanced interviewing.

Reliability. The crucial question practitioners might ask when deciding whether or not to adopt a new selection technology is whether they will have to sacrifice accurate judgments for the sake of efficiency (Blackman, 2002). As was pointed out in the review section, a high level of

structure is the best-known way of ensuring high levels of reliability in its various conceptualizations, including agreement and accuracy (Huffcutt et al., 2014). With respect to new technologies like asynchronous video interviews, Study 1 found that further increasing the level of structure by including a high level of response standardization is still effective for increasing the level of interrater agreement and accuracy in a constellation that already incorporates the highest levels of question and process standardization. As suggested by Levashina et al. (2014), a high level of structure is beneficial for phone or video interviews used for screening due to a better fit between the communication medium and the complexity of information that needs to be processed. The results presented here provide further evidence of the benefit of investing in the development of anchored rating scales in asynchronous video interviews, as they can significantly improve interrater agreement and reduce rating idiosyncrasies, which justifies the higher costs associated with developing such anchored rating scales (S. D. Maurer, 2002; Melchers et al., 2011).

Study 2 found no significant effects of the social bandwidth used for evaluation or of the construct domain assessed in different questions on rater accuracy in highly structured interviews, contrary to the developed hypotheses. Thus, this dissertation's results are in line with previous research identifying structure as very effective for achieving high levels of reliability; moreover, the results contribute to the literature by providing evidence that high levels of structure limit the potential effects of contamination by several media attributes to a minimum under the discussed methodological limitations.

Validity. As reliability sets the upper limit for validity, the points discussed in the last paragraph can be seen as evidence that technology-enhanced interview formats possess the necessary foundation in terms of reliability to enable valid assessments. In addition, this thesis

provides further important knowledge for selection practitioners by presenting the very first results from a field sample on the validity of asynchronous video interviews in high-stakes selection. The very few previous studies on this topic were either conducted in a laboratory setting (Gorman et al., 2018) or did not assess the relationship between AVI and criterion variables (Langer et al., 2017). First, Study 3 revealed that asynchronous video interviews in selection are substantially and significantly correlated with behavioral outcomes in assessment centers. This result is quite promising, as it provides evidence for criterion-related validity in the field with results that are in line with previous findings from a non-applicant sample (Gorman et al., 2018). One limitation is that there is still no research available investigating the relationship between the results of asynchronous video interviews and overall job performance.

The second important result of this study is the moderate correlation with cognitive ability that emerged for asynchronous video interviews but not face-to-face interviews. These results were quite surprising, as previous research found that response formats higher in social bandwidth are lower in cognitive saturation (Lievens et al., 2015). Third, the field study uncovered a significant but moderate correlation between asynchronous video interviews and face-to-face interviews. The small effect size of this result actually indicates that a large portion of variance is due to the strong differences in the alignment of media attributes in these two holistic classes of interviews. Thus, more research is needed here before meaningful guidelines for practice can be derived.

Finally, Study 5 found that a relatively small adjustment to the interview process in asynchronous video interviews influenced interview ratings with a quite substantial effect size. Unfortunately, more research is needed to test whether the differences in mean scores are a threat

to validity (Day & Carroll, 2003). Nevertheless, practitioners should be aware of the potential impacts of process adjustments that might seem minor at first glance.

Applicant reactions. Study 4 predicted that a social bandwidth manipulation would influence applicant reactions in a non-interactive procedure, but failed to confirm this hypothesis. Beyond the methodological limitations discussed above, this result also has important implications for practice. First of all, the overall levels of the different procedural justice dimensions were comparable to those of other non-interactive selection procedures like cognitive testing (Anderson et al., 2010). Thus, it seems that interview formats with no direct interactivity are considered more like initial tests, meaning that it could be very problematic to substitute face-to-face employment interviews in later selection stages with asynchronous, non-interactive ones. Gilliland (1993) already discussed process stage as an important moderator of perceived fairness, as candidates do not expect high levels of direct interaction in early stages of the selection process. On the other hand, the same argument means that candidates do expect direct interaction in other stages. Practitioners should also consider potential differences among subgroups of applicants. For example, using too much technology be might be problematic for individuals with low-computer self-efficacy, or may produce adverse impact when computer-self-efficacy is related to factors such as age. Previous research has also found that interviews are considered fundamental for job search success (Saks, 2006). Thus, the use of new technologies, especially asynchronous technologies that differ most drastically from individuals' expectations regarding how interviews typically look like and feel like, might also be problematic for subgroups like the long-term unemployed. Overall, practitioners should be aware that technology usage in interview processes has a strong impact on applicant reactions and should consider the context, including selection stage, selection purpose, and candidate pool heterogeneity, before

employing such technologies. In early selection stages with a high selection focus and potentially small candidate pool heterogeneity, like that found for entry-level positions for university graduates, highly technology-enhanced interview forms will be more accepted than in late selection stages with a strong recruiting focus and a heterogeneous candidate pool.

Adverse impacts. The avoidance of adverse impact is essential for practitioners due to potential lawsuits and subsequent reputation damage that can occur in the event that selection systems produce systematic biases against specific subgroups of candidates. The issue of potential discrimination due to race and gender was previously addressed in a laboratory study, which found no evidence that video-recorded interview ratings are vulnerable to adverse impact (Kroll & Ziegler, 2016). The studies presented here did not address this topic directly but repeatedly used control variables to identify any adverse impacts due to age or gender. Overall, a closer look at these control variables did not reveal any direct age or gender effects. Nevertheless, the data revealed some first-level correlations that might indirectly lead to adverse impact. For example, in line with previous research (e.g., Jorm, 1987) Study 4 revealed a first-level correlation between gender and neuroticism. As neuroticism is associated with more perceived stress in unfamiliar situations and negative adaptation (Costa & McCrae, 1992), such systematic mean differences might indirectly produce adverse impacts when (a) a trait acts as a moderator, e.g., by moderating the relationship between procedural justice dimensions and intention to accept a job offer, and (b) when a construct rating is contaminated by a higher or lower personality saturation due to media usage. Apart from these topics, which require in-depth further research, the results of the studies presented in this dissertation do not indicate to practitioners that technology usage in employment interviews leads to systematic adverse impacts per se.

Overall utility. As reviewed in the first part of this dissertation, overall utility considerations depend on various input variables (Cascio & Boudreau, 2011), which were not covered in detail by studies presented above. Nevertheless, some of these aspects were covered, providing some important implications for practice. First, Study 3 is the first study to present important evidence for the predictive and divergent validity of asynchronous video interviews in a field sample in a high-stakes selection process, one of the most important determinants of overall utility. The intercorrelations with other selection procedures can also provide some initial guidance for constructing multi-stage and multi-method selection systems, which are highly dependent on these intercorrelations (Sackett & Roth, 1996). The main findings were quite promising for the usage of asynchronous video technology in pre-selection. Second, Study 5 revealed potential stumbling blocks in designing interview processes with technology, which ultimately has an influence on overall utility. Finally, the conclusions drawn with respect to applicants reactions are important, as even well-designed and valid interview processes can decrease overall utility when they are not accepted by candidates because they are considered too invasive of privacy and cause top candidates to withdraw from the interview process (K. R. Murphy, 1986).

General perspective on selection procedures. Organizations could benefit from adopting a modular view of technology-enhanced interviews and interview features when designing or redesigning their recruitment processes. The working framework of media attributes based on Potosky (2008) and media synchronicity theory (Dennis et al., 2008) could increase organizations' awareness of the consequences of changing one or more design principles used in the interview process that affect one or more of the described dimensions. For example, an organization might need to optimize adverse impact and applicant reactions to the use of

technology high in interaction, contextualization and personalization in early pre-selection. On

the other hand, when saving time and costs are essential and the pool of applicants is much larger

than the number of positions to be filled, less interactivity and personalization might be adequate.

Overall, the practical implications discussed in this section should help to usher in a more

nuanced view of potential impacts of technology usage in employment interviews, even though

much more research is still required to catch up to the latest developments in the field. In the end,

this research should help to maximize the positive effects of using technology in selection

systems for individuals and organizations, while minimizing the negative effects (Cascio &

Montealegre, 2016).

6.5.   Concluding comments

This review and the five presented empirical studies importantly add to our understanding

of the impact of technology in employment interviews and make an empirical contribution to the

literature by providing empirical results on the psychometric properties of asynchronous video

interviews. One major limitation of the theories presented above is the question of how to treat

future applications, which are often based on advances in software rather than hardware or

physically observable media attributes. This is obviously the case for technologies like in-depth

machine learning, self-learning algorithms, or virtual reality applications. Media attribute

frameworks would have to be extended to include a theoretically unlimited number of software-

dependent dimensions. Another obstacle is a lack of understanding of the environmental drivers

that cause recruiters to adapt new technologies.

From a methodological perspective, there are also alternative approaches to studying the

impact of technology on interviews conducted in laboratory or field studies, in which the impact

of the actual interaction process between interviewees and interviewers remains a black box. One

alternative useful approach to the process of adopting new technology in selection comes from role theory (Barley, 2015; Cascio & Montealegre, 2016). The main argument of role theory is that technology triggers changes in what tasks are performed, the way these tasks are performed, and even the nature of interactions. A corresponding research design applied to the interview context might include, analogously to Barley (2015), the systematic documentation of repetitive behavioral patterns – for example, during the implementation of automated interview solutions over time – in order to investigate how the roles and interaction patterns between HR representatives and candidates change.

