

THE STRUCTURED EMPLOYMENT INTERVIEW: NARRATIVE AND QUANTITATIVE REVIEW OF THE RESEARCH LITERATURE

JULIA LEVASHINA
Kent State University

CHRISTOPHER J. HARTWELL
Purdue University

FREDERICK P. MORGESON
Michigan State University

MICHAEL A. CAMPION
Purdue University

In the 20 years since frameworks of employment interview structure have been developed, a considerable body of empirical research has accumulated. We summarize and critically examine this literature by focusing on the 8 main topics that have been the focus of attention: (a) the definition of structure; (b) reducing bias through structure; (c) impression management in structured interviews; (d) measuring personality via structured interviews; (e) comparing situational versus past-behavior questions; (f) developing rating scales; (g) probing, follow-up, prompting, and elaboration on questions; and (h) reactions to structure. For each topic, we review and critique research and identify promising directions for future research. When possible, we augment the traditional narrative review with meta-analytic review and content analysis. We concluded that much is known about structured interviews, but there are still many unanswered questions. We provide 12 propositions and 19 research questions to stimulate further research on this important topic.

The employment interview has been one of the most widely used selection methods in the past 100 years (Binet, 1911; Macan, 2009). Its ubiquity has prompted some to suggest that “it is rare, even unthinkable, for someone to be hired without some type of interview” (Huffcutt & Culbertson, 2010, p. 185). It is often the only method used to assess applicants for employment or can serve as the initial screen to the final step in a multiple-hurdle selection process.

The employment interview has also received wide interest from researchers. The first comprehensive review of research on the employment

Correspondence and requests for reprints should be addressed to Julia Levashina, College of Business Administration, Kent State University, Kent, OH 44242-0001; jlevashi@kent.edu.

interview was published by Wagner in 1949 and there have been regular reviews ever since in the top journals in the field (Arvey & Campion, 1982; Harris, 1989; Mayfield, 1964; Posthuma, Morgeson, & Campion, 2002; Schmitt, 1976; Ulrich & Trumbo, 1965; Wright, 1969). One of the most consistent findings in the history of research on the employment interview is that structured interviews are much more reliable and valid than unstructured interviews.

Twelve meta-analyses have been conducted on this topic, and they have consistently found strong evidence for the superiority of structured interviews compared to unstructured interviews (Conway, Jako, & Goodman, 1995; Huffcutt & Arthur, 1994; Huffcutt, Conway, Roth, & Klehe, 2004; Hunter & Hunter, 1984; Latham & Sue-Chan, 1999; Marchese & Muchinsky, 1993; McDaniel, Whetzel, Schmidt, & Maurer, 1994; Reilly & Chao, 1982; Schmidt & Rader, 1999; Schmidt & Zimmerman, 2004; Wiesner & Cronshaw, 1988; Wright, Lichtenfels, & Pursell, 1989). In addition, structured interviews often provide incremental validity over personality tests and cognitive ability tests because they are typically only weakly related to each other (e.g., Berry, Sackett, & Landers, 2007; Cortina, Goldstein, Payne, Davison, & Gilliland, 2000; Huffcutt, Roth, & McDaniel, 1996; Salgado & Moscoso, 2002). Finally, structured interviews can be designed to measure different constructs (Huffcutt, Conway, Roth, & Stone, 2001) and predict different criteria (e.g., ethical behaviors, Hollwitz & Pawlowski, 1997; maximum and typical performance, Klehe & Latham, 2006).

In the 1990s, several structured interview frameworks were developed (Campion, Palmer, & Campion, 1997; Huffcutt & Arthur, 1994; McDaniel et al. 1994). In the intervening 20 years, although a considerable amount of research has been conducted, there have not been any comprehensive reviews of the literature focused on the structured interview. Given this, we feel it is a good time to revisit this literature and assess the state of our collective knowledge about structured interviewing. Thus, our purpose is to provide a systematic review of the literature on the structured employment interview since 1996 (the last year summarized in the most recent previous review). The paper is organized around eight topics that have been the most frequently studied during this time, including (a) the definition of structure; (b) reducing group differences in interview ratings through structure; (c) impression management in structured interviews; (d) measuring personality via structured interviews; (e) comparing situational (SQ) versus past-behavior questions (PBQ); (f) developing rating scales; (g) probing, follow-up, prompting, and elaboration on questions; and (h) reactions to structure.

We define, review, and critique the published research for each topic. In areas where a meta-analysis has been conducted, we describe the

meta-analytic findings only; primary studies are discussed only if they were published after the meta-analysis. For parsimony, we focus on more methodologically rigorous studies. When discussing effect sizes, we follow Cohen's (1988) conventions, where bivariate correlations of .10, .30, and .50 represent small, medium, and large effect sizes, respectively. When possible, we augment our narrative review with quantitative analyses, such as meta-analytic and content analysis methodologies. The narrative review enhances conceptual understanding of the area and describes recommendations for future research. Content analysis complements the narrative review by providing quantitative analysis of the structured interview and its components. Meta-analysis provides a more powerful estimation of the true effect size of the different components of structure and a quantitative summary of the accumulated findings where a sufficient number of studies have been conducted. Finally, directions for future research are identified. To enhance our theoretical contribution, we develop formal propositions when theory or enough prior research was conducted to enable specific predictions. If there was not enough prior research, we propose research questions.

Several different means were utilized to identify the relevant articles for this review. First, a web-based search for articles from 1996 to the present was conducted in the PsychInfo, ABI/Inform, and Business Source Premier databases. Numerous search terms were used, including job interview, selection interview, employment interview, interview structure, and structured interview. Second, articles that have cited Campion et al. (1997) were examined for inclusion. Finally, as relevant articles were gathered from these searches, their reference sections were manually checked for any additional articles.

Definition of the Employment Interview

The employment interview has been defined as “a face-to-face interaction conducted to determine the qualifications of a given individual for a particular open position” (Huffcutt & Youngcourt, 2007, p. 182). The advancement of technology has led to changes in the media for conducting interviews. Today, the employment interview is no longer limited to face-to-face interaction, having been expanded to other media, including telephone (Oliphant, Hansen, & Oliphant, 2008) and computer-mediated video chat (Chapman & Rowe, 2002). Because interpersonal interaction is a central component of the employment interview, we define the 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*. According to this definition, 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 (Moore & Kearsley, 1996). In synchronous interaction, both the interviewee and the interviewer must participate in the interview at a fixed time, whereas in asynchronous interaction, they participate in the interview according to their own schedule. For example, asynchronous evaluation happens when the interview is recorded, and the decision maker watches the video to evaluate the applicant in order to make a final decision. A written interview would be excluded from our definition because it may be more appropriately defined as a written test.

Definition of Structure

Many definitions of structure have been proposed in the literature. For example, Campion et al. (1997) defined structure 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. 656). Huffcutt and Arthur (1994) defined structure as “the degree of discretion that an interviewer is allowed in conducting the interview” (p. 186) and proposed two dimensions: (a) standardization of interview questions and (b) standardization of response scoring.

Components of Interview Structure

One of the main themes in these definitions is that structure involves the establishment and deliberate application of predetermined rules for questions, observations, and evaluations. The most comprehensive typology of interview structure was proposed by Campion et al. (1997). It consists of *content* and *evaluation* dimensions, under which 15 distinct components of structure are grouped. The *content* dimension includes the components of (a) basing questions on a job analysis; (b) asking the same questions of each applicant; (c) limiting prompting, follow-up, and elaboration on questions; (d) using better types of questions; (e) using longer interviews or larger number of questions; (f) controlling ancillary information; and (g) not allowing questions from applicant until after the interview. The *evaluation* dimension includes the components of (a) rating each answer or using multiple scales; (b) using anchored rating scales; (c) taking notes; (d) using multiple interviewers; (e) using the same interviewer(s) across all applicants; (f) not discussing applicants/answers between interviews;

(g) providing interviewer training; and (h) using statistical, rather than clinical, prediction.

Validating the interview structure typology. Campion et al. (1997) developed their typology based on conceptual logic but did not offer any empirical evidence for its validity. Thus, we examined whether the 15 components are appropriately classified into their respective content and evaluation dimensions. To do this, we used the expert judgment categorization methodology proposed by Hinkin and Tracey (1999).¹ We identified 47 subject matter experts (SMEs) in personnel selection and the employment interview. The SMEs were selected based on their research activity in the domain and their status in the field, as reflected in their presence on the editorial boards of leading journals on personnel selection (e.g., *Personnel Psychology* and *Journal of Applied Psychology*). The SMEs were contacted via e-mail and invited to participate in a brief web-based survey. Completed surveys were received by 30 SMEs (64% response rate), who had an average of 18 years of academic experience, nine publications in the selection domain, and three publications on interviewing-related topics. The 15 components of structure were defined (drawing from the definitions provided in Campion et al., 1997) and presented in random order to control for potential order effects. The SMEs rated each of the 15 components on the extent to which they believed the components were consistent with each of the two structure dimensions. A five-point rating scale (1 = *not at all* to 5 = *completely*) was used.

To determine the validity of the typology, we conducted a one-way repeated ANOVA to compare a component's mean ratings on the *content* dimension to the mean ratings on the *evaluation* dimension (following Hinkin & Tracey, 1999). Support for the typology exists if the mean ratings are significantly higher for the expected structure dimension. As shown in Table 1, 14 out of the 15 interview components have mean scores that are significantly higher ($p > .05$) on the dimension proposed by Campion et al. Only one component, control ancillary information, has a mean that is marginally higher ($p = .092$) on a different dimension, suggesting that this component might be a part of the evaluation, and not content, dimension. Overall, this supports the interview structure typology developed by Campion et al.

Content analysis of the usage of structure components by past research. Given its comprehensiveness and influence on subsequent research, we content analyzed recent research on structured interviews in order to examine the usage of these 15 structure components. In addition, we included three potential new components not covered by Campion et al.,

¹We thank an anonymous reviewer for this suggestion.

TABLE 1
Expert Mean Ratings of Structure Components Across Content and Evaluation Dimensions

Structure components	Structure dimensions		<i>p</i> -value*
	Content <i>Mean (SD)</i>	Evaluation <i>Mean (SD)</i>	
1. Job analysis	4.77 (.43)	2.67 (1.24)	<.0001
2. Same questions	4.27 (1.14)	3.20 (1.54)	.02
3. Limit prompting	4.37 (1.00)	3.03 (1.27)	.0003
4. Better questions	4.73 (.52)	3.47 (1.36)	<.0001
5. Longer interview	4.13 (1.04)	3.03 (1.33)	.001
6. Control ancillary information	3.50 (1.41)	4.10 (1.03)	.09
7. No questions from applicant	3.90 (1.40)	3.03 (1.33)	.02
8. Rate each question	2.30 (1.34)	4.87 (.35)	<.0001
9. Anchored rating scales	2.33 (1.27)	4.90 (.31)	<.0001
10. Detailed notes	2.47 (1.38)	4.47 (.68)	<.0001
11. Multiple interviewers	2.63 (1.30)	4.47 (.97)	<.0001
12. Same interviewer(s)	2.93 (1.39)	4.33 (.99)	<.0001
13. No discussion between interviews	2.20 (1.47)	4.56 (.82)	<.0001
14. Training	4.23 (1.04)	4.73 (.52)	.01
15. Statistical prediction	1.70 (1.15)	4.77 (.67)	<.0001

Note. $N = 30$. Bolded means indicate that the components were rated significantly higher on the dimension proposed by Campion et al. (1997).

*Significance of the difference between means.

but identified in the recent research (limiting rapport building, interview transparency, and recording of interviews). Articles were included if they described or tested overall structured interviews or structure components. Articles were discarded if there was no definition or operationalization of structure. Multiple studies within the same article were coded differently only if there were specific differences in the definition of structure across studies (e.g., Roth, Van Iddekinge, Huffcutt, Eidson, & Schmit, 2005; Townsend, Bacigalupi, & Blackman, 2007). The use of each of the 15 components, and several potential new components, were coded with a "1" for presence and a "0" for absence of the component. Both empirical and nonempirical articles and book chapters were included in the analysis. Moreover, interview setting (field vs. mock interviews) and interview modality (face-to-face, phone, and technology-mediated interviews) were recorded as potential moderators of the structure component use. These steps allowed us to analyze a total of 104 interviews from 103 articles.²

The results of the content analysis are presented in Table 2. Interviews typically include six components of structure ($M = 5.74$, $SD = 2.83$),

²Articles included in this analysis are indicated with an asterisk (*) in the reference section.

TABLE 2
*Inclusion of Structure Components in the Recent Research Based on the Content
 Analysis of Studies Conducted From 1997 to 2010*

Structure components	Total (<i>n</i> = 104) <i>M</i> (<i>SD</i>)	Nonempirical (<i>n</i> = 35) <i>M</i> (<i>SD</i>)	Field interviews (<i>n</i> = 31) <i>M</i> (<i>SD</i>)	Mock interviews (<i>n</i> = 38) <i>M</i> (<i>SD</i>)	t-test ^a
Content (total)	2.92 (1.41)	3.06 (1.75)	2.84 (1.34)	2.87 (1.12)	.10
1. Job analysis	.71 (.46)	.66 (.48)	.65 (.49)	.82 (.39)	1.61
2. Same questions	.74 (.44)	.71 (.46)	.65 (.49)	.84 (.37)	1.91
3. Limit prompting	.27 (.45)	.29 (.46)	.29 (.46)	.24 (.43)	-.50
4. Better questions	.78 (.42)	.74 (.44)	.84 (.37)	.76 (.43)	-.77
5. Longer interview	.19 (.40)	.17 (.38)	.29 (.46)	.13 (.34)	-1.64
6. Control ancillary information	.17 (.38)	.37 (.49)	.10 (.30)	.05 (.23)	-.70
7. No questions from applicant	.06 (.23)	.11 (.32)	.03 (.18)	.03 (.16)	-.14
Evaluation (total)	2.81 (1.75)	2.69 (1.73)	3.26 (1.95)	2.58 (1.60)	-1.59
8. Rate each question	.71 (.46)	.71 (.46)	.81 (.40)	.63 (.49)	-1.60
9. Anchored rating scales	.66 (.47)	.57 (.50)	.68 (.48)	.74 (.45)	.53
10. Detailed notes	.19 (.40)	.26 (.44)	.16 (.37)	.16 (.37)	-.04
11. Multiple interviewers	.45 (.50)	.46 (.51)	.61 (.50)	.32 (.47)	-2.55*
12. Same interviewer(s)	.13 (.34)	.09 (.28)	.16 (.37)	.16 (.37)	-.04
13. No discussion between interviews	.04 (.19)	.03 (.17)	.06 (.25)	.03 (.16)	-.73
14. Training	.52 (.50)	.46 (.51)	.58 (.50)	.53 (.51)	-.45
15. Statistical prediction	.11 (.31)	.11 (.32)	.19 (.40)	.03 (.16)	-2.18*
Total structure	5.74 (2.82)	5.74 (3.16)	6.10 (2.97)	5.45 (2.40)	-1.00
New components					
16. Rapport building	.05 (.21)	.06 (.24)	0	.08 (.27)	1.78
17. Recording of interviews	.30 (.46)	.29 (.46)	.32 (.48)	.32 (.47)	-.06
18. Interview transparency	.05 (.21)	.03 (.17)	.06 (.25)	.05 (.23)	-.23

Note. The most frequently used components are in bold.

^at-test of the significance of the inclusion of structure components in high-stakes versus mock interviews.

**p* < .05.

with approximately equal numbers of components for content ($M = 2.93$, $SD = 1.41$) and evaluation ($M = 2.82$, $SD = 1.76$). The components utilized most frequently are *job analysis*, *same questions*, *better types of questions*, *anchored rating scales*, *rating each question*, and *training*. T-tests were performed to compare the utilization of various structure components in mock and field interviews. Slightly more components are used in field ($M = 6.10$, $SD = 2.97$) than in mock interviews ($M = 5.45$, $SD = 2.40$), although the difference is not statistically significant. To compare utilization of components across types of interview modality, we performed an analysis of variance with total structure components as the dependent variable and interview modality as an independent variable ($F(2, 65) = 5.26$, $p < .01$). Face-to-face interviews ($M = 6.50$, $SD = 2.65$, $n = 44$) utilized more components of structure than phone ($M = 5.50$, $SD = 2.43$, $n = 6$) and video ($M = 4.28$, $SD = 1.90$, $n = 18$). Yet, only the difference between face-to-face and video interviews was statistically significant ($d = .96$, $p < .01$). The results of this content analysis provide valuable information regarding the current state of research on structured interviews. It is important to recognize, however, that our results are valid only to the extent to which a complete list of structure components was fully described in the articles that were used in our analysis.

Discussion and Future Research

Motowidlo et al. (1992) noted that the consensus among researchers was that structured interviews were better than unstructured, but no consensus existed on what the term *structure* meant. The same trend continues in the reviewed literature. Many articles and book chapters fail to define structure (i.e., Moscoso, 2000; Ployhart, 2006), and for articles that do, it is often unclear whether the provided description adequately reflects the operationalization. For example, Van Iddekinge, Raymark, Roth, and Payne (2006) state that 11 of the 15 components of Campion et al. (1997) are used, but only 7 are explicitly indicated. We therefore call for more thorough descriptions of structure in future research to establish a common definition of structure. The results of our content analysis show that past research has used an average of six components of structure, with approximately equal numbers of content and evaluation components. This leads to the following question:

Research Question 1: (a) Does using six structured interview components represent sufficient structuring of the interview? (b) What incremental value is achieved by incorporating additional components of structure?

The results of the content analysis also suggest that face-to-face interviews are structured with more components than alternative forms of interviews, such as phone or video interviews. Yet, these alternative forms of interviewing do not provide interviewers with the full range of verbal and nonverbal cues and may benefit from structuring even more than face-to-face interviews. There was a surprising lack of research on alternative forms of interviews, even though these types of interviews are commonly used as an initial interview by companies with large applicant pools, and they influence the characteristics of the applicant pool that move to the next level of selection process.

According to media richness theory (Daft & Lengel, 1986; Fletcher & Major, 2006; McGrath & Hollingshead, 1993), communication media vary in their amount of inherent richness. Face-to-face is the richest medium of communication because verbal cues (the word spoken), nonverbal cues (gestures, etc.), and paraverbal cues (vocal inflections, vocal tone, etc.) are all present. Video communication reduces richness because some of the nonverbal cues are absent due to constrained viewable space and a move from three to two dimensions (Fletcher & Major, 2006). Telephone communication reduces richness even further by eliminating nonverbal cues altogether, thus leaving only verbal and paraverbal cues (McGrath & Hollingshead, 1993). This reduction in communication richness limits the effectiveness of an individual to process complex and subjective information. Importantly, however, media low in richness can still be effective in processing information that is standardized and well understood (Daft & Lengel, 1986). In the context of selection interviews, alternative forms of interviews low in media richness, such as phone or video interviews, would therefore likely see a significantly larger benefit from structuring than would face-to-face interviews, due to a better fit between the communication medium and the complexity of the information that needs to be processed.

Proposition 1: Alternative forms of interviews, such as phone or video interviews, will benefit more (i.e., have better psychometric properties) from structuring than will face-to-face interviews designed to assess the same constructs.

Next, we review three additional components that have emerged in recent research, including *limiting rapport building*, *interview transparency*, and *recording of interviews*, that could be incorporated into the existing taxonomy of interview structure. Rapport building occurs prior to an interview and includes “small talk” or casual conversation that is initiated by the interviewer in order to put applicants at ease. The most likely effect of *limiting rapport building* is enhanced validity via decreasing contamination. Social psychological research suggests that people form

impressions very quickly (Ambady, Bernieri, & Richeson, 2000; Bar, Neta, & Linz, 2006). As such, interviewers may form early impressions based on the rapport building phase where nonjob-related information is exchanged and before any structured questions are asked. These evaluations might bias information gathering during the main phase of a structured interview (Dipboye, 1994). Therefore, we suggest that rapport building should be eliminated, minimized, or standardized. We propose four levels of this component. The highest level is the prohibition of rapport building (McCarthy, Van Iddekinge, & Campion, 2010). The next level is to allow only limited or preplanned rapport building. For example, standardized questions for rapport building may be provided, or interviewers may be allowed to ask a predetermined number of questions during specific time limits. The third level involves imposing limitations on the time of rapport building but no limitations on the questions or the theme of “small talk” (Barrick, Swider, & Stewart, 2010; Chapman & Zweig, 2005). The lowest level is providing no guidance on rapport building, where some interviewers may use it and others may not.

Little research has been published on rapport building in structured interviews. Chapman and Zweig (2005) indicated that rapport building is one of the four structure dimensions and is positively related to interviewer affective reactions. Thus eliminating rapport building might lead to negative interviewer reaction. Yet, Barrick et al. (2010) found that initial impressions from rapport building were related to mock-interview ratings and actual job offers.

Research Question 2: (a) Does limiting rapport building increase structured interview validity? (b) What level of rapport building leads to a better acceptance of structured interviews by the interviewers and applicants?

Another potentially new component is *interview transparency*. This is defined as the extent to which the interview dimensions assessed are made known to the applicants before or during the interview. When applicants are told what interview questions measure, they can better recall relevant past work experiences and use them in the answers. We propose four levels of transparency. The first level involves making sure that applicants do not get any information about the assessed dimensions (König, Melchers, Kleinmann, Richter, & Klehe, 2007). The second level is making applicants aware of the dimensions on which they would be rated in interviews (Allen, Facticeau, & Facticeau, 2004; Klehe, König, Richter, Kleinmann, & Melchers, 2008; Maurer, Solamon, & Lippstreu, 2008;

McFarland, Ryan, Sacco, & Kriska, 2004). The third level is giving applicants a list of questions before the interview (Day & Carroll, 2003; Personnel Decisions International, 1997). The fourth level is giving applicants a list of questions before the interview and explaining what dimensions are measured with each question. A challenge of this component and these levels of transparency is that it is unclear what level represents the higher level of structure. Because of this, exploring these different levels is an important area of future research.

Four studies (Allen et al., 2004; Day & Carroll, 2003; Klehe et al., 2008; Maurer et al., 2008) compared transparent and nontransparent interviews and found that transparency led to higher interview ratings. Transparency was also shown to increase (a) construct validity (Klehe et al., 2008); (b) reliability and predictive validity (Maurer et al., 2008); (c) applicants' perception of fairness; and (d) it did not decrease criterion-related validity of interviews (Day & Carroll, 2003). Yet, König et al. (2007) found that applicants' ability to identify criteria being measured was related to their ratings, suggesting that interviews should not be made transparent.

Research Question 3: (a) What transparency level represents the higher level of structure? (b) What is the impact of different levels of transparency on interview validity and faking?

Beyond intentional strategies organizations use to provide applicants with information prior to an interview, interview questions may become known to applicants through other means. For example, there are various websites that list common questions asked at specific companies (e.g., www.glassdoor.com, www.vault.com), and job applicants may post their interview questions online after their interviews. Interviews with high levels of structure may increase the likelihood of question leakage because there is a predetermined and relatively small set of questions that might become easily available to applicants without organizational "approval."

Research Question 4: (a) What is the extent of question leakage? (b) What is the impact of question leakage on the reliability and validity of structured interviews?

A final potentially new component is *recording of interviews*. The advantage of structured interviews may depend on the extent to which interviewers follow the prescribed structured format. Yet, we know that

interviewers do not always do so (Dipboye, 1997; Lievens & De Paepe, 2004). Recording of interviews may hold interviewers accountable for the way they conduct structured interviews to ensure that they follow the prescribed structured format (Sedikides, Herbst, Hardin, & Dardis, 2002), which has been shown to increase interview validity (Brtek & Motowidlo, 2002). We propose three levels of this component. The first and highest level includes video and audio recording of interviews (Brtek & Motowidlo, 2002). The second level includes audio recording. The third level includes interviews that are not recorded or there is no guidance on recording.

Research Question 5: (a) How much recording is needed? (b) What recording level leads to a better acceptance of structured interviews by the interviewers and applicants?

Reducing Group Differences in Interview Ratings Through Structure

Unstructured interviews have been criticized for their low reliability, low validity, and susceptibility to different biases, such as race, gender, and disability (Arvey & Campion, 1982). Biases may occur when interviewers gather and evaluate nonjob-related information about applicants. Because interviewers conduct unstructured interviews in an idiosyncratic way and have discretion in what they ask and how they evaluate responses (Dipboye, Wooten, & Halverson, 2004), the content and evaluation process in unstructured interviews may be more reflective of the interviewers' implicit theories of the job requirements than the actual job requirements. Structured interviews have stronger psychometric properties because structure links the decision process to job-related factors and limits the influence of extraneous information, such as disability, gender, or race. Seventeen studies were conducted on this topic since 1996. Race, gender, and disability were the most frequently examined group differences.