6.6.   Conclusions

In recent decades, considerable research efforts have been expended on understanding the impacts of communication technology and making its impacts or usage more predictable (Daft & Lengel, 1986; Dennis et al., 2008; Short et al., 1976). Moreover, as technological progress has led to massive changes in the recruitment and selection domain, more specific frameworks have been developed for this field (Arthur et al., 2018; Potosky, 2008). In this work, I reviewed the existing literature on technology in employment interviews and conducted multiple studies to test specific hypotheses about the psychometric properties of asynchronous video interviews. My integrated review argues that a modular approach is more adequate for understanding nuanced differences between assessment technologies, as distinct categories or holistic views cannot adequately represent the multiple interviewing apps in today's workplace. Thus, on the basis of the existing literature, it was suggested that the media attributes of transparency, social bandwidth, interactivity, surveillance, parallelism, rehearsability, and reprocessability have distinct impacts on interview reliability, validity and applicant and recruiter reactions.

The results obtained from five unique samples on asynchronous video interviews found that these non-interactive formats possess a sufficient level of interrater agreement as well as promising findings regarding criterion validity and incremental validity over cognitive ability tests. Surprisingly, rating accuracy did not depend on the social bandwidth of the media format used to rate recorded answers in a highly structured format. Next, the introduction of simple personal video messages did not affect procedural justice perceptions, even though organizations might profit from candidates' perceptions. Finally, rehearsability, which was manipulated by providing candidates with preparation time before recording their answers, as is typically the case in asynchronous video interviews, raise some concerns about rating inflation, but more research is needed on the processes behind this phenomenon and its potential impacts on criterion and construct validity. Academics and practitioners alike are encouraged to conduct further research in the field of technology usage in employment interviews, as this area is still highly understudied in relation to its considerable relevance for practice.

Overall, this dissertation makes several contributions to the existing literature on technology usage in selection: First, the review section provides an updated and comprehensive review of research practice on technology usage in employment interviews and integrates several theoretical directions in proposing a more nuanced, modular conceptual framework that allows for the development of testable hypotheses on factors beyond the measurement equivalence of different broad classes of conceptualizations like telephone interviews vs. video interviews. Second, the five empirical studies presented in the second part of this dissertation provide in-depth knowledge on the psychometric properties of asynchronous video interviews, where research is still in its infancy and only a few investigations exist (Gorman et al., 2018; Langer et al., 2017). Third, the general discussion highlights potential avenues for further research to

increase our knowledge on the impact of technology on selection and hopefully provides

practitioners with some additional guidance on using technology in employment interviews. As a

final conclusion, both the review and empirical studies reveal that the ongoing technological

revolution in the I/O domain involves more than just bringing highly specialized applications to

the market. It involves complex interactions between technological attributes, structural

interview components, assessment dimensions, and the organizational and social environments in

which selection systems are embedded. Future research should find a balance between

complexity reduction, generalizability and applicability when attempting to answer the question

of how and why new technology alters selection outcomes. Finally, the critical importance of

structure for interviews' psychometric properties was confirmed once more, regardless of

whether interviews are conducted face-to-face, via synchronous video devices, or even

asynchronously.

References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179-211. https://doi.org/10.1016/0749-5978(91)90020-T

Albarracin, D., Johnson, B. T., Fishbein, M., & Muellerleile, P. A. (2001). Theories of reasoned action and planned behavior as models of condom use: A meta-analysis. *Psychological Bulletin, 127*(1), 142-161. https://doi.org/10.1037//0033-2909.127.1.142

Albarracin, D., & Shavitt, S. (2018). Attitudes and attitude change. *Annual Review of Psychology, Vol 69, 69*, 299-327. https://doi.org/10.1146/annurev-psych-122216-011911

Allen, D. G., Mahto, R. V., & Otondo, R. F. (2007). Web-based recruitment: Effects of information, organizational brand, and attitudes toward a website on applicant attraction. *Journal of Applied Psychology, 92*(6), 1696-1708. https://doi.org/10.1037/0021-9010.92.6.1696

Ambady, N., Bernieri, F. J., & Richeson, J. A. (2000). Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. *Advances in Experimental Social Psychology, 32*, 201-271.

Ambady, N., Krabbenhoft, M. A., & Hogan, D. (2006). The 30-sec sale: Using thin-slice judgments to evaluate sales effectiveness. *Journal of Consumer Psychology, 16*(1), 4-13.

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal aonsequences: A metaanalysis. *Psychological Bulletin, 111*(2), 256-274. https://doi.org/10.1037/0033-2909.111.2.256

Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology, 64*(3), 431.

Anderson, N. (2003). Applicant and recruiter reactions to new technology in selection: A critical review and agenda for future research. *International Journal of Selection and Assessment, 11*(2-3), 121-136. https://doi.org/10.1111/1468-2389.00235

Anderson, N., Salgado, J. F., & Hulsheger, U. R. (2010). Applicant reactions in selection: Comprehensive meta-analysis into reaction generalization versus situational specificity. *International Journal of Selection and Assessment, 18*(3), 291-304.

Arthur, W., Jr., & Glaze, R. (2011). Cheating and response distortion of remotely delievered assessments. In N. Tippins & S. Adler (Eds.), *Technology-enhanced asessments of talent* (pp. 99-152). San Francisco, CA: Jossey-Bass.

Arthur, W., Jr., Keiser, N. L., & Doverspike, D. (2018). An information-processing-based conceptual framework of the effects of unproctored internet-based testing devices on scores on employment-related assessments and tests. *Human Performance, 31*(1), 1-32. https://doi.org/10.1080/08959285.2017.1403441

Arvey, R. D., & Campion, J. E. (1982). The employment interview: A summary and review of recent research. *Personnel Psychology, 35*(2), 281-322. https://doi.org/10.1111/j.1744-6570.1982.tb02197.x

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, Vol 63, 63*, 1-29. https://doi.org/10.1146/annurev-psych-120710-100422

Banerjee, P. (2016). Attracting job-seekers through online job advertisements: Application of RJPs, blogs and video podcasts. *International Journal of Technology and Human Interaction (IJTHI), 12*(3), 1-17. https://doi.org/10.4018/IJTHI.2016070101

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

Barley, S. R. (2015). Why the Internet makes buying a car less loathsome: How technologies change role relations. *Academy of Management Discoveries, 1*(1), 5-35. https://doi.org/10.5465/amd.2013.0016

Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment, 9*(1-2), 9-30. https://doi.org/10.1111/1468-2389.00160

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

Barrick, M. R., Swider, B. W., & Stewart, G. L. (2010). Initial evaluations in the interview: Relationships with subsequent interviewer evaluations and employment offers. *Journal of Applied Psychology, 95*(6), 1163-1172. https://doi.org/10.1037/a0019918

Bartels, L. K., & Nordstrom, C. R. (2013). Too big to hire: Factors impacting weight discrimination. *Management Research Review, 36*(9), 868-881.

Bartram, D. (2000). Internet recruitment and selection: Kissing frogs to find princes. *International Journal of Selection and Assessment, 8*(4), 261-274. https://doi.org/10.1111/1468-2389.00155

Basil, M. D. (2012). Multiple resource theory. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 2384-2385). Boston, MA: Springer US.

Batrinca, L. M., Mana, N., Lepri, B., Pianesi, F., & Sebe, N. (2011). *Please, tell me about yourself: Automatic personality assessment using short self-presentations*. Paper

presented at the ICMI '11 Proceedings of the 13th international conference on multimodal

interfaces, New York, NY, USA.

Bauer, T. N., Maertz, C. P., Dolen, M. R., & Campion, M. A. (1998). Longitudinal assessment of

applicant reactions to employment testing and test outcome feedback. *Journal of Applied*

Psychology, 83(6), 892-903. https://doi.org/10.1037//0021-9010.83.6.892

Bauer, T. N., Truxillo, D. M., Paronto, M. E., Weekley, J. A., & Campion, M. A. (2004).

Applicant reactions to different selection technology: Face-to-face, interactive voice

response, and computer-assisted telephone screening interviews. *International Journal of*

Selection and Assessment, 12(1-2), 135-148. https://doi.org/10.1111/j.0965-

075X.2004.00269.x

Bauer, T. N., Truxillo, D. M., Sanchez, R. J., Craig, J. M., Ferrara, P., & Campion, M. A. (2001).

Applicant reactions to selection: Development of the Selection Procedural Justice Scale

(SPJS). *Personnel Psychology, 54*(2), 387-419. https://doi.org/10.1111/j.1744-

6570.2001.tb00097.x

Bauer, T. N., Truxillo, D. M., Tucker, J. S., Weathers, V., Bertolino, M., Erdogan, B., &

Campion, M. A. (2006). Selection in the information age: The impact of privacy concerns

and computer experience on applicant reactions. *Journal of Management, 32*(5), 601-621.

https://doi.org/10.1177/0149206306289829

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the

active self a limited resource? *Journal of Personality and Social Psychology, 74*(5),

1252-1265. https://doi.org/10.1037/0022-3514.74.5.1252

Beach, L. R. (1993). Broadening the definition of decision-making: The role of prechoice screening of options. *Psychological Science, 4*(4), 215-220. https://doi.org/10.1111/j.1467-9280.1993.tb00264.x

Behrend, T., Toaddy, S., Thompson, L. F., & Sharek, D. J. (2012). The effects of avatar appearance on interviewer ratings in virtual employment interviews. *Computers in Human Behavior, 28*(6), 2128-2133. https://doi.org/10.1016/j.chb.2012.06.017

Berry, C. M., Sackett, P. R., & Landers, R. N. (2007). Revisiting interview-cognitive ability relationships: Attending to specific range restriction mechanisms in meta-analysis. *Personnel Psychology, 60*(4), 837-874. https://doi.org/10.1111/j.1744-6570.2007.00093.x

Blackman, M. C. (2002). The employment interview via the telephone: Are we sacrificing accurate personality judgments for cost efficiency? *Journal of Research in Personality, 36*(3), 208-223. https://doi.org/10.1006/jrpe.2001.2347

Blacksmith, N., & Poeppelman, T. (2014). Video-based technology: The next generation of recruitment and hiring. *TIP: The Industrial-Organizational Psychologist, 52*(2), 84-88.