Main Recent Findings

Huffcutt and Roth (1998) conducted a meta-analysis of 31 studies and found that both Black and Hispanic applicants receive slightly lower interview ratings than White applicants, with White-Black mean differences of .25 and White-Hispanic differences of .26. High-structure interviews had lower group differences than low-structure interviews ($\bar{d} = .23$, and $\bar{d} = .32$, respectively), and past-behavior interviews had lower group differences than situational interviews ($\bar{d} = .10$, and $\bar{d} = .20$,

respectively). Five studies addressed Posthuma et al.'s (2002) call to investigate the effects of demographic similarity, or the match between the interviewee and the interviewer demographic characteristics on interview ratings. Most of these were field studies, utilizing large applicant samples and rigorously designed structured interviews. Small or no effects of race and gender similarity were found in one-on-one interviews (Chapman & Rowe, 2001; Sacco, Scheu, Ryan, & Schmitt, 2003), and in two-person (McCarthy et al., 2010), three-person (McFarland et al., 2004), and four-person panels structured interviews (Buckley, Jackson, Bolino, Veres, & Field, 2007).

Meta-analysis of Recent Findings on Group Differences in Structured Interviews

We meta-analyzed the findings across recent primary studies on group differences. We followed the Huffcutt and Roth (1998) procedures to facilitate a direct comparison of our meta-analytic results with theirs. Their article includes studies up until 1996, which is where our study picks up. In order to be as comprehensive as possible, we also examined the sources cited by Huffcutt and Roth and included 11 studies from eight sources in our analysis that met the following criteria: some evidence of structure in the interviews, some form of interview rating as a criterion variable, and adequate statistical information reported by race and/or gender conditions. There were no pre-1996 studies that examined group differences in structured interviews for applicants with disabilities. Following Huffcutt and colleagues (Huffcutt & Roth 1998; Huffcutt et al., 1996), studies with a sample size of less than 100 were weighted as "1" ($n = 4$), studies with sample sizes between 100 and 500 were weighted as "2" ($n = 3$), and studies with a sample size of over 500 were weighted as "3" ($n = 3$). This procedure helped to mitigate the dominance of large studies on the results of the meta-analysis. The results are reported in Table 3 separately for the pre- and post-1996 studies, as well as combined.³ There were no studies published in 1996. Pre-1996 studies showed a group difference in which the interviewees that were racial minorities and/or female were rated lower than their nonminority and/or male counterparts ($\bar{d} = .20$). This difference was more pronounced for real ($\bar{d} = .28$) than for mock interviews, which actually showed ratings that slightly favored minorities over nonminorities ($\bar{d} = -.05$). Race differences ($\bar{d} = .20$) were more pronounced than gender differences ($\bar{d} = .14$). Post-1996 studies showed minimal to nonexistent group differences. The overall effect size

³The pre-1996 articles included in this analysis are indicated with an "a" superscript (a) in the reference section, while the post-1996 studies are indicated using a "b" superscript (b).

TABLE 3
Meta-analysis of Group Differences in Structured Interview Ratings

	<i>k</i>	<i>N</i>	<i>d</i>	<i>SD_{res}</i>	95% CI
Pre-1996 studies					
Overall interview type	11	5,342	.20	.21	.02 to .38
Mock	3	763	-.05	.16	-.29 to .20
Field	8	4,579	.28	.24	.12 to .45
Group difference type					
Race	11	5,342	.20	.21	.02 to .38
Gender	9	3,734	.14	.40	-.06 to .33
Disability	-	-	-	-	-
Post-1996 studies					
Overall interview type	10	121,571	-.02	.07	-.06 to .01
Mock	7	730	-.03	.00	-.42 to .36
Field	3	120,841	-.02	.07	-.04 to .00
Group difference type					
Race	5	121,044	-.01	.08	-.04 to .01
Gender	4	121,986	-.03	.18	-.05 to .01
Disability	3	431	-.14	.00	-.47 to .19
All studies					
Overall interview type	21	126,913	.10	.18	.05 to .15
Mock	10	1,493	-.04	.06	-.36 to .29
Field	11	125,420	.18	.26	.15 to .22
Group difference type					
Race	16	126,386	.13	.21	.09 to .17
Gender	13	125,720	.09	.30	.05 to .13
Disability	3	431	-.14	.00	-.47 to .19

Note. *k* = number of studies; *N* = total sample size; *d* = mean effect size (negative *d* indicates that the raters were biased in favor of minority races, females, and/or individuals with disabilities); *SD_{res}* = residual standard deviation in *d* after removing sampling error variance; 95% CI = 95% confidence interval.

($\bar{d} = -.02$) indicates that minority applicants, women, and individuals with disabilities were actually rated slightly higher than other applicants, though the effect size is very small. In fact, confidence intervals for all group differences include zero, which indicates no meaningful differences. There was virtually no difference between mock and real interviews ($\bar{d} = -.03$ and $-.02$, respectively). Most interesting in these findings are the differences between the findings of the pre-1996 and post-1996 studies. We conducted several post hoc analyses in an attempt to explain these disparate results. These analyses showed no significant difference in the number of structure components utilized. Yet, a binominal test showed that the proportion of studies that required interviewers to rate applicants at the end of the interview (as opposed to after each question) was significantly higher in pre-1996 than in post-1996 studies ($p < .03$),

which might lead to increased susceptibility of ratings to different biasing factors.

Discussion and Future Research

The reviewed studies and the results of our meta-analysis suggest that structure reduces the impact of different biasing factors on interview ratings. Large-scale rigorously designed studies of structured interviews conducted in high-stakes contexts show little to no effect of demographic variables and demographic similarity in either panel or one-on-one interviews. Studies conducted in laboratory contexts also show that structured interviews are more resistant to different biasing factors (e.g., disability) than unstructured interviews. Our meta-analytic results suggest that group differences found in structured interviews were mostly negligible. However, the small number of studies included in our meta-analysis necessitates caution in interpreting these results.

Research should examine the amount of structure needed to reduce group differences. Kutcher and Bragger (2004) showed that interviews that are structured around both content and evaluation components of structure eliminate the bias against overweight applicants compared with unstructured interviews. Yet, interviews that are structured around only the content, but not the evaluation, components of the interview will mitigate, but do not eliminate, the bias against overweight applicants. In addition, our post hoc analysis suggests that requiring interviewers to rate applicants after each question (as opposed to rating at the end of the interview) might better decrease group differences.

Research Question 6: How much structure is enough or what components should be used to minimize group differences?

All research that examined the impact of disability on interview ratings in structured interviews was conducted in laboratory settings. Typically, undergraduate students were asked to watch videotaped interviews where confederates played disabled applicants. It is unclear whether or not the reported results will generalize to real high-stakes selection contexts. Although we are sensitive to the difficulty of conducting this kind of research in field settings, more realistic research designs are sorely needed in this area. If a real selection environment is not possible, serious attempts must be made to create high levels of psychological fidelity so that the same type of perceived accountability and meaningfulness is simulated with the research design. This could be enhanced in future research by conducting studies with authentically disabled individuals, and real interviewers, possibly in videotaped interviews.

Proposition 2: The impact of disability on ratings will be lower in structured field than in unstructured field interviews.

Impression Management in Structured Interviews

Impression management (IM) has been defined as a process by which people attempt to influence the images others form of them during social interaction (Schlenker, 1980). The employment interview may offer job applicants an opportunity to engage in IM due to its interpersonal nature, short duration, and high stakes. Past research has shown that applicants frequently engage in different types of verbal and nonverbal IM behaviors in unstructured interviews (Stevens & Kristof, 1995). Applicants may use nonverbal tactics, such as smiling or frequent eye contact. They may also use assertive IM to proactively construct images of being a good job applicant, including self-promotion IM tactics, such as self-promotion (e.g., claiming the responsibility for accomplishments), or other-focused IM tactics, such as ingratiation (e.g., evoking interpersonal attraction) and opinion conformity (e.g., claiming values held by the interviewer). Defensive verbal tactics (e.g., apologies) may be used to reactively repair negative images. Eighteen recent articles have been published on IM in structured interviews.

Main Recent Findings

The reviewed literature has explored the effect of interview structure on the effectiveness of applicant IM. Past research on unstructured interviews has demonstrated that applicants can increase their interview ratings when they engage in IM during interviews (Gilmore & Ferris, 1989; Stevens & Kristof, 1995). There is some belief that structure may reduce the effectiveness of IM because the evaluation process of structured interviews (e.g., using behaviorally anchored rating scales) narrows interviewer focus to job-related information, such that the influence of extraneous information (e.g., IM tactics) on interviewer decisions is minimized. One meta-analysis (Barrick, Shaffer, & DeGrassi, 2009) and eight primary studies since have explored this issue. Barrick et al.'s (2009) meta-analysis showed that structure moderated the effect of IM on interview scores, such that the relationship between verbal IM and interview ratings decreased from low-structure (.27) to high-structure interviews (.17). The authors also showed that the relationship between nonverbal IM and interview outcomes decreased as interview structure increased from low-structure (.40) to high-structure interviews (.30). To further explore this issue, we conducted a meta-analysis of primary studies to estimate the impact of

TABLE 4
Meta-analysis of Effects of Different Types of Impression Management on Interview Ratings

IM type	<i>k</i>	<i>N</i>	<i>r</i>	<i>SD_r</i>
Self-promotion	13	2,202	.26	.12
Other-focused IM	12	2,024	.13	.11
Defensive IM	12	2,075	.12	.09
Nonverbal IM	7	1,562	.18	.08

Note. IM = impression management; *k* = number of studies; *N* = total sample size; *r* = observed sample-weighted mean correlation; *SD_r* = observed sample-weighted mean standard deviation.

IM tactics on interview ratings.⁴ In an effort to be comprehensive in our meta-analysis, we also considered for inclusion the citations in Barrick et al.'s (2009) meta-analysis and other articles published prior to 1996. Only one of those studies (Riggio & Throckmorton, 1988) met our requirements: evidence of structure in the interview, measurement of IM, some form of interview score as a criterion, and adequate statistical information. The results of the meta-analysis (Table 4) indicate that self-promotion IM had the strongest impact on ratings ($r = .26$), followed by nonverbal IM ($r = .18$), other-focused IM ($r = .13$), and defensive IM ($r = .12$).

One of the main themes in the reviewed literature was that structure may reduce the use of IM tactics. The standardized process and content of structured interviews might provide applicants with less opportunity to direct or change the course of the interview through IM use. Five studies have been conducted to explore this issue (Ellis, West, Ryan, & DeShon, 2002; Lievens & Peeters, 2008; McFarland, Ryan, & Kriska, 2003; Peeters & Lievens, 2006; Tsai, Chen, & Chiu, 2005). Contrary to these expectations, the reviewed research showed that almost all applicants use some form of IM in structured selection interviews (Ellis et al., 2002). Yet, fewer applicants engage in IM in structured interviews used for promotion purposes (McFarland et al., 2003), likely because of probability of potential immediate verification of those claims.

Five studies (Ellis et al., 2002; Kleinmann & Klehe, 2011; McFarland et al., 2003; Peeters & Lievens, 2006; Van Iddekinge, McFarland, & Raymark, 2007) have also examined how the content of structured interviews, such as the use of SQ and PBQ, may offer applicants less of an opportunity to use IM. We meta-analyzed the findings across four studies that

⁴Articles included in this analysis are indicated with a "c" superscript (°) in the reference section.

TABLE 5
*Meta-analysis of Effects of Past-Behavior and Situational Questions on
 Impression Management Usage*

IM type	<i>k</i>	<i>N</i>	<i>M</i> _{PBQ}	<i>SD</i> _{PBQ}	<i>M</i> _{SQ}	<i>SD</i> _{SQ}	<i>d</i>	95% CI
Self-promotion	4	467	2.93	1.90	1.99	1.46	.48	.01 to .95
Other-focused IM	4	467	1.80	1.06	2.18	1.42	-.27	-.91 to .38
Defensive IM	4	467	1.63	1.16	1.30	.68	.46	-.08 to 1.00

Note. IM = impression management; *k* = number of studies; *N* = total sample size; *M* = mean; PBQ = past-behavior questions; *SD* = standard deviation; SQ = situational questions; *d* = mean effect size, positive values of *d* indicate that the IM tactics were used more often in past-behavior interviews, whereas negative values of *d* indicate that the IM tactics were used more often in situational interviews; 95% CI = 95% confidence interval.

compared SQs and PBQs,⁵ and the results are reported in Table 5. Other-focused IM tactics were found to be used more often in SQs ($M = 2.18$, $SD = 1.42$), whereas self-promotion IM and defensive IM tactics were found to be used more often in PBQs ($M = 2.93$, $SD = 1.90$, $M = 1.63$, $SD = 1.16$, respectively). However, only the finding for self-promotion IM had a confidence interval that did not include zero. These findings suggest that applicants' use of different IM tactics may depend on cues inherent in each question type. For instance, PBQs require applicants to focus on past job-related behaviors and accomplishments, and thus, promote their positive job-related qualifications. Conversely, SQs may instead encourage the use of other-focused IM, where applicants frame their answers to fit attitudes and values that, they believe, are held by the interviewer or the organization. Finally Peeters and Lievens (2006) found question type did not affect the use of nonverbal IM.

In 1989, Gilmore and Ferris called for the investigation of deceptive IM in interviews, but only recently have researchers answered that call. Four studies examined deceptive IM in structured interviews. Levashina and Campion (2007) argued that deceptive IM (when applicants describe nonexistent accomplishments) should be separated from honest IM (when applicants describe their real accomplishments). They argued that applicants can engage in four types of deceptive IM: (a) *slight image creation*—distorting or embellishing prior experiences or qualifications; (b) *extensive image creation* (EIC)—intentional fabrication or inventing experiences or qualifications; (c) *image protection* (IP)—intentionally omitting or masking undesirable experiences or qualifications; and (d) *deceptive ingratiation*—insincere praise of the interviewer or organization.

⁵In an effort to be comprehensive, we also searched for articles prior to 1996 that examined this topic, but no studies were found that met our criteria. Articles included in this analysis are indicated with a "d" superscript (d) in the reference section.

They showed that the majority of undergraduate applicants exhibited deceptive IM, that structure decreased faking, that applicants engaged more often in deceptive IM when answering SQs rather than PBQs, and that EIC positively impacted interview outcomes, whereas IP negatively impacted the interview outcomes. Swider, Barrick, Harris, and Stoverink (2011) found that self-promotion was positively related to mock-interview ratings, whereas EIC was negatively related to mock-interview ratings. Further, the authors showed that the relationships between IM tactics were moderated by interviewees' initial impressions of the interview such that interviewees who perceived they were seen by the interviewer as less suitable during rapport building were more effective when using IM tactics. Based on the videotaped interviews of students randomly assigned to three faking conditions (honest, realistic faking, and realistic faking with disclosure of what is being measured), Allen et al. (2004) showed that there were no statistically significant differences in interview scores among the conditions. Yet, Van Iddekinge, Raymark, and Roth (2005) showed that participants in the applicant-like condition received higher interview ratings than participants in the honest condition, suggesting that applicants can fake structured interviews. The authors also showed that the extent of faking in interviews is lower than faking on personality measures, suggesting that it is more difficult for applicants to fake structured interviews than personality tests.

Discussion and Future Research

Research on IM in structured interviews is still a relatively undeveloped area. As such, there are many provocative ideas as well as methodological issues that should be addressed by future research. First, it is unclear whether IM represents a bias in interviews. Researchers almost uniformly considered IM as a contamination variable that should be reduced or eliminated through structure. Yet, IM can take many forms, and some forms are less likely than others to represent a bias. For example, when applicants describe and emphasize the skills and experiences that they possess in their employment interviews, these behaviors are labeled as self-focused or self-promotion IM tactics (Ellis et al., 2002; Kristof-Brown, Barrick, & Franke, 2002; Tsai et al., 2005). On the other hand, some applicants may describe and emphasize skills and experiences that they do not possess (e.g., "I made positive events I was responsible for appear better than they actually were," Higgins & Judge, 2004, p. 632). These behaviors have also been labeled as self-focused or self-promotion IM. The first example does not illustrate a bias but rather represents an honest IM or the means by which job applicants describe their true job-related credentials during a short interaction with an interviewer. However, the second example represents deceptive IM or faking and does illustrate

a bias that should be controlled. Initial support for the notion that not all IM should be considered as a bias is provided by McFarland et al. (2003). They found that IM use in interviews related to both interpersonal and technical dimensions of performance. Moreover, some jobs, such as restaurant servers and casino employees, require job incumbents to only briefly interact with customers. During short interactions, employees must be perceived as competent, likable, and caring, which could be achieved when employees engage in “relevant” IM tactics (Barrick et al., 2010). For these jobs, IM could be considered to be job related, and thus it may be desirable to assess applicant IM use in interviews and use it later when making hiring decisions. Due to the nature of these jobs, it could be even more relevant to evaluate the first impressions interviewers form about applicants at the beginning of the interview (Barrick et al., 2010). For other jobs (e.g., engineers), such IM attempts would not be considered to be job related and thus would be deemed undesirable.

Future research should also examine when job-related and desirable IM impacts the validity of structured interviews. Job-related and desirable IM might influence the constructs being measured during the interview (e.g., planning, problem solving) because the constructs being measured could be distorted to some unknown extent by applicant IM behaviors. The challenge for interview research in this case is how to assess the influence of job-related and desirable IM tactics on constructs being measured during interviews.

Research Question 7: What type of IM reflects bias in structured interviews and what type reflects job-related competencies?

Second, it is unclear who is in the best position to accurately judge the presence of IM in interviews. The reviewed research has adopted two approaches. One approach involves asking trained coders to assess the frequency of IM tactics in audio or videotaped interviews (e.g., Ellis et al., 2002). Another approach involves asking job applicants to self-report IM use (Levashina & Campion, 2007). Yet, interviewers are the target of applicant IM, so it is surprising that interviewers’ perceptions of applicants IM have not been examined (Lievens & Peeters, 2008). Thus, more research is needed to examine interviewers’ perceptions of applicant IM use.

Research Question 8: (a) What structure components make interviewers more or less susceptible to honest and deceptive applicant IM? (b) How does interviewer’s perception of applicant IM impact interview ratings? (c) Can interviewers accurately perceive applicant deceptive and honest IM?

Third, more research is needed that explains why IM influences interviewer evaluations. The reviewed research showed that structure reduced the impact of IM on interview ratings, and different IM tactics have a different impact on interview outcomes. Yet, the reviewed research has failed to explore why IM impacts interviewers' ratings. Past research suggested that verbal IM might indicate applicant motivation or interest in a job (Bozeman & Kacmar, 1997), applicant social skills (Kristof-Brown et al., 2002), job-related skills (Levashina & Campion, 2006), integrity (Turnley & Bolino, 2001), or likelihood of future deviant behaviors at work (Griffith & McDaniel, 2006). Yet, more empirical research is needed to explore these underlying processes. Even less is known about why nonverbal IM influences interview outcomes. Recent research on emotional contagion (the process through which a person "catches" another's emotions; Hatfield, Cacioppo, & Rapson, 1994; Ilies, Wagner, & Morgeson, 2007) might be useful here. When applicants engage in positive nonverbal IM behaviors (e.g., smiling, positive facial expressions), the interviewers might "catch" the positive emotional state of the applicant and be more lenient in their evaluations.

Research Question 9: What processes underlie IM impact on ratings?

Fourth, a measure of honest IM should be developed and validated. Past research has frequently used a five-item self-report measure of self-promotion (Higgins & Judge, 2004, p. 632) that combines items measuring both deceptive (e.g., "I made positive events I was responsible for appear better than they actually were") and honest IM (e.g., "I described my skills and abilities in an attractive way"). Because recent research has started exploring the differences between deceptive and honest IM, separate measures of these constructs are needed. A measure of deceptive IM has been proposed (Levashina & Campion, 2007) and requires additional validation. A measure of honest IM should also be developed and validated. The proposed taxonomy of IM (Ellis et al., 2002) could serve as a starting point because Ellis et al. emphasized that they studied IM that "does not imply fabrication" (p. 1201).

Research Question 10: What is a valid self-report measure of honest IM?

Finally, research should examine other components of structure that lead to more or less IM, such as note taking, probing, interviewer training, interview length, and constructs being measured in interviews (e.g., communication skills, cognitive ability). For example, Peeters and Lievens (2006) argued that interview duration can impact IM use. IM could be more effective in short interviews, when interviewers have to evaluate applicants based on a minimal amount of information (Gilmore &

Ferris, 1989). Tsai et al. (2005) found that, when the interview was longer, the effects of applicant self-promotion tactics on interview outcomes was nonsignificant.

Proposition 3: Longer structured interviews decrease IM use and effectiveness.

Research Question 11: What other structure components decrease IM use and effectiveness?

Measuring Personality via Structured Interviews

Eighteen recent articles have been published on personality in structured interviews.

Main Recent Findings

In two meta-analyses (Cortina et al., 2000; Huffcutt, Conway et al., 2001), researchers content analyzed interview questions and concluded that personality was frequently assessed in interviews. Cortina et al. found that all of the studies included questions that measured some facets of conscientiousness such as initiative, work ethic, and thoroughness. Huffcutt, Conway et al. (2001) found that one-third of all interview questions in 47 actual interviews measured personality. Conscientiousness was the most frequently assessed trait.

Three meta-analyses and six primary studies (Barrick, Patton, & Haugland, 2000; Conway & Peneno, 1999; DeGroot & Kluemper, 2007; Huffcutt, Conway et al., 2001; Roth et al., 2005; Van Iddekinge, Raymark, Eidson, & Attenweiler, 2004) have examined the extent to which structured interview ratings are saturated with personality, as evidenced by examining the magnitude of the correlation between self-reports of personality and interview ratings. Cortina et al. (2000) reported a meta-analytic correlation of .26 between conscientiousness and ratings. Salgado and Moscoso's (2002) meta-analysis found mean observed correlations ranging from .04 for openness to experience and emotional stability to .10 for extraversion in structured interviews. Roth et al. (2005) conducted a meta-analysis of field studies and reported small mean correlations between self-reports and ratings: agreeableness ($r = .01$), openness to experience ($r = .03$), emotional stability ($r = .01$), extraversion ($r = .08$), and conscientiousness ($r = .12$). The results reported in primary studies suggest also relatively little saturation of structured interview ratings by personality.

Four primary studies have directly attempted to measure applicant personality via a structured interview. Barrick et al. (2000) asked experienced

interviewers to interview undergraduates and then assess their personality. Statistically significant self-interviewer correlations were found for extraversion ($r = .42$), openness to experience ($r = .34$), and agreeableness ($r = .30$). Interviewer ratings were also found to correlate higher with self-ratings ($r = .28$) than stranger's ratings ($r = .09$) but lower than ratings from close friends ($r = .39$). Blackman (2002) also reported significant positive correlations between interviewer-self and interviewer-close friend ratings (r s ranged between .43 and .61). Van Iddekinge et al. (2005) developed structured interviews to measure personality and asked experienced interviewers to interview undergraduate students instructed to respond honestly or as a job applicant. In the honest condition, statistically significant self-interviewer correlations were found for altruism (a facet of agreeableness, $r = .43$), self-discipline (a facet of conscientiousness, $r = .32$), and vulnerability (a facet of emotional stability, $r = .20$).

In the applicant-like condition, no significant self-interviewer correlations were found. DeGroot and Gooty (2009) explored interviewers' personality attributions based upon cognitive processing of nonverbal information. Conscientiousness attributions mediated the relationship between visual cues and ratings, whereas extraversion attributions mediated the relationship between vocal cues and ratings. Neuroticism attributions had a suppressing effect for both visual and vocal cues.

Researchers have suggested that the degree of interview structure should affect interviewer assessments of personality. It has been argued that applicant personality should vary more and be more readily manifest in an unstructured interview in part because it is not constrained by structure (restrictions on probing, question types, etc.). Two studies (Barrick et al., 2000; Blackman, 2002) examined how structure impacted an interviewer's assessment of applicant personality, but results were inconsistent. Barrick et al. reported small positive effects for more structured interviews by showing that interviewer's ratings converge with applicant self-ratings more in structured ($r = .37$) than unstructured ($r = .22$) interviews. Yet, Blackman reported positive effects for unstructured interviews by showing higher average self-interviewer ($r = .61$) and interviewer-close friend correlations ($r = .56$) for unstructured interviews than for structured interviews ($r = .50$ and $r = .43$, respectively).

Discussion and Future Research

The growth in research on personality in structured interviews is likely to have occurred for at least two reasons. First, because of its positive relation to job performance and generally low levels of adverse impact, personality-oriented selection has become quite popular in the last 20 years. Second, structured interviews designed to measure personality

might be a better way to measure personality than self-report measures because interviewers can observe and rate job applicant personality, thereby minimizing the socially desirable responding thought to occur with self-report measures (Morgeson et al., 2007).