Blacksmith, N., Willford, J. C., & Behrend, T. S. (2016). Technology in employment interviews: A meta-analysis and future research agenda. *Personnel Assessment and Decisions:, 2*(1), 12-20. https://doi.org/10.25035/pad.2016.002

Bolino, M. C., Kacmar, K. M., Turnley, W. H., & Gilstrap, J. B. (2008). A multi-level review of impression management motives and behaviors. *Journal of Management, 34*(6), 1080-1109. https://doi.org/10.1177/0149206308324325

Borkenau, P., Mauer, N., Riemann, R., Spinath, F. M., & Angleitner, A. (2004). Thin slices of behavior as cues of personality and intelligence. *Journal of Personality and Social Psychology, 86*(4), 599-614. https://doi.org/10.1037/0022-3514.86.4.599

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.* https://doi.org/10.3389/fpsyg.2016.00863

Burke, M. J., & Dunlap, W. P. (2002). Estimating interrater agreement with the average deviation index: A user's guide. *Organizational Research Methods, 5*(2), 159-172. https://doi.org/10.1177/1094428102005002002

Burke, M. J., Finkelstein, L. M., & Dusig, M. S. (1999). On average deviation indices for estimating interrater agreement. *Organizational Research Methods, 2*(1), 49-68. https://doi.org/10.1177/109442819921004

Burnett, J. R., & Motowidlo, S. J. (1998). Relations between different sources of information in the structured selection interview. *Personnel Psychology, 51*(4), 963.

Cable, D. M., & Judge, T. A. (1997). Interviewers' perceptions of person-organization fit and organizational selection decisions. *Journal of Applied Psychology, 82*(4), 546-561. https://doi.org/10.1037/0021-9010.82.4.546

Cammio. (2019). Our platform. Retrieved from https://cammio.com/our-platform/

Camp, R., Schulz, E., Vielhaber, M., & Wagner-Marsh, F. (2004). Human resource professionals' perceptions of team interviews. *Journal of Managerial Psychology, 19*(5), 490-505. https://doi.org/10.1108/02683940410543588

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81-105. https://doi.org/10.1037/h0046016

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

Carney, D. R., Colvin, C. R., & Hall, J. A. (2007). A thin slice perspective on the accuracy of first impressions. *Journal of Research in Personality, 41*(5), 1054-1072. https://doi.org/10.1016/j.jrp.2007.01.004

Cascio, W. F., & Boudreau, J. W. (2011). *Investing in People. Financial impact of human resource initiatives* (2 ed.). Upper Saddle River, NJ: Pearson.

Cascio, W. F., & Montealegre, R. (2016). How technology is changing work and organizations. *Annual Review of Organizational Psychology and Organizational Behavior, Vol 3, 3*, 349-375. https://doi.org/10.1146/annurev-orgpsych-041015-062352

Chamorro-Premuzic, T., Winsborough, D., Sherman, R. A., & Hogan, R. (2016). New talent signals: Shiny new objects or a brave new world? *Industrial and Organizational Psychology-Perspectives on Science and Practice, 9*(3), 621-640. https://doi.org/10.1017/iop.2016.6

Chan, D., & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology, 82*(1), 143-159. https://doi.org/10.1037/0021-9010.82.1.143

Chan, D., & Schmitt, N. (2004). An agenda for future research on applicant reactions to selection procedures: A construct-oriented approach. *International Journal of Selection and Assessment, 12*(1-2), 9-23. https://doi.org/10.1111/j.0965-075X.2004.00260.x

Chapman, D. S., & Rowe, P. M. (2001). The impact of videoconference technology, interview structure, and interviewer gender on interviewer evaluations in employment interviews: A field experiment. *Journal of Occupational and Organizational Psychology, 74*, 279-298. https://doi.org/10.1348/096317901167361

Chapman, D. S., & Rowe, P. M. (2002). The influence of videoconference technology and interview structure on the recruiting function of the employment interview: A field experiment. *International Journal of Selection and Assessment, 10*(3), 185-197. https://doi.org/ 10.1111/1468-2389.00208

Chapman, D. S., Uggerslev, K. L., & Webster, J. (2003). Applicant reactions to face-to-face and technology-mediated interviews: A field investigation. *Journal of Applied Psychology, 88*(5), 944-953. https://doi.org/10.1037/0021-9010.88.5.944

Chapman, D. S., & Webster, J. (2001). Rater correction processes in applicant selection using videoconference technology: The role of attributions. *Journal of Applied Social Psychology, 31*(12), 2518-2537.

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

Chen, C. C., Wen‐Fen Yang, I., & Lin, W. C. (2010). Applicant impression management in job interview: The moderating role of interviewer affectivity. *Journal of Occupational and Organizational Psychology, 83*(3), 739-757.

Chen, Y.-Y., & Fang, W. (2008). The moderating effect of impression management on the organizational politics–performance relationship. *Journal of Business Ethics, 79*(3), 263-277.

Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational judgment tests: Constructs assessed and a meta‐analysis of their criterion‐related validities. *Personnel Psychology, 63*(1), 83-117.

Compeau, D. R., & Higgins, C. A. (1995). Computer self-efficacy: Development of a measure and initial test. *Mis Quarterly, 19*(2), 189-211. https://doi.org/10.2307/249688

Conway, J. M., Jako, R. A., & Goodman, D. F. (1995). A meta-analysis of interrater and internal consistency reliability of selection interviews. *Journal of Applied Psychology, 80*(5), 565-579.

Cooper, R., & Foster, M. (1971). Sociotechnical Systems. *American Psychologist, 26*(5), 467-474. https://doi.org/10.1037/h0031539

Cornerstone OnDemand Ltd. (2019). Cornerstone + Hirevue. Retrieved from https://www.cornerstoneondemand.com/marketplace/hirevue

Cortina, J. M., Goldstein, N. B., Payne, S. C., Davison, H. K., & Gilliland, S. W. (2000). The incremental validity of interview scores over and above cognitive ability and conscientiousness scores. *Personnel Psychology, 53*(2), 325-351. https://doi.org/10.1111/j.1744-6570.2000.tb00204.x

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor (NEO-FFI) Inventory professional manual*. Odessa, FL: Psychological Assessment Resources.

CU-SeeMe Development Team. (1995). CU SeeMe. Retrieved from http://ftp.icm.edu.pl/packages/cu-seeme/html/Welcome.html

Cuddy, A. J. C., Wilmuth, C. A., Yap, A. J., & Carney, D. R. (2015). Preparatory Power Posing Affects Nonverbal Presence and Job Interview Performance. *Journal of Applied Psychology, 100*(4), 1286-1295. https://doi.org/10.1037/a0038543

cut-e. (2018a). *scales lst*. *Quality quaracteristics*. cut-e Group. Hamburg.

cut-e. (2018b). *scales numerical & verbal*. *Quality characteristics*. cut-e Group. Hamburg.

cut-e. (2019). Data-driven video assessment. Retrieved from https://www.cut-e.com/solutions/video-assessment/

Daft, R. L., & Lengel, R. H. (1984). Information richness: A new approach to managerial behavior and organizational design. *Research in Organizational Behavior, 6*, 191-233.

Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Management Science, 32*(5), 554-571. https://doi.org/10.1287/mnsc.32.5.554

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *Mis Quarterly, 13*(3), 319-340. https://doi.org/10.2307/249008

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer-technology: A comparison of 2 theoretical models. *Management Science, 35*(8), 982-1003.

Day, A. L., & Carroll, S. A. (2003). Situational and patterned behavior description interviews: A comparison of their validity, correlates, and perceived fairness. *Human Performance, 16*(1), 25-47. https://doi.org/10.1207/S15327043hup1601_2

Deary, I. J. (2012). Intelligence. *Annual Review of Psychology, Vol 63, 63*, 453-482. https://doi.org/10.1146/annurev-psych-120710-100353

DeGroot, T., & Gooty, J. (2009). Can nonverbal cues be used to make meaningful personality attributions in employment interviews? *Journal of Business and Psychology, 24*(2), 179-192. https://doi.org/10.1007/s10869-009-9098-0

DeGroot, T., & Kluemper, D. (2007). Evidence of predictive and incremental validity of personality factors, vocal attractiveness and the situational interview. *International Journal of Selection and Assessment, 15*(1), 30-39.

DeGroot, T., & Motowidlo, S. J. (1999). Why visual and vocal interview cues can affect interviewers' judgments and predict job performance. *Journal of Applied Psychology, 84*(6), 986-993. https://doi.org/10.1037/0021-9010.84.6.986

Dennis, A. R., Fuller, R. M., & Valacich, J. S. (2008). Media, tasks, and communication processes: A theory of media synchronicity. *Mis Quarterly, 32*(3), 575-600.

Dennis, A. R., & Valacich, J. S. (1999). *Rethinking media richness: Towards a theory of media synchronicity*. Paper presented at the 32nd Hawaii International Conference on System Sciences, Hawaii.

Dipboye, R. L. (1982). Self-Fulfilling prophecies in the selection-recruitment interview. *Academy of Management Review, 7*(4), 579-586. https://doi.org/10.5465/AMR.1982.4285247

Dipboye, R. L. (1994). Structured and unstructured selection interviews: Beyond the job-fit model. *Research in personnel and human resources management, 12*, 79-123.

Donovan, J. J., Dwight, S. A., & Schneider, D. (2014). The impact of applicant faking on selection measures, hiring decisions, and employee performance. *Journal of Business and Psychology, 29*(3), 479-493. https://doi.org/10.1007/s10869-013-9318-5

Eid, M., Gollwitzer, M., & Schmitt, M. (2011). *Statistik und Forschungsmethoden Lehrbuch*. Weinheim: Beltz.

Ellis, A. P. J., West, B. J., Ryan, A. M., & DeShon, R. P. (2002). The use of impression management tactics in structured interviews: A function of question type? *Journal of Applied Psychology, 87*(6), 1200-1208. https://doi.org/10.1037//0021-9010.87.6.1200

Fast, L. A., & Funder, D. C. (2008). Personality as manifest in word use: Correlations with self-report, acquaintance report, and behavior. *Journal of Personality and Social Psychology, 94*(2), 334-346. https://doi.org/10.1037/0022-3514.94.2.334

Feingold, A. (1992). Good-looking people are not what we think. *Psychological Bulletin, 111*(2), 304-341. https://doi.org/10.1037//0033-2909.111.2.304

Fenn, J., & Raskino, M. (2008). Mastering the Hype Cycle: How to choose the right innovation at the right time. *Harvard Business School Press Books*, 1.

Frauendorfer, D., & Mast, M. S. (2015). The impact of nonverbal behavior in the job interview. In *The Social Psychology of Nonverbal Communication* (pp. 220-247): Springer.

Frieder, R. E., Van Iddekinge, C. H., & Raymark, P. H. (2016). How quickly do interviewers reach decisions? An examination of interviewers' decision-making time across applicants. *Journal of Occupational and Organizational Psychology, 89*(2), 223-248. https://doi.org/10.1111/joop.12118

Funder, D. C., & Colvin, C. R. (1988). Friends and strangers: acquaintanceship, agreement, and the accuracy of personality judgment. *Journal of Personality and Social Psychology, 55*(1), 149.

Funder, D. C., & Dobroth, K. M. (1987). Differences between traits: properties associated with interjudge agreement. *Journal of Personality and Social Psychology, 52*(2), 409.

Funke, U., & Schuler, H. (1998). Validity of stimulus and response components in a video test of social competence. *International Journal of Selection and Assessment*, 115-123.

Gelléri, P., & Kiefer, C. (2018, 17.09.2018). *Krch …Krch … [Bildstörung] – Der Einfluss von technischen Störungen in Videointerviews zur Personalauswahl*. Paper presented at the Vortrag auf dem 51. Kongress der Deutschen Gesellschaft für Psychologie (DGPS), Frankfurt.

Gibson, C., Vaughn, D., & Hudy, M. (2017). Beyond empirical equivalence. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 10*(4), 676-680. https://doi.org/10.1017/iop.2017.75

Gilliland, S. W. (1993). The perceived fairness of selection systems: An organizational justice perspective. *Academy of Management Review, 18*(4), 694-734.

Gorman, C. A., Robinson, J., & Gamble, J. S. (2018). An investigation into the validity of asynchronous web-based video employment-interview ratings. *Consulting Psychology Journal: Practice and Research, 70*(2), 129-146. https://doi.org/10.1037/cpb0000102

Graves, L. M., & Karren, R. J. (1992). Interviewer decision-processes and effectiveness: An experimental policy-capturing investigation. *Personnel Psychology, 45*(2), 313-340.

Greb, C., & Linnenbürger, A. (2013). *Ein psycholinguistisches IT-Programm für Diagnostik, Beratung & Coaching: Grundlagen, Anwendung und weitere Entwicklung*. Psyware GmbH. Aachen.

Guchait, P., Ruetzler, T., Taylor, J., & Toldi, N. (2014). Video interviewing: A potential selection tool for hospitality managers - A study to understand applicant perspective. *International Journal of Hospitality Management, 36*, 90-100. https://doi.org/10.1016/j.ijhm.2013.08.004

Guilford, J. P. (1954). *Psychometric methods* (2d ed.). New York,: McGraw-Hill.

Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.

Hamdani, M. R., Valcea, S., & Buckley, M. R. (2014). The relentless pursuit of construct

validity in the design of employment interviews. *Human Resource Management Review*,

24(2), 160-176. https://doi.org/10.1016/j.hrmr.2013.07.002

Harris, M. M. (1989). Reconsidering the employment interview: A review of recent literature

and suggestions for future-research. *Personnel Psychology, 42*(4), 691-726.

https://doi.org/ 10.1111/j.1744-6570.1989.tb00673.x

Harris, M. M. (1999). What is being measured? In R. W. Eder & M. M. Harris (Eds.), *The*

employment interview handbook (pp. 143-157). Thousand Oaks, CA: Sage.

Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection

procedures: An updated model and meta-analysis. *Personnel Psychology, 57*(3), 639-683.

https://doi.org/10.1111/J.1744-6570.2004.00003.X

Herrnstein, R. J., & Murray, C. A. (1994). *The bell curve: Intelligence and class structure in*

American life. New York: Free Press.

Hiemstra, A. M. F., Derous, E., Serlie, A. W., & Born, M. P. (2012). Fairness perceptions of

video resumes among ethnically diverse applicants. *International Journal of Selection &*

Assessment, 20(4), 423-433. https://doi.org/10.1111/ijsa.12005

HireVue. (2019). HireVue Assessments. Retrieved from

https://www.hirevue.com/products/assessments

Hirschmuller, S., Egloff, B., Nestler, S., & Back, M. D. (2013). The dual lens model: A

comprehensive framework for understanding self-other agreement of personality

judgments at zero acquaintance. *Journal of Personality and Social Psychology, 104*(2), 335-353. https://doi.org/10.1037/a0030383

Horn, R. G., & Behrend, T. S. (2017). Video killed the interview star: Does picture-in-picture affect interview performance? *Personnel Assessment and Decisions:, 3*(1), 51-59. https://doi.org/10.25035/pad.2017.005

Hough, L. M. (1998). Effects of intentional distortion in personality measurement and evaluation of suggested palliatives. *Human Performance, 11*(2-3), 209-244. https://doi.org/10.1207/s15327043hup1102&3_6

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling-a Multidisciplinary Journal, 6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Huffcutt, A. I. (2011). An empirical review of the employment interview construct literature. *International Journal of Selection and Assessment, 19*(1), 62-81.

Huffcutt, A. I., & Arthur, W., Jr. (1994). Hunter and Hunter (1984) Revisited - interview validity for entry-level jobs. *Journal of Applied Psychology, 79*(2), 184-190. https://doi.org/10.1037/0021-9010.79.2.184

Huffcutt, A. I., Conway, J. M., Roth, P. L., & Klehe, U.-C. (2004). The impact of job complexity and study design on situational and behavior description interview validity. *International Journal of Selection and Assessment, 12*(3), 262-273. https://doi.org/10.1111/j.0965-075X.2004.280_1.x

Huffcutt, A. I., Conway, J. M., Roth, P. L., & Stone, N. J. (2001). Identification and meta-analytic assessment of psychological constructs measured in employment interviews.

Journal of Applied Psychology, 86(5), 897-913. https://doi.org/10.1037//0021-9010.86.5.897

Huffcutt, A. I., Culbertson, S. S., & Weyhrauch, W. S. (2013). Employment interview reliability: New meta-analytic estimates by structure and format. *International Journal of Selection and Assessment, 21*(3), 264-276. https://doi.org/10.1111/ijsa.12036

Huffcutt, A. I., Culbertson, S. S., & Weyhrauch, W. S. (2014). Moving forward indirectly: Reanalyzing the validity of employment interviews with indirect range restriction methodology. *International Journal of Selection and Assessment, 22*(3), 297-309. https://doi.org/10.1111/ijsa.12078

Huffcutt, A. I., & Roth, P. L. (1998). Racial group differences in employment interview evaluations. *Journal of Applied Psychology, 83*(2), 179-189. https://doi.org/10.1037/0021-9010.83.2.179

Huffcutt, A. I., Roth, P. L., & McDaniel, M. A. (1996). A meta-analytic investigation of cognitive ability in employment interview evaluations: Moderating characteristics and implications for incremental validity. *Journal of Applied Psychology, 81*(5), 459-473. https://doi.org/10.1037/0021-9010.81.5.459

Huffcutt, A. I., & Youngcourt, S. S. (2007). Employment interviews. In D. L. Whetzel & G. R. Wheaton (Eds.), *Applied measurement: Industrial psychology in human resource management* (pp. 181–199). London: Psychology Press.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job-performance. *Psychological Bulletin, 96*(1), 72-98. https://doi.org/10.1037//0033-2909.96.1.72

Imada, A. S., & Hakel, M. D. (1977). Influence of nonverbal communication and rater proximity on impressions and decisions in simulated employment interviews. *Journal of Applied Psychology, 62*(3), 295.