Research findings that suggest relatively little saturation of structured interview ratings by personality factors do not necessarily mean that interviewers cannot or do not measure personality. Structured interviews are typically designed to measure multiple job-related skills and abilities, where personality requirements may play only a small part. Thus, personality saturation in interview ratings might exist only for jobs which require applicants to possess given personality traits and those specific traits are actually measured in the interview. Because the structured interview is a method that can be used to measure different constructs (e.g., personality, knowledge, skills, ability, etc.), researchers must be cognizant of the construct and method distinction (Arthur & Villado, 2008) when examining personality in structured interviews. In addition, because a majority of the reviewed studies on saturation were conducted in field settings where applicants would be very motivated to distort their responses on self-report personality tests, the small correlations between applicant self-reports of a personality trait and interview ratings of the same trait could be due to faked self-reports. We recommend less research on personality saturation in structured interviews in favor of more research that examines direct assessment of personality in structured interviews, as described below.

Research should further examine personality-based structured interviews as an alternative method of applicant personality self-assessment. Conceptually, personality is expected to predict various important job-related outcomes, such as job performance (Hurtz & Donovan, 2000), turnover (Barrick & Mount, 1996), and absenteeism (Judge, Martocchio, & Thoresen, 1997). Yet, the validity of self-report personality measures in predicting these outcomes has been low. A group of former journal editors suggested that future research should focus on identifying alternatives to self-reports (Morgeson et al., 2007). We believe that the personality-based structured interview may represent a viable alternative method. One of the main issues with self-reports is faking. Van Iddekinge et al. (2005) provided some evidence that personality-based structured interviews are more resistant to applicant faking than self-reports. Interviews allow separation of the description of behaviors from the evaluation of described behaviors. On self-reports, applicants are asked to evaluate and assign ratings to themselves. In interviews, applicants are asked to describe their behaviors, and interviewers are asked to evaluate described behaviors and to assign ratings. Thus, job applicants have less control over interview scores compared to self-report personality measure scores, potentially resulting in less response distortion.

Research Question 12: Are personality based structured interviews more resistant to applicant faking than self-report personality measures?

Recent meta-analytic results show that observer ratings of personality are more predictive of job performance than self-reports, with observer validity about twice the magnitude of the self-report validity (Connelly & Ones, 2010; Oh, Wang, & Mount, 2011). In the context of structured interviews, similar meta-analytic results were reported by Huffcutt, Conway et al. (2001), showing that correlations between structured interview and job performance ratings varied from .16 for openness to experience to .31 for emotional stability. The underlying factor structure of interview ratings was also unaffected by faking, unlike self-reports (Van Iddekinge et al., 2005).

Proposition 4: When designed to measure the same personality traits, personality based structured interviews have higher criterion-related validity than self-report personality measures.

Future research should also explore how to design personality-based interviews in terms of questions, ratings, and other factors. Personality can be defined as a consistency of behaviors across times, situations, and circumstances. As such, personality-based interview questions could be designed to measure the specific job-related behaviors that are presumed to underlie a particular personality trait. For example, conscientiousness could be decomposed into such behaviors as high levels of effort, goal-directed behavior, meeting deadlines and obligations, and/or detail-oriented behavior. In addition, structured interviews could be designed in a way that allows one to assess behavioral consistency across time and circumstances. For example, personality-based interviews could include both PBQs and SQs. Interviews that include both types of questions provide an opportunity to assess both past and future behaviors and to evaluate their consistency. To measure consistency of behaviors across situations, personality-based interviews could also have interviewers provide an overall evaluation of a particular behavior based on several questions regarding the same behavior in different situations.

Interviews may also offer a unique opportunity to assess consistency between reported and displayed behaviors. Personality-based structured interviews could be used to elicit a *description* of trait-congruent behaviors (Haaland & Christiansen, 2002). In addition, *displayed* trait-congruent behaviors during the interview could be evaluated. For instance, a PBQ such as “Tell me about a time when you had a work-related disagreement

with your team members. How did you deal with the situation?” could be used to measure agreeableness and to elicit an applicant *description* of agreeableness-congruent behaviors. At the same time, the interviewer could monitor *displayed* agreeableness-congruent behaviors, such as being compassionate, good-natured, cooperative, and avoiding conflicts. Any observed inconsistencies of reported behaviors across times, situations, and inconsistencies between reported and displayed behaviors could be used as an indicator of faking or response distortion. Research could also explore which traits are most effectively measured in interviews. Research suggests that some personality traits are easier to observe than others (Funder, 1995, 1999). Indeed, personality research has found that judgment accuracy is higher for more observable traits such as extraversion and conscientiousness than for less visible traits such as emotional stability and openness to experience (e.g., Borkenau & Liebler, 1993; Connelly & Ones, 2010). Yet, some researchers suggest that an interview is an example of a “strong situation” in which behaviors that are congruent with applicant personality may be replaced by more planned situation-specific behavior (Cortina et al., 2000). For example, even if extraversion is a more visible trait, applicants who do not possess that trait can manage their answers in interviews to make an impression that they possess that trait.

Proposition 5: Reliability and validity of personality-based structured interviews will be higher when personality-based interview questions are designed to (a) measure the specific job-related behaviors that are presumed to underlie a particular personality trait; (b) assess behavioral consistency across time and circumstances; (c) assess consistency between reported and displayed behaviors; and (d) measure more observable (e.g., extraversion, conscientiousness) than less observable (e.g., emotional stability, openness to experience) traits.

Comparing Situational Versus Past-Behavioral Questions

Situational (Latham, Saari, Pursell, & Campion, 1980) and past-behavior (Janz, 1982) questions have emerged as the most popular types of structured interview questions. SQs are based on goal-setting theory and rely on the assumption that intentions predict future behavior (Latham, 1989; Locke & Latham, 1990). As such, SQs ask applicants to describe what they would do in hypothetical job-related situations. One of the core aspects of SQs is a dilemma, which requires an applicant to choose between two or more exclusive courses of action (Latham & Sue-Chan,

1999). Conversely, PBQs are based on the premise that past behavior predicts future behavior (Janz, 1982). As such, PBQs ask applicants to describe what they did in past job-related situations. Campion et al. (1997) reported 17 studies with mixed results regarding reliability and validity of SQs and PBQs, and called for more research to compare (a) reliability, (b) construct and criterion-related validity, (c) fakability, and (d) experience requirements of SQs and PBQ.

Main Recent Findings

Nineteen articles have been published on this topic since Campion et al. (1997). One of the main themes of the reviewed research was an examination of the reliability, construct, criterion-related, and incremental validity of SQs and PBQs. A meta-analytic comparison found quite similar levels of interrater reliability, with estimates of .79 and .76 for SQs and PBQs, respectively (Taylor & Small, 2002).

Four primary studies explored the construct equivalence of SQs and PBQs designed to assess the same job-related characteristics. Conway and Peneno (1999) found that SQs reflect basic job knowledge whereas PBQs reflect experience. Huffcutt, Weekley, Wiesner, DeGroot, and Jones (2001) found that PBQs were related to extraversion ($r = .30$), but neither question type was related to either mental ability or job experience. Yet, Day and Carroll (2003) found that PBQs were significantly related to experience ($r = .33$) and SQs were significantly related to cognitive ability ($r = .22$). Finally, Krajewski, Goffin, McCarthy, Rothstein, and Johnston (2006) found that PBQs were related to cognitive ability and to the personality traits of achievement orientation and dominance, whereas SQs were not. In sum, the reviewed studies seem to suggest that when SQs and PBQs are carefully designed to assess the same characteristics, they tend to measure different constructs, with SQs primarily measuring job knowledge or cognitive ability and PBQs primarily measuring experience and perhaps some personality facets.

Recent research has also examined the equivalence of SQs and PBQs with respect to their capacity to predict job performance. Three meta-analyses and three primary studies (Day & Carroll, 2003; Gibb & Taylor, 2003; Klehe & Latham, 2006) have explored the criterion-related validity of SQs and PBQs. Meta-analytic results suggest that both question types have criterion-related validity, with PBQs demonstrating slightly higher validity. Latham and Sue-Chan (1999) reported the uncorrected mean validity of SQs as .35. Taylor and Small (2002) reported the mean uncorrected validity of SQs and PBQs as .25 and .31, respectively. Similarly, Huffcutt et al. (2004) reported the validity of SQs and PBQs as .26 and .31, respectively. Finally, Klehe and Latham (2006) showed that SQs

predicted both typical ($r = .41$) and maximum performance ($r = .25$), but PBQs predicted only typical ($r = .34$) and not maximum performance ($r = .11$).

Researchers have suggested that the criterion-related validity of SQs and PBQs might be moderated by several variables, such as job complexity, the use of anchored rating scales, prior work-related experience, and study design. Two meta-analyses and three primary studies (Huffcutt, Weekley et al., 2001; Klehe & Latham, 2005; Krajewski et al., 2006) have examined the moderating effect of job complexity, with inconsistent results. A meta-analysis conducted by Taylor and Small (2002) found that job complexity slightly moderates the validity of both SQs and PBQs such that the validity of SQs and PBQs is decreased for high-complexity jobs. They found a mean uncorrected validity for SQs of .26 for low-complexity jobs and .23 for high-complexity jobs, and for PBQs of .40 for low-complexity jobs and .31 for high-complexity jobs. A meta-analysis conducted by Huffcutt et al. (2004) found that job complexity moderates the validity of SQs but not PBQs. They found a mean uncorrected validity for SQs of .27 for low-complexity jobs and .18 for high-complexity jobs, but similar numbers for low- and high-complexity jobs for PBQs (.30 and .31, respectively). Later, Klehe and Latham (2005) showed that both SQs ($r = .41$) and PBQs ($r = .34$) predict team playing behavior equally well for high-complexity jobs. Yet, Krajewski et al. (2006) showed that PBQs significantly predicted job performance ($r = .32$) for complex jobs, whereas SQs did not ($r = .09$, *ns*).

One study has addressed Campion et al.'s (1997) call to examine the moderating effect of prior job experience. Gibb and Taylor (2003) found no evidence that job experience moderates the validity of PBQs, suggesting that applicants with little or no prior job experience may use nonwork experience in responding to PBQs, which is common in practice.

One meta-analysis explored the moderating effect of study design on the validity of SQs and PBQs. Huffcutt et al. (2004) reported that predictive studies have .10 lower validity on average for both SQs and PBQs than concurrent studies. This could be due to the fact that incumbents in concurrent studies used experience from their current positions when answering interview questions, leading to the increased relationships with job performance. Lower validity of predictive studies could be also due to applicant intentional response distortion in interviews.

Finally, two studies have examined the incremental validity of SQs and PBQs with opposite results. Klehe and Latham (2005) found that SQs had incremental validity over PBQs, but PBQs did not have incremental validity over SQs. The opposite results were found by Krajewski et al. (2006).

Discussion and Future Research

The reviewed research suggests that SQs have slightly lower criterion-validity than PBQs, especially for high-complexity jobs. Yet, this finding does not necessarily mean that only PBQs should be used, as both SQs and PBQs have good validity for most jobs. Instead, researchers commonly recommend the use of both question types, in part because SQs and PBQs tend to measure different constructs and may complement each other (e.g., Campion et al., 1997; Day & Carroll, 2003; Huffcutt, Weekley et al., 2001; Krajewski et al., 2006) to increase overall criterion-related validity and also because they allow some variety in question format. As such, future research should examine SQs and PBQs as separate testing devices and how they can complement each other to improve criterion-related validity of and user reactions to structured interviews that include both types of questions.

Proposition 6: Structured interviews that include both SQs and PBQs will have (a) higher criterion-related validity and (b) more positive user reactions than structured interviews that include either SQs or PBQs alone.

Future research should also examine SQ design. Originally, Latham and colleagues (Latham et al., 1980) argued that posing a dilemma between competing values is an essential component of the SQ design. For example, sawmill workers were asked to choose between attending work and caring for ill family members (Latham et al., 1980). However, as noted in two meta-analyses (Huffcutt et al., 2004; Taylor & Small, 2002), many researchers have used SQs that contained no dilemmas. Thus, it is important to explore the potential differences between SQs with and without dilemmas for several reasons. First, Klehe and Latham (2005) argued that SQs with and without dilemmas are two different types of interview questions and as such should have different labels; SQs without dilemma can be thought of as future questions (FQs; e.g., Campion, Campion, & Hudson, 1994). Second, a dilemma is included to minimize the transparency of the assessed dimensions during interviews and, thus, to decrease applicant faking in interviews (Latham et al., 1980). However, this assumption has yet to be tested.

Third, SQs and FQs might measure different constructs. Originally, Latham and colleagues argued that SQs should measure intentions or motivation to behave in a particular way on the job. However, most of the reviewed research suggests that SQs measure job knowledge. Thus, instead of reflecting an applicant's intentions to behave in a particular way, it assesses an applicant's knowledge of how to deal with situations

described in SQs (Motowidlo, 1999). In addition, it is possible that SQs measure motivation, as originally suggested, and FQs measure job knowledge. Fourth, SQs and FQs might require different anchored rating scales. For instance, several researchers pointed out that some SQ ratings were designed to reward a single action whereas other ratings rewarded more elaborated and sophisticated responses (Huffcutt, Weekley et al., 2001; Taylor & Small, 2002). SQs designed to measure intentions may use rating scales that give high scores to applicants who describe an appropriate single action. FQs designed to measure job knowledge may use rating scales that give high scores to applicants who describe not only the action but why they would take that action, how they come up with that action, and the presumed advantages and disadvantages of other possible alternatives.

Proposition 7: SQs will show discriminate validity from FQs when they are designed to measure the same constructs.

Probing, Follow-up, Prompting, and Elaboration on Questions

When the initial response of the applicant to the interview question is incomplete or inadequate, the interviewer may ask further probing questions to help the applicant complete his or her answer. Probing has been examined in four recent articles.

Main Recent Findings

Blackman (2002) found that probing is used more often in unstructured ($M = 4.19, SD = 2.64$) than in structured mock interviews ($M = .60, SD = 1.10$). Levashina and Campion (2007) found that probing increased faking for both PBQs and SQs. Two surveys described the frequent use of probing in real interviews in the United States (Chapman & Zweig, 2005) and in Belgium (Lievens & De Paepe, 2004).

Content Analysis of the Use of Probing

Campion et al. (1997) proposed four levels of probing structure. The first and highest is the prohibition of any probing. The second level is to allow only limited or preplanned probing. The third level is to allow and encourage unlimited probing. The fourth and lowest level is no guidance on probing. Using these levels, we content analyzed 14 recent empirical articles (17 total probing conditions) to identify levels of probing that were used in structured interviews. Articles were included if they discussed or tested probing. Meta-analyses, surveys, and conceptual articles were excluded. The highest level was used in eight

articles (Blackman, 2002; Buckley et al., 2007; Ellis et al., 2002; Huffcutt, Weekley et al., 2001; Klehe & Latham, 2006; Levashina & Campion, 2007; Sue-Chan & Latham, 2004; Townsend et al., 2007). Eight articles reported using the second level (Klehe & Latham, 2006; Levashina & Campion, 2007; Lievens, Harris, Van Keer, & Bisqueret, 2003; McCarthy et al., 2010; Morgeson, Reider, & Campion, 2005; Sacco et al., 2003; Tross & Maurer, 2008; Van Iddekinge, Sager et al., 2006). The third level was used once (Blackman, 2002). It is unclear whether the lowest level was used in all other articles that did not describe this component or if researchers just omitted the description of probing.

Discussion and Future Research

Many researchers have argued that controlling probing may be a defining element of structured interviews (Dipboye et al., 2004; Huffcutt & Arthur, 1994). It may help to control interviewer biases in conducting interviews and collecting information about applicants (Schwab & Heneman, 1969). Yet, many practitioners and consulting firms have argued that probing may be a means of improving the accuracy of information gathered in an interview (Schmidt & Conaway, 1999; targeted selection, <http://www.ddiworld.com>; behavioral description interview, <http://www.pdinh.com>) because it provides interviewers with an opportunity to delve deeper into applicant responses to seek additional explanations, clarifications, and justification. It is unclear what impact probing has on interview outcomes because the research on probing in structured interviews is almost nonexistent. Following Campion et al. (1997), we call for more research on this component.

First, future research should provide a better understanding of what probing is. Research has not provided a definition of probing. Multiple terms such as *follow-up questions*, *probing*, *push-back*, and *prompting* have been used. It is unclear whether these words are used interchangeably or whether they refer to different types of probing. As a starting point, we offer a definition of probing as *a follow-up question that is intended to augment an inadequate or incomplete response provided by the applicant, or to seek additional or clarifying information*.

Second, following Campion et al. (1997) we call for more research exploring the extent to which probing should be allowed in structured interviews. The results of our content analysis suggest that limited probing was used as often as no probing in structured interviews. Yet, researchers often failed to define how probing was limited. For example, Klehe and Latham (2006) did not use probing in SQs. Yet, they use probing in PBQs when applicants responded that they had never experienced a given situation.

Third, future research should examine the impact of probing on interview outcomes. The use of unlimited probing may lead to low reliability, validity, and increased susceptibility to different biases because interviewers will have more discretion in what they ask and how they ask it. Schwab and Heneman (1969) showed that interrater agreement was higher when interviewers were not allowed to probe. On the other hand, the use of planned or limited probing may help interviewers collect more job-related information during the interview, which may lead to increased interview validity. Planned probing could be used to help applicants who might be shy or speak in succinct ways to clarify their answers and provide more detailed job-related information. Probing might also help applicants recall and identify the most relevant experiences they need to describe in order to best present their true job-related credentials.

Proposition 8: Planned probing will increase (a) validity and (b) lead to more positive user reactions than unlimited probing or no probing.

Fourth, future research should explore the effect of probing on faking and other forms of IM in interviews. Some researchers have argued that probing can decrease faking. For example, probing can be used to make interview questions more difficult. McCarthy et al. (2010) used probing for SQs that challenged the applicants by eliminating obvious answers. In addition, probing can be used to ask applicants to elaborate on their responses by providing additional evidence to verify their responses. Research shows that requiring elaboration for biodata items reduces faking (Levashina, Morgeson, & Campion, 2012; Schmitt & Kuncze, 2002). Yet, some empirical findings suggest that probing increases faking. For example, Levashina and Campion (2007) hypothesized that probing would be a response verification mechanism that would inhibit faking but found that probing actually increased faking in both SQs and PBQs. Informal debriefing with applicants revealed that probing was perceived as a cue signaling what types of answers were important and critical. Probing may also lead to faking in situations when the applicant lacks prior experience related to the question but is then encouraged by the interviewer to provide an example, regardless of context. Research on public opinions (e.g., Sanchez & Morchio, 1992) suggests that probing “don’t know” responses to knowledge and opinion questions leads to guessing and inflated responses. Similarly, probing “never experienced” responses may force applicants to create situations in order to please the interviewer and provide the requested response.

Research Question 13: Does probing increase or decrease faking and other forms of IM?

Finally, a typology of probing should be developed and used to accumulate knowledge regarding which types of probing can be used to obtain more job-related information without increasing applicant faking. For instance, enhancing difficulty (e.g., “What would you do if this course of action does not work?”), verifying (e.g., “Who can verify your accomplishments?”), and quantifying (e.g., “You mentioned that you were successful in managing your team. How many members were in your team? What outcomes were achieved?”) are more likely to reduce faking by making the question more difficult or increasing accountability.

Proposition 9: Enhancing difficulty, verifying, and quantifying probing decreases faking in structured interviews.

Until research demonstrates whether probing increases or decreases faking and validity, a practical recommendation is to only use planned neutral probes that will not cue the applicant to desired answers, to probe equally across all applicants, allow the same time limit for answers, and only probe if the answer is clearly deficient, or off track, or the applicant is shy.

Developing Anchored Rating Scales

Anchored rating scales (ARSs) have been developed to assist interviewers with applicant evaluation. ARSs can simplify the complex judgmental tasks by providing behavioral, descriptive, or evaluative examples to illustrate points on the rating scale. To evaluate responses, the interviewer matches the applicant's responses with the written anchors for each question. Campion et al. (1997) reported 10 studies that have examined ARSs and suggested that ARSs should increase reliability and accuracy, reduce bias, and lead to positive court decisions. Yet, the authors explained the popularity of ARSs by their logical appeal rather than strong research support and called for more research on ARSs. Four recent articles have explored ARSs.

Main Recent Findings

Taylor and Small (2002) meta-analyzed 19 past-behavior interview studies to estimate criterion-related validity and interrater reliability for interviews with and without ARSs. They found that the use of ARSs resulted in higher validity estimates ($r_{xy} = .35$) and interrater reliability ($r_{xx} = .77$) than when rating scales were not used ($r_{xy} = .26$, $r_{xx} = .73$). Improved validity could be produced by increased reliability, greater job relatedness, or a combination of these two factors (e.g., Anastasi & Urbina, 1997). Because anchors are based on and include examples of job behaviors or requirements, they also might increase job relatedness,

which is likely to enhance validity. Maurer (2002) found that job experts do not produce better ratings than naïve raters when behaviorally anchored ratings scales (BARSs) were used, and BARSs enhanced interrater agreement and accuracy of ratings. Reilly, Bocketti, Maser, and Wennet (2006) explored the susceptibility of interview ratings to disability bias in interviews with and without ARSs. They demonstrated that the use of ARSs reduces the impact of disability bias on interview ratings by providing interviewers with objective behavior standards and, thus, enhancing interviewers' attention to relevant job-related behaviors. The authors also found that the number of anchors used in ARSs affects susceptibility of interview ratings to disability bias, such that the use of a five-point scale with anchors at each point results in interview ratings that are more resilient to disability bias than interview ratings obtained with the use of a five-point scale with two anchors (for the highest and lowest points), or no anchors.

Discussion and Future Research

The reviewed studies consistently demonstrate that ARSs enhance reliability and validity of structured interviews by controlling different biases, making the same job-related information salient to all interviewers, and helping to ensure that applicant responses are rated consistently across interviewers. Despite the importance of ARSs to structured interviews, the science of rating scales is still surprisingly underresearched. As such, following Campion et al. (1997), we call for more research on this component.

Future research should explore whether examples or descriptions are better anchors. Example anchors include the language that job applicants may use in their responses (e.g., "My work ethic helped set the standard for others"). Description anchors include broad categories or definitions of applicant answers without using potential applicant words (e.g., "Described how he/she models the desired behaviors for others"). Both types of anchors were used in recent research, but it is unclear whether one is better than the other. For instance, the use of examples as anchors might lead to deficiency because it is difficult to identify examples of answers that will represent all potential applicant answers. On the other hand, the use of descriptions as anchors might increase cognitive complexity of the rating task because raters need to match specific phrases to the general description of answers (Green, Alter, & Carr, 1993).

Research Question 14: Are examples or descriptions better anchors?

Future research should also explore how many anchors should be used to enhance psychometric properties of the interview. Are five anchors on

a five-point scale needed or will three anchors at the low, middle, and high points be sufficient? For example, Sacco et al. (2003) used a nine-point rating scale with only three anchors for ineffective, effective, and highly effective responses. Yet, Reilly et al. (2006) demonstrated that interview ratings were less susceptible to bias when all points on a scale were defined by anchors. When one anchor is provided for multiple points on a rating scale, or only some of the points have their own anchors, interviewers have more discretion in their ratings which might lead to bias and decreased reliability and validity. Given the potential difficulties interviewers have in making fine distinctions, anchoring every point would seem advantageous, although it is true that it is more difficult to develop anchors for each scale point. In addition, in practice, answers often fall in between anchors.

Proposition 10: The use of ARSs with all anchors defined with behaviors, descriptions, or examples will lead to higher validity and reliability of structured interviews than the use of ARSs when not all of the anchors are defined.

Future research should also explore whether extensive anchors are better than brief anchors. On the one hand, extensive anchors are more likely to cover all possible applicant answers and thus reduce contamination and deficiency of ratings. On the other hand, extensive anchors can increase cognitive complexity of the rating process because interviewers need to scan extensive anchors to match the applicant responses to the scale. The use of extensive anchors might lead to negative interviewer reactions and interviewers' unwillingness to use them in interviews (Green et al., 1993). On the other hand, extensive anchors are more useful for training and may require less note taking because interviewers can circle answers.

Proposition 11: The use of extensive anchors will yield higher reliability and validity of structured interviews than the use of brief anchors.

Research Question 15: Is there an upper limit on the use of extensive anchors such that there are diminishing returns on increased reliability and validity of structured interviews?

Reactions to Structured Interviews

Twelve recent articles have been published on this topic, six articles on applicant reactions and six articles on interviewer reactions. Campion et al. (1997) noted too little empirical research examining user reactions

at the time to truly understand much about the topic. They speculated about the possible influence each of the 15 components may have on user reactions and proposed 22 research questions, which led to the research reviewed below.

Main Recent Findings

Applicants' perception of fairness of the structured interview was analyzed in three studies. Conway and Peneno (1999) found no differences in procedural justice perceptions between structured and unstructured interview questions in an actual high-stakes interview setting, yet Chapman and Zweig (2005) found that applicants had more negative perceptions of procedural justice with higher levels of structure. Day and Carroll (2003) found no difference in fairness perceptions between situational and past-behavior interviews.