Ingold, P. V., Kleinmann, M., Konig, C. J., Melchers, K. G., & Van Iddekinge, C. H. (2015). Why do situational interviews predict job performance? The role of interviewees' ability to identify criteria. *Journal of Business and Psychology, 30*(2), 387-398. https://doi.org/10.1007/s10869-014-9368-3

James, L. R., Demaree, R. G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology, 69*(1), 85-98. https://doi.org/10.1037/0021-9010.69.1.85

Jansen, A., Melchers, K. G., Lievens, F., Kleinmann, M., Brändli, M., Fraefel, L., & König, C. J. (2013). Situation assessment as an ignored factor in the behavioral consistency paradigm underlying the validity of personnel selection procedures. *Journal of Applied Psychology, 98*(2), 326-341. https://doi.org/10.1037/a0031257

John, O. P., & Robins, R. W. (1993). Determinants of interjudge agreement on personality traits: The Big Five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of personality, 61*(4), 521-551.

Jones, E. E., & Sigall, H. (1971). The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin, 76*(5), 349-364. https://doi.org/10.1037/h0031617

Jorm, A. F. (1987). Sex differences in neuroticism: A quantitative synthesis of published research. *Australian and New Zealand Journal of Psychiatry, 21*(4), 501-506. https://doi.org/10.3109/00048678709158917

Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist, 28*(2), 107-128. https://doi.org/10.1037/h0034225

Kenny, D. A., Horner, C., Kashy, D. A., & Chu, L.-c. (1992). Consensus at zero acquaintance: replication, behavioral cues, and stability. *Journal of Personality and Social Psychology, 62*(1), 88.

King, W. R., & He, J. (2006). A meta-analysis of the technology acceptance model. *Information & Management, 43*(6), 740-755. https://doi.org/10.1016/j.im.2006.05.003

Kleinmann, M. (1993). Are rating dimensions in assessment centers transparent for participants? Consequences for criterion and construct validity. *Journal of Applied Psychology, 78*(6), 988-993.

Kohn, L. S., & Dipboye, R. L. (1998). The effects of interview structure on recruiting outcomes. *Journal of Applied Social Psychology, 28*(9), 821-843. https://doi.org/10.1111/j.1559-1816.1998.tb01733.x

König, C. J., Melchers, K. G., Kleinmann, M., Richter, G. M., & Klehe, U.-C. (2007). Candidates' ability to identify criteria in nontransparent selection procedures: Evidence from an assessment center and a structured interview. *International Journal of Selection and Assessment, 15*(3), 283-292. https://doi.org/10.1111/j.1468-2389.2007.00388.x

Kroeck, K. G., & Magnusen, K. O. (1997). Employer and job candidate reactions to videoconference job interviewing. *International Journal of Selection and Assessment, 5*(2), 137-142. https://doi.org/10.1111/1468-2389.00053

Kroll, E., & Ziegler, M. (2016). Discrimination due to ethnicity and gender: How susceptible are video-based job interviews? *International Journal of Selection and Assessment, 24*(2), 161-171. https://doi.org/10.1111/ijsa.12138

Kuncel, N. R., Klieger, D. M., Connelly, B. S., & Ones, D. S. (2013). Mechanical versus clinical data combination in selection and admissions decisions: A meta-analysis. *Journal of Applied Psychology, 98*(6), 1060-1072. https://doi.org/10.1037/a0034156

Kutcher, E. J., & Bragger, J. D. (2004). Selection interviews of overweight job applicants: Can structure reduce the bias? *Journal of Applied Social Psychology, 34*(10), 1993-2022. https://doi.org/10.1111/j.1559-1816.2004.tb02688.x

Landers, R. N., Sackett, P. R., & Tuzinski, K. A. (2011). Retesting after initial failure, coaching rumors, and warnings against faking in online personality measures for selection. *Journal of Applied Psychology, 96*(1), 202-210. https://doi.org/10.1037/a0020375

Lang, F. R., Ludtke, O., & Asendorpf, J. B. (2001). Validity and psychometric equivalence of the German version of the big five inventory in young, middle-aged and old adults. *Diagnostica, 47*(3), 111-121. https://doi.org/10.1026//0012-1924.47.3.111

Langer, M., Konig, C. J., & Fitili, A. (2018). Information as a double-edged sword: The role of computer experience and information on applicant reactions towards novel technologies for personnel selection. *Computers in Human Behavior, 81*, 19-30. https://doi.org/10.1016/j.chb.2017.11.036

Langer, M., Konig, C. J., Gebhard, P., & Andre, E. (2016). Dear computer, teach me manners: Testing virtual employment interview training. *International Journal of Selection and Assessment, 24*(4), 312-323. https://doi.org/10.1111/ijsa.12150

Langer, M., Konig, C. J., & Krause, K. (2017). Examining digital interviews for personnel selection: Applicant reactions and interviewer ratings. *International Journal of Selection and Assessment, 25*(4), 371-382. https://doi.org/10.1111/ijsa.12191

Latane, B. (1981). The psychology of social impact. *American Psychologist, 36*(4), 343-356. https://doi.org/10.1037/0003-066x.36.4.343

Latham, G. P., & Sue-Chan, C. (1999). A meta-analysis of the situational interview: An enumerative review of reasons for its validity. *Canadian Psychology/Psychologie canadienne, 40*(1), 56-67. https://doi.org/10.1037/h0086826

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature-review and 2-component model. *Psychological Bulletin, 107*(1), 34-47. https://doi.org/10.1037/0033-2909.107.1.34

LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods, 11*(4), 815-852. https://doi.org/10.1177/1094428106296642

Levashina, J., & Campion, M. A. (2007). Measuring faking in employment interviews: Development and validation of an Interview Faking Behavior scale. *Journal of Applied Psychology, 92*(6), 1638-1656. https://doi.org/10.1037/0021-9010.92.6.1638

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

Lievens, F., De Corte, W., & Westerveld, L. (2015). Understanding the building blocks of selection procedures: Effects of response fidelity on performance and validity. *Journal of Management, 41*(6), 1604-1627. https://doi.org/10.1177/0149206312463941

Lievens, F., & Patterson, F. (2011). The validity and incremental validity of knowledge tests, low-fidelity simulations, and high-fidelity simulations for predicting job performance in

advanced-level high-stakes selection. *Journal of Applied Psychology, 96*(5), 927-940. https://doi.org/10.1037/a0023496

Lievens, F., & Sackett, P. R. (2006). Video-based versus written situational judgment tests: A comparison in terms of predictive validity. *Journal of Applied Psychology, 91*(5), 1181-1188. https://doi.org/10.1037/0021-9010.91.5.1181

Lievens, F., & Sackett, P. R. (2007). Situational judgment tests in high-stakes settings: Issues and strategies with generating alternate forms. *Journal of Applied Psychology, 92*(4), 1043-1055. https://doi.org/10.1037/0021-9010.92.4.1043

Lievens, F., & Sackett, P. R. (2017). The effects of predictor method factors on selection outcomes: A modular approach to personnel selection procedures. *Journal of Applied Psychology, 102*(1), 43-66. https://doi.org/10.1037/apl0000160

Maas, J. B. (1965). Patterned scaled expectation interview: Reliability studies on a new technique. *Journal of Applied Psychology, 49*(6), 431-433. https://doi.org/10.1037/h0022803

Macan, T. (2009). The employment interview: A review of current studies and directions for future research. *Human Resource Management Review, 19*(3), 203-218. https://doi.org/10.1016/j.hrmr.2009.03.006

Maier, C., Laumer, S., Eckhardt, A., & Weitzel, T. (2013). Analyzing the impact of HRIS implementations on HR personnel's job satisfaction and turnover intention. *Journal of Strategic Information Systems, 22*(3), 193-207. https://doi.org/10.1016/j.jsis.2012.09.001

Marchese, M. C., & Muchinsky, P. M. (1993). The validity of the employment interview: A meta‐analysis. *International Journal of Selection and Assessment, 1*(1), 18-26. https://doi.org/10.1111/j.1468-2389.1993.tb00080.x

Marr, J. C., & Cable, D. M. (2014). Do interviewers sell themselves short? The effects of selling orientation on interviewers' judgments. *Academy of Management Journal, 57*(3), 624-651. https://doi.org/10.5465/amj.2011.0504

Maurer, S. D. (2002). A practitioner-based analysis of interviewer job expertise and scale format as contextual factors in situational interviews. *Personnel Psychology, 55*(2), 307-327.

Maurer, T. J., Solamon, J., & Troxtel, D. (1998). Relationship of coaching with performance in situational employment interviews. *Journal of Applied Psychology, 83*(1), 128-136. https://doi.org/10.1037/0021-9010.83.1.128

Mayfield, E. C. (1964). The selection interview: A re-evaluation of published research. *Personnel Psychology, 17*(3), 239-260. https://doi.org/10.1111/j.1744-6570.1964.tb00065.x

McCarthy, J. M., Bauer, T. N., Truxillo, D. M., Anderson, N., Costa, A. C., & Ahmed, S. M. (2017). Applicant perspectives during selection: A review addressing 'So what?,' 'What's new?,' and 'Where to next?'. *Journal of Management, 43*(6), 1693-1725. https://doi.org/10.1177/0149206316681846

McCarthy, J. M., & Goffin, R. (2004). Measuring job interview anxiety: Beyond weak knees and sweaty palms. *Personnel Psychology, 57*(3), 607-637. https://doi.org/10.1111/j.1744-6570.2004.00002.x

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.

McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. D. (1994). The validity of employment interviews: A comprehensive review and metaanalysis. *Journal of Applied Psychology, 79*(4), 599-616. https://doi.org/10.1037/0021-9010.79.4.599

Melchers, K. G., Lienhardt, N., Von Aarburg, M., & Kleinmann, M. (2011). Is more structure really better? A comparison of frame-of-reference training and descriptively anchored rating scales to improve interviewers' rating quality. *Personnel Psychology, 64*(1), 53-87. https://doi.org/10.1111/j.1744-6570.2010.01202.x

Messick, S. (1995). Validity of psychological-assessment:Validation of inferences from persons responses and performances as Scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749. https://doi.org/10.1037/0003-066x.50.9.741

Middendorf, C. H., & Macan, T. H. (2002). Note-taking in the employment interview: Effects on recall and judgments. *Journal of Applied Psychology, 87*(2), 293-303. https://doi.org/10.1037//0021-9010.87.2.293

Miller, J., & Haden, P. (2006). *Statistical analysis with the general linear model*. USA: Creative Commons Attribution.

Mitchell, G. (2012). Revisiting truth or triviality: The external validity of research in the psychological laboratory. *Perspectives on Psychological Science, 7*(2), 109-117. https://doi.org/10.1177/1745691611432343

Monahan, E. L., Hoffman, B. J., Lance, C. E., Jackson, D. J. R., & Foster, M. R. (2013). Now you see them, now you do not: The influence of indicator-factor ratio on support for assessment center dimensions. *Personnel Psychology, 66*(4), 1009-1047. https://doi.org/10.1111/peps.12049

Morelli, N. A., Potosky, D., Arthur, W., Jr., & Tippins, N. (2017). A call for conceptual models of technology in I/O psychology: An example from technology-based talent assessment. *Industrial and Organizational Psychology-Perspectives on Science and Practice, 10*(4), 634-653. https://doi.org/10.1017/iop.2017.70

Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology, 60*(3), 683-729. https://doi.org/10.1111/j.1744-6570.2007.00089.x

Moscoso, S. (2000). Selection interview: A review of validity evidence, adverse impact and applicant reactions. *International Journal of Selection and Assessment, 8*(4), 237-247. https://doi.org/0.1111/1468-2389.00153

Motowidlo, S. J., & Burnett, J. R. (1995). Aural and visual sources of validity in structured employment interviews. *Organizational Behavior and Human Decision Processes, 61*(3), 239-249.

Motowidlo, S. J., Carter, G. W., Dunnette, M. D., Tippins, N., Werner, S., Burnett, J. R., & Vaughan, M. J. (1992). Studies of the structured behavioral interview. *Journal of Applied Psychology, 77*(5), 571.

Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure - the low-fidelity simulation. *Journal of Applied Psychology, 75*(6), 640-647. https://doi.org/10.1037//0021-9010.75.6.640

Murphy, K. R. (1986). When your top choice turns you down: Effect of rejected offers on the utility of selection tests. *Psychological Bulletin, 99*(1), 133-138. https://doi.org/10.1037/0033-2909.99.1.133

Murphy, N. A. (2007). Appearing smart: The impression management of intelligence, person

perception accuracy, and behavior in social interaction. *Personality and Social*

Psychology Bulletin, 33(3), 325-339. https://doi.org/10.1177/0146167206294871

Nguyen, L. S., & Gatica-Perez, D. (2015). *I would hire you in a minute: Thin slices of nonverbal*

behavior in job interviews. Paper presented at the Proceedings of the ACM International

Conference on Multimodal Interaction.

O'Neill, T. A., Lewis, R. J., & Carswell, J. J. (2011). Employee personality, justice perceptions,

and the prediction of workplace deviance. *Personality and Individual Differences, 51*(5),

595-600. https://doi.org/10.1016/j.paid.2011.05.025

Oliphant, G. C., Hansen, K., & Oliphant, B. J. (2008). A review of a telephone-administered

behavior-based interview technique. *Business Communication Quarterly, 71*, 383–386.

Oostrom, J. K., Born, M. P., Serlie, A. W., & van der Molen, H. T. (2010). Webcam testing:

Validation of an innovative open-ended multimedia test. *European Journal of Work and*

Organizational Psychology, 19(5), 532-550. https://doi.org/Pii 917059591

10.1080/13594320903000005

Oostrom, J. K., Melchers, K. G., Ingold, P. V., & Kleinmann, M. (2016). Why do situational

interviews predict performance? Is it saying how you would behave or knowing how you

should behave? *Journal of Business and Psychology, 31*(2), 279-291.

https://doi.org/10.1007/s10869-015-9410-0

Orlikowski, W. J. (2007). Sociomaterial practices: Exploring technology at work. *Organization*

Studies, 28(9), 1435-1448. https://doi.org/10.1177/0170840607081138

Parton, S. R., Siltanen, S. A., Hosman, L. A., & Langenderfer, J. (2002). Employment interview outcomes and speech style effects. *Journal of Language and Social Psychology, 21*(2), 144-161.

Ployhart, R. E., & Holtz, B. C. (2008). The diversity-validity dilemma: Strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology, 61*(1), 153-172. https://doi.org/10.1111/j.1744-6570.2008.00109.x

Ployhart, R. E., Schmitt, N., & Tippins, N. T. (2017). Solving the Supreme Problem: 100 years of selection and recruitment at the Journal of Applied Psychology. *Journal of Applied Psychology, 102*(3), 291-304. https://doi.org/10.1037/apl0000081 10.1037/apl0000081.supp (Supplemental)

Posthuma, R. A., Morgeson, F. P., & Campion, M. A. (2002). Beyond employment interview validity: A comprehensive narrative review of recent research and trends over time. *Personnel Psychology, 55*(1), 1-81. https://doi.org/10.1111/j.1744-6570.2002.tb00103.x

Potosky, D. (2008). A conceptual framework for the role of the administration medium in the personnel assessment process. *Academy of Management Review, 33*(3), 629-648.

Ree, M. J., Carretta, T. R., Earles, J. A., & Albert, W. (1994). Sign changes when correcting for range restriction: A note on Pearsons and Lawleys Selection Formulas. *Journal of Applied Psychology, 79*(2), 298-301. https://doi.org/10.1037/0021-9010.79.2.298

Reilly, R. R., & Chao, G. R. (1982). Validity and fairness of some alternative employee selection procedures. *Personnel Psychology, 35*(1), 1-62. https://doi.org/10.1111/j.1744-6570.1982.tb02184.x

Rockmann, K. W., & Northcraft, G. B. (2008). To be or not to be trusted: The influence of media richness on defection and deception. *Organizational Behavior and Human Decision Processes, 107*(2), 106-122. https://doi.org/10.1016/j.obhdp.2008.02.002

Roese, N. J., & Jamieson, D. W. (1993). Twenty years of bogus pipeline research: A critical review and meta-analysis. *Psychological Bulletin, 114*(2), 363-375. https://doi.org/10.1037/0033-2909.114.2.363

Roth, P. L., Bobko, P., Van Iddekinge, C. H., & Thatcher, J. B. (2016). Social media in employee-selection-related decisions: A research agenda for uncharted territory. *Journal of Management, 42*(1), 269-298. https://doi.org/10.1177/0149206313503018

Roth, P. L., & Huffcutt, A. I. (2013). A meta-analysis of interviews and cognitive ability back to the future? *Journal of Personnel Psychology, 12*(4), 157-169. https://doi.org/10.1027/1866-5888/a000091

Roulin, N., Bangerter, A., & Levashina, J. (2015). Honest and deceptive impression management in employment interviews: Can it be detected and how does it impact evaluations? *Personnel Psychology, 68*(2), 395-444. https://doi.org/10.1111/peps.12079

Ryan, A. M., Daum, D., Bauman, T., Grisez, M., Mattimore, K., Nalodka, T., & Mccormick, S. (1995). Direct, indirect, and controlled observation and rating accuracy. *Journal of Applied Psychology, 80*(6), 664-670. https://doi.org/10.1037//0021-9010.80.6.664

Ryan, A. M., & Ployhart, R. E. (2000). Applicants' perceptions of selection procedures and decisions: A critical review and agenda for the future. *Journal of Management, 26*(3), 565-606.