Four studies explored basic applicant reactions to structured interviews. Conway and Peneno (1999) found that applicants have less positive affective reactions toward structured than unstructured interviews. Chapman and Rowe (2002) report less satisfaction with interview performance in structured than unstructured interviews. Chapman and Zweig (2005) found that applicants perceive structured interviews to be more difficult than unstructured interviews. In a case study conducted by Molgaard and Lewis (2008), applicants viewed the structured interview as relevant, straightforward, uncomplicated, and easy to understand.

Another important applicant reaction is the perception of the organization or interviewer as a result of increased interview structure. Two studies (Chapman & Rowe, 2002; Kohn & Dipboye, 1998) found that structured interviews were associated with less positive evaluations of the organization, but the two studies found differing results for applicants evaluating the interviewer. Kohn and Dipboye found that structured interviews were associated with less positive evaluations of the interviewer, but Chapman and Rowe found no significant difference. Molgaard and Lewis (2008) found that applicants had positive perceptions of the interviewers.

Three studies also examined moderators of applicant reactions. Chapman and Rowe (2002) examined the effects of the mode of interviewing, using both face-to-face and videoconference. The evaluation of the organization did not significantly change between structured conditions in the videoconference interviews but was significantly lower when structured interviews were used in the face-to-face interviews. Two studies examined the impact of information on applicant reactions. Day and Carroll (2003) found that giving applicants advanced knowledge of interview questions resulted in significantly higher perceptions of fairness. Kohn and

Dipboye (1998) found that information on the job and organization reduced the negative effect of interview structure on evaluations of the organization, interviewer, and interview. Younger participants had significantly more negative reactions to interview structure, though the differences were small. Negative reactions were also more pronounced for individuals with high job-related self-esteem and low openness to experience.

Interviewer reactions. Four studies have examined interviewers' perceptions of structured interviews with inconsistent findings. Two studies found that interviewers have a positive view of structured interviews and incorporate at least a modest degree of structure in their interviews (Camp, Schulz, Vielhaber, & Wagner-Marsh, 2004; Lievens & De Paepe, 2004); yet, van der Zee, Bakker, and Bakker (2002) indicated that managers' intentions to use structured interviews were weaker than intentions to use unstructured interviews. Terpstra and Rozell (1997) reported that the most dominant reason for interviewers and organizations not using structured interviews was the failure to see them as useful or effective.

Three studies examined moderators of interviewers' reactions, such as experience, cognitive style, values, and race. Four studies found that interviewers with training or prior experience conducting structured interviews have positive perceptions of structured interviews (Barclay, 2001; Camp et al., 2004; Chapman & Zweig, 2005; Molgaard & Lewis, 2008). Yet, one study found that experienced interviewers had more negative reactions toward structured interviews, presumably because they perceived highly structured interviews to be more rigid (Chen, Tsai, & Hu, 2008). Finally, one study found that interviewer experience did not have a significant effect on the degree of interview structure used (Lievens & De Paepe, 2004).

Chen et al. (2008) found that interviewers' cognitive style was related to their reactions to structured interviews such that interviewers with an analytic orientation (vs. interviewers with an intuitive orientation) were more likely to have positive reactions to structured interviews. No significant relation was found between interviewers' need for power and their reactions to structured interviews. Lievens and De Paepe (2004) found that interviewers that value a high level of discretion when conducting interviews are less inclined to use structured interviews. Finally, Camp et al. (2004) found that race was a moderator in interviewers' perceptions of the process dynamics of panel interviews but not in their perception of organizational outcomes or their affective reactions to interviews.

The focus of the interview was found to have an effect on interviewers' intentions, in that interviewers with a selection focus tend to favor the use of structured interviews more than those with a recruitment focus (Chapman & Zweig, 2005; Lievens & De Paepe, 2004). Job complexity, organizational interviewing norms, and organizational size were also

found to moderate other individual differences (cognitive style and need for power) and the interviewers' reactions to structured interviews (Chen et al., 2008).

Discussion and Future Research

Interviews are consistently viewed by applicants across cultures as one of the most preferred selection methods (Anderson, Salgado, & Hülsheger, 2010). From the perspective of both validity and legal defensibility, it is probably an unwise decision not to use structured interviews. Yet, despite this evidence, many interviewers fail to use structured interviews. The research on interviewer and applicant reactions has the potential to reveal and address barriers that might exist in the implementation of structured interviews, including concerns about applicant preferences for unstructured interviews, interviewer desire for discretion, recruiting impact of structured interviews, decreased job offer acceptance intentions, decreased interviewer compliance with structured interview protocols, and compatibility of structured interview with the organizational or national culture.

First, the structured interview often functions as both a recruiting and selection tool. The reviewed research suggests that interviewers and organizations perceive structured interviews to be less effective in recruiting than unstructured interviews, yet it is unclear what components of structured interview contribute to such perceptions. For example, candidates may react negatively to being unable to ask questions during the interview but may not even be aware of the rating scales used. Future research that examines potential trade-offs between psychometric improvements and user reactions must be conducted by examining the reactions to the specific components of structure. By examining components separately, it may be possible to identify those with the most negative expected impact, examine their contribution to reliability and validity, and decide to either eliminate the component in order to increase positive reactions to structured interviews or keep it in order to increase psychometric properties of the interview.

Research Question 16: What are the potential trade-offs between psychometric improvements and user reactions to individual components of the structured interview?

Second, one of the most practical applicant reactions for organizations to understand is whether or not using structured interviews affects an applicant's intentions of accepting a job offer. Given the high practical significance of this question, it was disappointing to find that only one article examined it. Using college students interviewing for short-term

positions, Chapman and Zweig (2005) found that applicants had lower intentions of accepting a job offer when higher levels of interview structure were used. Yet, they did not examine whether the intentions of rejecting a job offer led to actual job offer rejection.

Research Question 17: Do structured interviews affect applicant's intentions of accepting a job offer and/or actual job offer rejections in a real selection context?

Future research should also examine the potential moderators of this relationship, including applicant qualifications. It is possible that applicants with low levels of job-related qualifications might resent the focus of structured interviews on measuring those qualifications, whereas applicants with high-levels of job-related qualifications might prefer structured interviews. If research finds that lower acceptance intentions are a function of attitudes of the lower skilled applicants, this might support the use of structured interviews.

Proposition 12: Applicant qualifications will moderate the impact of interview structure on applicant reactions such that applicants with low levels of job-related qualifications will have negative perceptions of structured interviews, whereas applicants with high levels of job-related qualifications will have positive perceptions of structured interviews.

Third, structured interviews can only be successfully implemented when interviewers comply with the established structure. Even when organizations invest in the development and implementation of structured interviews, interviewers may choose to deviate from the established procedures. Although interviewers' compliance with structured interview protocols is very important, it is often overlooked in practice. Interviewers may deviate from the structured protocol when they perceive loss of control or view the structured interview as a boring and monotonous exercise (Dipboye et al., 2004). Interviewers may follow the structured protocol when they perceive structure as a way to reduce their cognitive load and to increase the quality of their decisions about applicants. Finally, interviewers may follow the structured protocol when they are monitored (e.g., interviews are recorded). Future research should examine the reasoning behind the interviewers' compliance or noncompliance with interview structure.

Research Question 18: When and why will interviewers fail to follow structured interview protocols?

TABLE 6
Major Findings and Recommendations for Future Research on Structured Interviews

Major findings	Major recommendations for future research
<p>Definition of structure</p>	
<p>(1) Campion et al.'s (1997) 15 component-taxonomy of interview structure is the most comprehensive typology of interview structure.</p>	<p>(1) A more consistent conceptualization and operationalization of the structured interview is needed. At a minimum, articles should describe the presence or absence of Campion et al.'s (1997) 15 components of structure in order to make comparisons of structured interviews between articles more apparent.</p>
<p>(2) Interviews are usually structured with six components, such as using job analysis, same questions, better types of questions, using anchored rating scales, rating each question, using multiple interviewers, and providing training.</p>	<p>(2) Limiting rapport building, interview transparency, and recording of interviews should be included into the interview structure framework.</p> <p>(3) Direct research is needed to understand reliability and the incremental validity of the different components of structure.</p> <p>(4) In addition to traditional face-to-face interviews, research should also examine other forms of structured interviews that are becoming increasingly common, such as computer-mediated and phone interviews.</p>
<p>Reducing group differences through structure</p>	
<p>(1) The use of structure in interviews greatly reduces group differences based on race, gender, and disability in structured interview ratings.</p>	<p>(1) Higher quality research and more field research is needed to better understand the generalizability of previous lab-based interview research.</p> <p>(2) Future research should examine how much structure is enough to reduce group differences.</p>
<p>Impression management in structured interviews</p>	
<p>(1) Applicants use IM extensively in structured interviews.</p>	<p>(1) Research should examine when IM should or should not be considered a bias (i.e., honest vs. deceptive IM).</p>

continued

TABLE 6 (continued)

Major findings	Major recommendations for future research
<p>(2) Interview question type affects the type of the IM tactics used by applicants. Situational questions lead to more other-focused IM tactics, whereas past-behavior questions lead to more self-promotion and defensive IM.</p> <p>(3) The effect of applicant IM on interview ratings is reduced through the use of structured interviews.</p>	<p>(2) Future research should explore who should rate IM behaviors (applicant and/or interviewer).</p> <p>(3) Research should examine how the applicants' IM behaviors affect interviewers' ratings.</p> <p>(4) More research is needed to understand how different components of structure impact IM use and impact on interview ratings.</p>
<p>Measuring personality via structured interviews</p>	<p>(1) There is relatively little saturation of structured interview ratings by personality factors unless the interview is specifically designed to measure personality.</p> <p>(2) Research could explore factors such as interview structure and the visibility of personality traits that may enhance or impede interviewer judgments of applicant personality.</p> <p>(3) Research could also explore differential interviewer ability to measure applicant personality.</p>
<p>Comparing situational versus past-behavioral questions</p>	<p>(1) Researchers should stop examining the relative superiority of situational questions and past-behavioral questions. Instead, research should focus on how these two question types can best complement each other when used together in structured interviews.</p>

continued

TABLE 6 (continued)

Major findings	Major recommendations for future research
(2) Situational and past-behavioral questions both exhibit acceptable levels of criterion-related validity.	(2) The operationalization of interview question types needs further examination. Do dilemmas in situational questions actually serve the assumed purpose of minimizing transparency? Are dilemmas necessary for situational questions? Would dilemmas add value to past-behavioral questions? (3) Outside of situational and past-behavioral questions, research should examine what other types of questions can add validity to the structured interviews.
Probing, follow-up, prompting, and elaboration on questions (1) The effects of probing in structured interviews are still unclear. The debate as to whether probing increases answer accuracy (and, therefore, more accurate ratings) or introduces contamination and bias has not been sufficiently settled.	(1) A better conceptual understanding of probing, including different types of probing (e.g., verifying, quantifying, testing, etc.) is needed. (2) An initial definition and typology of probing has been proposed in the current review, but the information requires further development and testing. (3) Research should examine how and when probing adds and/or detracts from interview outcomes.
Developing anchored rating scales	(1) More research is needed on this important but surprisingly underresearched component of structure.
(1) Anchored rating scales enhance validity and reliability of structured interviews.	(2) Research examining how anchored rating scales are best designed and operationalized is needed, including how many anchors are needed, what types of anchors are best (i.e., examples, descriptions), how anchors should be determined (i.e., past applicant answers, SMEs), and how anchors are used by interviewers.
Reactions to structured interviews	(1) Applicant and interviewer reactions to specific components of structure should be measured to give a more in-depth look at overall reactions to structure.
(1) Applicant and interviewer reactions to structured interviews are not consistent.	(2) Research should examine the potential trade-offs between psychometric improvements and user reactions to specific components of structure and ways to positively influence applicant and interviewer reactions.

Finally, future research should also examine ways to positively influence applicant and interviewer reactions. Previous research has shown that applicant reactions can be improved by advanced knowledge of the interview questions or dimensions to be measured in the interview (i.e., Day & Carroll, 2003). Other research has found that interviewer reactions can be positively influenced by participation in interview training (Chapman & Zweig, 2005; Lievens & De Paepe, 2004) or direct experience conducting structured interviews (Barclay, 2001; Camp et al., 2004). Examining other ways to increase applicant and interviewer reactions, such as explaining the purposes and advantages of the structured interview to applicants or including interviewers in the interview development process, could be beneficial.

Research Question 19: What are the ways to positively influence applicant and interviewer reactions?