Rynes, S. L., & Boudreau, J. W. (1986). College recruiting in large organizations: Practice, evaluation, and research implications. *Personnel Psychology, 39*(4), 729-757. https://doi.org/10.1111/j.1744-6570.1986.tb00592.x

Sackett, P. R., Borneman, M. J., & Connelly, B. S. (2008). High stakes testing in higher education and employment: Appraising the evidence for validity and fairness. *American Psychologist, 63*(4), 215-227. https://doi.org/10.1037/0003-066x.63.4.215

Sackett, P. R., Dahlke, J. A., Shewach, O. R., & Kuncel, N. R. (2017). Effects of predictor weighting methods on incremental validity. *Journal of Applied Psychology, 102*(10), 1421-1434. https://doi.org/10.1037/apl0000235

Sackett, P. R., & Roth, L. (1996). Multi-stage selection strategies: A Monte Carlo investigation of effects on performance and minority hiring. *Personnel Psychology, 49*(3), 549-572. https://doi.org/10.1111/j.1744-6570.1996.tb01584.x

Sackett, P. R., Schmitt, N., Ellingson, J. E., & Kabin, M. B. (2001). High-stakes testing in employment, credentialing, and higher education: Prospects in a post-affirmative-action world. *American Psychologist, 56*(4), 302-318. https://doi.org/10.1037/0003-066X.56.4.302

Sackett, P. R., Shewach, O. R., & Keiser, H. N. (2017). Assessment centers versus cognitive ability tests: Challenging the conventional wisdom on criterion-related validity. *Journal of Applied Psychology, 102*(10), 1435-1447. https://doi.org/10.1037/apl0000236

Sackett, P. R., & Yang, H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Psychology, 85*(1), 112-118. https://doi.org/10.1037/0021-9010.85.1.112

Saks, A. M. (2006). Multiple predictors and criteria of job search success. *Journal of Vocational Behavior, 68*(3), 400-415. https://doi.org/10.1016/j.jvb.2005.10.001

Salgado, J. F., & Moscoso, S. (2002). Comprehensive meta-analysis of the construct validity of the employment interview. *European Journal of Work and Organizational Psychology, 11*(3), 299-324. https://doi.org/10.1080/13594320244000184

Salvemini, N. J., Reilly, R. R., & Smither, J. W. (1993). The influence of rater motivation on assimilation effects and accuracy in performance ratings. *Organizational Behavior and Human Decision Processes, 55*(1), 41-60. https://doi.org/10.1006/obhd.1993.1023

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*(2), 262-274. https://doi.org/10.1037//0033-2909.124.2.262

Schmidt, F. L., & Rader, M. (1999). Exploring the boundary conditions for interview validity: Meta-analytic validity findings for a new interview type. *Personnel Psychology, 52*(2), 445-464. https://doi.org/10.1111/j.1744-6570.1999.tb00169.x

Schmidt, F. L., & Zimmerman, R. D. (2004). A counterintuitive hypothesis about employment interview validity and some supporting evidence. *Journal of Applied Psychology, 89*(3), 553-561.

Schmitt, N. (1976). Social and situational determinants of interview decisions: Implications for the employment interview. *Personnel Psychology, 29*(1), 79-101. https://doi.org/10.1111/j.1744-6570.1976.tb00404.x

Schwab, D. P., & Heneman, H. G. (1969). Relationship between interview structure and interinterviewer reliability in an employment situation. *Journal of Applied Psychology, 53*(3p1), 214-&. https://doi.org/10.1037/h0027409

Sears, G. J., Zhang, H., Wiesner, W. H., Hackett, R. D., & Yuan, Y. (2013). A comparative assessment of videoconference and face-to-face employment interviews. *Management Decision, 51*(8), 1733-1752. https://doi.org/10.1108/MD-09-2012-0642

Shaw, A. (2012). Using chatbots to easily create interactive and intelligent FAQ webpages. *Journal of Applied Global Research, 5*(15), 10-15.

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunication*. London: Wiley.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations - Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420-428. https://doi.org/10.1037/0033-2909.86.2.420

Silvester, J., & Anderson, N. (2003). Technology and discourse: A comparison of face-to-face and telephone employment interviews. *International Journal of Selection and Assessment, 11*(2-3), 206-214. https://doi.org/10.1111/1468-2389.00244

Silvester, J., Anderson, N., Haddleton, E., Cunningham-Snell, N., & Gibb, A. (2000). A cross-modal comparison of telephone and face-to-face selection interviews in graduate recruitment. *International Journal of Selection and Assessment, 8*(1), 16-21. https://doi.org/10.1111/1468-2389.00127

Smith, D. B., & Ellingson, J. E. (2002). Substance versus style: A new look at social desirability in motivating contexts. *Journal of Applied Psychology, 87*(2), 211-219. https://doi.org/10.1037//0021-9010.87.2.211

Smith, D. B., Hanges, P. J., & Dickson, M. W. (2001). Personnel selection and the five-factor model: Reexamining the effects of applicant's frame of reference. *Journal of Applied Psychology, 86*(2), 304-315. https://doi.org/10.1037/0021-9010.86.2.304

Software Advice. (2018). Video interview software. Retrieved from

http://www.softwareadvice.com/hr/video-interviewing-comparison/

Spannagel, C., & Bescherer, C. (2009). Computerbezogene Selbstwirksamkeitserwartung in

Lehrveranstaltungen mit Computernutzung. *Notes on Educational Informatics - Section

A: Concepts and Techniques,, 5*(1), 23-43.

Steiner, D. D., & Gilliland, S. W. (1996). Fairness reactions to personnel selection techniques in

France and the United States. *Journal of Applied Psychology, 81*(2), 134-141.

Stevens, C. K., & Kristof, A. L. (1995). Making the right impression: A field-study of applicant

impression management during job interviews. *Journal of Applied Psychology, 80*(5),

587-606. https://doi.org/10.1037/0021-9010.80.5.587

Straus, S. G., Miles, J. A., & Levesque, L. L. (2001). The effects of videoconference, telephone,

and face-to-face media on interviewer and applicant judgments in employment

interviews. *Journal of Management, 27*(3), 363-381. https://doi.org/10.1016/S0149-

2063(01)00096-4

Sulsky, L. M., & Balzer, W. K. (1988). Meaning and measurement of performance rating

accuracy: Some methodological and theoretical concerns. *Journal of Applied Psychology,

73*(3), 497-506. https://doi.org/10.1037/0021-9010.73.3.497

Swaab, R. I., Galinsky, A. D., Medvec, V., & Diermeier, D. A. (2012). The communication

orientation model: Explaining the diverse effects of sight, sound, and synchronicity on

Negotiation and group decision-making outcomes. *Personality and Social Psychology

Review, 16*(1), 25-53. https://doi.org/10.1177/1088868311417186

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning

and Instruction, 4*(4), 295-312. https://doi.org/10.1016/0959-4752(94)90003-5

Swider, B. W., Barrick, M. R., & Harris, T. B. (2016). Initial impressions: What they are, what they are not, and how they influence structured interview outcomes. *Journal of Applied Psychology, 101*(5), 625-638. https://doi.org/10.1037/apl0000077

Taylor, P. J., Pajo, K., Cheung, G. W., & Stringfield, P. (2004). Dimensionality and validity of a structured telephone reference check procedure. *Personnel Psychology, 57*(3), 745-772. https://doi.org/10.1111/j.1744-6570.2004.00006.x

Taylor, P. J., & Small, B. (2002). Asking applicants what they would do versus what they did do: A meta-analytic comparison of situational and past behaviour employment interview questions. *Journal of Occupational and Organizational Psychology, 75*, 277-294.

Thorndike, R. L. (1949). Personnel selection; test and measurement techniques. Oxford, England: Wiley.

Thorsteinson, T. J. (2018). A meta-analysis of interview length on reliability and validity. *Journal of Occupational and Organizational Psychology, 91*(1), 1-32. https://doi.org/10.1111/joop.12186

Tiller, D., O'Mara, D., Rothnie, I., Dunn, S., Lee, L., & Roberts, C. (2013). Internet-based multiple mini-interviews for candidate selection for graduate entry programmes. *Medical Education, 47*(8), 801-810. https://doi.org/10.1111/medu.12224

Tippins, N. T. (2015). Technology and assessment in selection. *Annual Review of Organizational Psychology and Organizational Behavior, 2*, 551-582. https://doi.org/10.1146/annurev-orgpsych-031413-091317

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. http://dx.doi.org/10.1016/j.ijhm.2016.10.012

Tourangeau, R., Steiger, D. M., & Wilson, D. (2002). Self-administered questions by telephone - Evaluating interactive voice response. *Public Opinion Quarterly, 66*(2), 265-278.

Tsai, W. C., Huang, T. C., Wu, C. Y., & Lo, I. (2010). Disentangling the effects of applicant defensive impression management tactics in job interviews. *International Journal of Selection and Assessment, 18*(2), 131-140.

Tullar, W. L. (1989). Relational control in employment interviews. *Journal of Applied Psychology, 74*(6), 971-977. https://doi.org/10.1037/0021-9010.74.6.971

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124-1131. https://doi.org/10.1126/science.185.4157.1124

Ulrich, L., & Trumbo, D. (1965). The selection interview since 1949. *Psychological Bulletin, 63*(2), 100-116. https://doi.org/10.1037/h0021696

Vallejo, G., Fernández, P., Cuesta, M., & Livacic-Rojas, P. E. (2015). Effects of modeling the heterogeneity on inferences drawn from multilevel designs. *Multivariate behavioral research, 50*(1), 75-90.

Van Iddekinge, C. H., Raymark, P. H., & Roth, P. L. (2005). Assessing personality with a structured employment interview: Construct-related validity and susceptibility to response inflation. *Journal of Applied Psychology, 90*(3), 536-552. https://doi.org/10.1037/0021-9010.90.3.536

Van Iddekinge, C. H., Raymark, P. H., Roth, P. L., & Payne, H. S. (2006). Comparing the psychometric characteristics of ratings of face-to-face and videotaped structured interviews. *International Journal of Selection and Assessment, 14*(4), 347-359.

Vanhove, A. J., Gibbons, A. M., & Kedharnath, U. (2016). Rater agreement, accuracy, and experienced cognitive load: Comparison of distributional and traditional assessment

approaches to rating performance. *Human Performance, 29*(5), 378-393.
https://doi.org/10.1080/08959285.2016.1192632

Venkatesh, V., & Bala, H. (2008). Technology acceptance model 3 and a research agenda on
interventions. *Decision Sciences, 39*(2), 273-315. https://doi.org/10.1111/j.1540-
5915.2008.00192.x

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the Technology Acceptance
Model: Four longitudinal field studies. *Management Science, 46*(2), 186-204.
https://doi.org/10.1287/Mnsc.46.2.186.11926

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of
information technology: Toward a unified view. *Mis Quarterly, 27*(3), 425-478.

viasto GmbH. (2019). Die interview suite für unsere Kunden. Retrieved from
https://www.viasto.com/kunden/

Wagner, S. M., Rau, C., & Lindemann, E. (2010). Multiple informant methodology: A critical
review and recommendations. *Sociological Methods & Research, 38*(4), 582-618.
https://doi.org/10.1177/0049124110366231

Walther, J. B. (1996). Computer-mediated communication: Impersonal, interpersonal, and
hyperpersonal interaction. *Communication Research, 23*(1), 3-43. https://doi.org/
10.1177/009365096023001001

Waung, M., Hymes, R., Beatty, J. E., & McAuslan, P. (2015). Self-promotion statements in
video resumes: Frequency, intensity, and gender effects on job applicant evaluation.
International Journal of Selection and Assessment, 23(4), 345-360.
https://doi.org/10.1111/ijsa.12119

Whetzel, D. L., Baranowski, L. E., Petro, J. M., Curtin, P. J., & Fisher, J. L. (2003). A written

structured interview by any other name is still a selection instrument. *Applied H.R.M.*

Research, 8(1-2), 1-16.

Wiesner, W. H., & Cronshaw, S. F. (1988). A meta-analytic investigation of the impact of

interview format and degree of structure on the validity of the employment interview.

Journal of Occupational Psychology, 61, 275–290.

Williams, K. Z., Schaffer, M. M., & Ellis, L. E. (2013). Legal risk in selection: An analysis of

processes and tools. *Journal of Business and Psychology, 28*(4), 401-410.

https://doi.org/10.1007/s10869-013-9299-4

Wright, O. R., Jr. (1969). Summary of research on the selection interview since 1964. *Personnel*

Psychology, 22(4), 391-413. https://doi.org/10.1111/j.1744-6570.1969.tb00340.x

Wright, P. M., Lichtenfels, P. A., & Pursell, E. D. (1989). The structured interview: Additional

studies and a meta-analysis. *Journal of Occupational Psychology, 62*(3), 191-199.

https://doi.org/10.1111/j.2044-8325.1989.tb00491.x

Zuckerman, M., Hodgins, H., & Miyake, K. (1990). The vocal attractiveness stereotype:

Replication and elaboration. *Journal of Nonverbal Behavior, 14*(2), 97-112.

List of Publications

<u>Journal Articles (with Peer Review)</u>

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

<u>Journal Articles (without Peer Review)</u>

Brenner, F. (2014). Keine Angst vor Big Data: Der wahre Nutzen hinter dem Hype. HR

    Performance 7/2014, 12-15.

Brenner, F. (2014). Personalauswahl: Die passenden Tools nutzen. Personal im Fokus 6/14, 14-

16.

Brenner, F. (20.11.2013). Methodenmix in der Personalauswahl. Retrieved from

    http://www.haufe.de/personal/hr-management/methodenmix-in-der-

    personalauswahl/methodenmix-in-der-personalauswahl_80_205810.html (verschiedene

    Nachdrucke)

<u>Book Chapters</u>

Langer, M., Brenner, F. S., Green, N., & König, C. J. (2018). Digitale Kompetenzbewertung in

    der Personalauswahl- und Entwicklung. In W. Appel & M. Wahler (eds.), *Die digitale*

    HR-Organisation: Wo wir stehen, was wir brauchen (S. 211-220). Neuwied, Germany:

    Personalwirtschaft/Wolters Kluwer.

Brenner, F. (2016). Bridging the Scientist-Practitioner Gap: Einflussfaktoren auf die

Bewerberakzeptanz bei neuen Technologien am Beispiel zeitversetzter Videointerviews.

In T. Verhoeven (Hrsg.), *Candidate Experience* (S. 71 -89). Wiesbaden: Springer

Fachmedien. DOI 10.1007/978-3-658-08896-5.

Brenner, F. (2015). Online-Assessments und digitale Medien in der Personalauswahl. In L.

Gooßens, M. Kersting, & S. Koch (Hrsg.), *Auf die richtigen Mitarbeiter kommt es an:*

Personaldiagnostik und ihre Anwendung (S. 201 – 212). Stuttgart: Deutscher Sparkassen

Verlag.

Conference Contributions

Brenner, F. (2018). Untersuchungen über die psychometrischen Gütekriterien von

zeitversetzten Videointerviews. Präsentation auf der Konferenz der Deutschen Gesellschaft für

Psychologie (DGPS) 2018, Frankfurt am Main, Deutschland.

Diese Seite ist aus datenschutzrechtlichen Gründen Online nicht verfügbar

Zusammenfassung (Deutsch)

Die Nutzung neuer digitaler Technologien zur Durchführung von Einstellungsinterviews ist sehr populär geworden. Dies umfasst auch synchrone und asynchrone Formen von Videointerviews. Entgegen ihrer weiten Verbreitung in der Praxis ist bisher wenig darüber bekannt, wie diese Technologien psychometrische Gütekriterien und Auswahlergebnisse beeinflussen. Erkenntnisse, die in Bezug auf das persönliche Einstellungsgespräche gewonnen wurden, können nicht einfach auf digitale Formen des Interviews übertragen werden. Darüber hinaus wurde unter Wissenschaftlern jüngst der Ruf nach vermehrten Anstrengungen in der Theorieentwicklung laut, um verschiedene Limitationen denen Cross-Medienvergleiche unterliegen, zu überwinden und zu erklären *warum* und *wie* Technologie Auswahlergebnisse beeinflusst. Obwohl verschiedene Theorien über Mediennutzung, Medienauswahl und Medienadaption existieren, ist deren Anwendbarkeit und Erklärungswert in Bezug auf Technologie in Einstellungsinterviews kaum adressiert. Sowohl ein systematischer Übersichtsartikel über konzeptuelle Modelle für den Technologieeinsatz in Einstellungsinterviews als auch empirische Ergebnisse im wenig beforschten Feld digitaler Interviewformen adressieren eine erhebliche Forschungslücke, weisen Wege für zukünftige Forschung und würden HR-Praktikern Hinweise für die Ausgestaltung von Auswahlsystemen geben.

In der vorliegenden Dissertation werden daher eine umfassende Literaturübersicht bezüglich Technologieeinsatzes in Einstellungsinterviews vorgestellt und fünf empirische Studien über die spezifische Form zeitversetzter Videointerviews präsentiert. Die Literaturübersicht integriert verschiedene Perspektiven auf technologieunterstützte Interviews, theoretische Richtungen der Medienforschung, Forschung bezüglich Unterschiede in

psychometrischen Gütekriterien und Auswahlergebnissen durch Technologie und Theorieintegration in ein kompaktes Arbeitsmodell.

Der empirische Teil der Dissertation präsentiert fünf explorative Untersuchungen über zeitversetzte Videointerviews, in denen ausgewählte Forschungsfragen untersucht werden, die blinde Flecke in der Literatur adressieren oder drängende Fragen für den Einsatz in der Praxis.

Die erste Studie adressiert die Beobachterübereinstimmung und den Einfluss von strukturierten Evaluationsformaten in einer Stichprobe von hypothetischen Bewerbern mit $N = 111$ Teilnehmern. Die zweite Studie untersucht den Einfluss der sozialen Bandbreite auf die Genauigkeit von Interviewbewertungen in einer Stichprobe von hypothetischen Bewerbern mit $N = 279$ Teilnehmern. Die dritte Studie legt die ersten Ergebnisse bezüglich der Validität von zeitversetzten Videointerviews im Feld vor mit einer Stichprobe von $N = 899$ Bewerbern. Die vierte Studie untersucht den Einfluss von personalisierter Kommunikation über Videobotschaften mit Bewerbern auf die Bewerberwahrnehmung in zeitversetzten Videointerviews in einer Stichprobe von hypothetischen Bewerbern mit $N = 98$ Teilnehmern. Zuletzt untersucht die fünfte Studie eine Inflation von Interviewbewertungen aufgrund von Vorbereitungszeiten in einer Stichprobe von hypothetischen Bewerbern mit $N = 51$ Teilnehmern.

Diese Dissertation trägt zur Fachliteratur in verschiedenen Punkten bei: Die Literaturübersicht bietet einen aktualisierten und mehrperspektivischen Überblick auf das Feld und integriert verschiedene Forschungsrichtungen in ein Modell für den Technologieeinsatz in Einstellungsinterviews. Die empirischen Studien zeigen erste und vielversprechende Ergebnisse bezüglich der psychometrischen Gütekriterien zeitversetzter Videointerviews für Reliabilität und Validität, aber betonen auch mögliche Risiken wie Bewertungsinflationen, die Auftreten können, wenn Vorbereitungszeiten im Prozessdesign integriert werden. Zuletzt betonen die Ergebnisse

der Studien ein weiteres Mal die herausgehobene Bedeutung von Struktur in

Einstellungsinterviews, auch wenn ein tieferes Verständnis für die Beziehung zwischen

Interviewstruktur und Technologieeinsatz in Einstellungsinterviews noch gewonnen werden

muss.

Eidesstattliche Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit allein verfasst und keine anderen als die angegebenen Hilfsmittel verwendet habe. Diese Arbeit ist in keinem früheren Promotionsverfahren angenommen oder abgelehnt worden.

Berlin, 17.03.2020

Falko Steffen Brenner

Diese Seite ist aus datenschutzrechtlichen Gründen Online nicht verfügbar