Conclusions

Within each section above we have summarized recent findings and discussed recommendations for future research in many promising areas. Although much is known about structured interviews, there are many unanswered questions. Table 6 summarizes some of the major findings and areas needing additional studies identified in our review. Structured employment interviews are an important area of research because they are more valid than unstructured interviews, can improve decision making, and they are widely used in practice. In addition, they are easy to use, the techniques are well known, and they are simple and low-cost to implement. We hope that our review stimulates further research on this important topic.

REFERENCES

References marked with an asterisk indicate studies included in the content analysis.

^aIndicates a pre-1996 article included in the meta-analysis of subgroup differences in structured interview ratings.

^bIndicates a post-1996 article included in the meta-analysis of subgroup differences in structured interview ratings.

^cIndicates an article included in the meta-analysis of effects of different types of impression management on structured interview ratings.

^dIndicates an article included in the meta-analysis of effects of past-behavior and situational questions on impression management usage.

*Allen TD, Fecteau JD, Fecteau CL. (2004). Structured interviewing for OCB: Construct validity, faking, and the effects of question type. *Human Performance*, 17, 1–24.

- Ambady N, Bernieri FJ, Richeson JA. (2000). Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. *Advances in Experimental Social Psychology*, 32, 201–271.
- Anastasi A, Urbina S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall, Inc.
- Anderson N, Salgado JF, Hülsheger UR. (2010). Applicant reactions in selection: Comprehensive meta-analysis into reaction generalization versus situational specificity. *International Journal of Selection and Assessment*, 18, 291–304.
- Arthur W Jr., Villado AJ. (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection research and practice. *Journal of Applied Psychology*, 93, 435–442.
- Arvey RD, Campion JE. (1982). The employment interview: A summary and review of recent research. *PERSONNEL PSYCHOLOGY*, 35, 281–322.
- Bar M, Neta M, Linz H. (2006). Very first impressions. *Emotion*, 6, 269–278.
- *Barclay JM. (2001). Improving selection interviews with structure: Organisations' use of "behavioural" interviews. *Personnel Review*, 30, 81–101.
- Barrick MR, Mount MK. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of Applied Psychology*, 81, 261–272.
- *Barrick MR, Patton GK, Haugland SN. (2000). Accuracy of interviewer judgments of job applicant personality traits. *PERSONNEL PSYCHOLOGY*, 53, 925–951.
- *Barrick MR, Shaffer JA, DeGrassi SW. (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, 1394–1411.
- Barrick MR, Swider BW, Stewart GL. (2010). Initial evaluations in the interview: Relationships with subsequent interviewer evaluations and employment offers. *Journal of Applied Psychology*, 95, 1163–1172.
- ^aBerkley S. (1984). *VII. Validation report corrections officer trainee*. Commonwealth of Pennsylvania, State Civil Service Commission.
- Berry CM, Sackett PR, Landers RN. (2007). Revisiting interview–cognitive ability relationships: Attending to specific range restriction mechanisms in meta-analysis. *PERSONNEL PSYCHOLOGY*, 60, 837–874.
- Binet A. (1911). Nouvelles recherches sur la mesure du niveau intellectuel chez les enfants d'école. *L'Annuaire Psychologique*, 17, 145–201.
- *Blackman MC. (2002). Personality judgment and the utility of the unstructured interview. *Basic and Applied Social Psychology*, 24, 241–250.
- *Blackman MC, Funder DC. (2002). Effective interview practices for accurately assessing counterproductive traits. *International Journal of Selection and Assessment*, 10, 109–116.
- *Bobrow W, Leonards JS. (1997). Development and validation of an assessment center during organizational change. *Journal of Social Behavior and Personality*, 12, 217–236.
- Borkenau P, Liebler A. (1993). Convergence of stranger ratings of personality and intelligence with self-ratings, partner ratings, and measured intelligence. *Journal of Personality and Social Psychology*, 65, 546–553.
- Bozeman DP, Kacmar KM. (1997). A cybernetic model of impression management processes in organizations. *Organizational Behavior and Human Decision Processes*, 69, 9–30.
- ^bBragger JD, Kutcher E, Morgan J, Firth P. (2002). The effects of the structured interview on reducing biases against pregnant job applicants. *Sex Roles*, 46, 215–226.
- *Brannan MJ, Hawkins B. (2007). London calling: Selection as a pre-emptive strategy for cultural control. *Employee Relations*, 29, 178–191.

- ^{†b}Brecher E, Bragger J, Kutcher E. (2006). The structured interview: Reducing biases toward applicants with physical disabilities. *Employee Responsibilities and Rights Journal*, 18, 155–170.
- Brtek MD, Motowidlo SJ. (2002). Effects of procedure and outcome accountability on interview validity. *Journal of Applied Psychology*, 87, 185–191.
- ^{†b}Buckley MR, Jackson KA, Bolino MC, Veres JG III, Field HS. (2007). The influence of relational demography on panel interview ratings: A field experiment. *PERSONNEL PSYCHOLOGY*, 60, 627–646.
- *Burnett JR, Motowidlo SJ. (1998). Relations between different sources of information in the structured selection interview. *PERSONNEL PSYCHOLOGY*, 51, 963–983.
- *Camp R, Schulz E, Vielhaber M, Wagner-Marsh F. (2004). Human resource professionals' perceptions of team interviews. *Journal of Managerial Psychology*, 19, 490–505.
- Campion MA, Campion JE, Hudson JP. (1994). Structured interviewing: A note on incremental validity and alternative question types. *Journal of Applied Psychology*, 79, 998–1002.
- Campion MA, Palmer DK, Campion JE. (1997). A review of structure in the selection interview. *PERSONNEL PSYCHOLOGY*, 50, 655–702.
- [†]Campion MA, Pursell ED, Brown BK. (1988). Structured interviewing: Raising the psychometric properties of the employment interview. *PERSONNEL PSYCHOLOGY*, 41, 25–42.
- *Carson KD, Carson PP, Fontenot G, Burdin JJ Jr. (2005). Structured interview questions for selecting productive, emotionally mature, and helpful employees. *The Health Care Manager*, 24, 209–215.
- *Chapman DS, Rowe PM. (2001). The impact of videoconference technology, interview structure, and interviewer gender on interviewer evaluations in the employment interview: A field experiment. *Journal of Occupational and Organizational Psychology*, 74, 279–298.
- *Chapman DS, Rowe PM. (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, 185–197.
- *Chapman DS, Zweig DI. (2005). Developing a nomological network for interview structure: Antecedents and consequences of the structured selection interview. *PERSONNEL PSYCHOLOGY*, 58, 673–702.
- *Chen Y, Tsai W, Hu C. (2008). The influences of interviewer-related and situational factors on interviewer reactions to high structured job interviews. *International Journal of Human Resource Management*, 19, 1056–1071.
- Cohen J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Connelly JJ, Ones DS. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin*, 6, 1092–1122.
- Conway JM, Jako RA, Goodman DF. (1995). A meta-analysis of interrater and internal consistency reliability of selection interviews. *Journal of Applied Psychology*, 80, 565–579.
- *Conway JM, Penno GM. (1999). Comparing structured interview question types: Construct validity and applicant reactions. *Journal of Business and Psychology*, 13, 485–506.
- Cortina JM, Goldstein NB, Payne SC, Davison HK, Gilliland SW. (2000). The incremental validity of interview scores over and above g and conscientiousness scores. *PERSONNEL PSYCHOLOGY*, 53, 325–351.
- ^{†b}Cunningham J, Macan T. (2007). Effects of applicant pregnancy on hiring decisions and interview ratings. *Sex Roles*, 57, 497–508.

- Daft RL, Lengel R. (1986). Organizational information requirements, media richness, and structural design. *Management Science*, 32, 554–571.
- *Day AL, Carroll SA. (2003). Situational and patterned behavior description interviews: A comparison of their validity, correlates, and perceived fairness. *Human Performance*, 16, 25–47.
- *DeGroot T, Gooty J. (2009). Can nonverbal cues be used to make meaningful personality attributions in employment interviews? *Journal of Business and Psychology*, 24, 179–192.
- *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, 30–39.
- Dipboye RL. (1994). Structured and unstructured selection interviews: Beyond the job-fit model. *Research in Personnel and Human Resources Management*, 12, 79–123.
- Dipboye RL. (1997). Structured selection interviews: Why do they work? Why are they underutilized? In Anderson N, Herriot P (Eds.), *International handbook of selection and assessment* (pp. 455–473). New York: Wiley.
- *Dipboye RL. (2005). The selection/recruitment interview: Core processes and contexts. In Evers A, Anderson N, Voskuijl O (Eds.), *The Blackwell handbook of personnel selection* (pp. 121–142). Malden, MA: Blackwell.
- *Dipboye RL, Wooten K, Halverson SK. (2004). Behavioral and situational interviews. In Thomas JC (Ed.), *Comprehensive handbook of psychological assessment: Industrial and organizational assessment* (Vol. 4, pp. 297–316). Hoboken, NJ: Wiley.
- *Dixon M, Wang S, Calvin J, Dineen B, Tomlinson E. (2002). The panel interview: A review of empirical research and guidelines for practice. *Public Personnel Management*, 31, 397–428.
- *^{cd}Ellis APJ, West BJ, Ryan AM, DeShon RP. (2002). The use of impression management tactics in structured interviews: A function of question type? *Journal of Applied Psychology*, 87, 1200–1208.
- *Engle R. (1998). Multi-step, structured interviews to select sales representatives in Russia and Eastern Europe. *European Management Journal*, 16, 476–484.
- Fletcher TD, Major DA. (2006). The effects of communication modality on performance and self-ratings of teamwork components. *Journal of Computer-Mediated Communication*, 11, 557–576.
- Funder DC. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, 102, 652–670.
- Funder DC. (1999). *Personality judgment: A realistic approach to person perception*. San Diego, CA: Academic Press.
- *^bGibb JL, Taylor PJ. (2003). Past experience versus situational employment: Interview questions in a New Zealand social service agency. *Asia Pacific Journal of Human Resources*, 41, 371–383.
- Gilmore DC, Ferris GR. (1989). The effects of applicant impression management tactics on interviewer judgments. *Journal of Management*, 15, 557–564.
- *Goho J, Blackman A. (2006). The effectiveness of academic admission interviews: An exploratory meta-analysis. *Medical Teacher*, 28, 335–340.
- Green PC, Alter P, Carr AF. (1993). Development of standard anchors for scoring generic past-behavior questions in structured interviews. *International Journal of Selection and Assessment*, 1, 203–212.
- Griffith RL, McDaniel M. (2006). The nature of deception and applicant faking behavior. In Griffith RL, Peterson MH (Eds.), *A closer examination of applicant faking behavior* (pp. 1–19). Greenwich, NC: Information Age.

- *Grove DA. (1981). A behavioral consistency approach to decision making in employment selection. *PERSONNEL PSYCHOLOGY*, 34, 55–64.
- Haaland S, Christiansen ND. (2002). Implications of trait-activation theory for evaluating the construct validity of assessment center ratings. *PERSONNEL PSYCHOLOGY*, 55, 137–163.
- *Hackett RD, Lapierre LM, Gardiner HP. (2004). A review of Canadian human rights cases involving the employment interview. *Canadian Journal of Administrative Sciences*, 21, 215–228.
- *Harel GH, Arditi-Vogel A, Janz T. (2003). Comparing the validity and utility behavior description interview versus assessment center ratings. *Journal of Managerial Psychology*, 18, 84–104.
- Harris MM. (1989). Reconsidering the employment interview: A review of recent literature and suggestions for future research. *PERSONNEL PSYCHOLOGY*, 42, 691–726.
- Hatfield E, Cacioppo JT, Rapson RL. (1994). *Emotional contagion*. New York: Cambridge University Press.
- Higgins CA, Judge TA. (2004). The effect of applicant influence tactics on recruiter perceptions of fit and hiring recommendations: A field study. *Journal of Applied Psychology*, 89, 622–632.
- Hinkin TR, Tracey JB. (1999). An analysis of variance approach to content validation. *Organizational Research Methods*, 2, 175–186.
- Hollwitz J, Pawlowski DR. (1997). The development of a structured ethical integrity interview for pre-employment screening. *The Journal of Business Communication*, 34, 203–219.
- *Huffcutt AI. (2010). From science to practice: Seven principles for conducting employment interviews. *Applied H.R.M. Research*, 12, 121–136.
- Huffcutt AI, Arthur W Jr. (1994). Hunter & Hunter (1984) revisited: Interview validity for entry-level jobs. *Journal of Applied Psychology*, 79, 184–190.
- *Huffcutt AI, Conway JM, Roth PL, Klehe U. (2004). The impact of job complexity and study design on situational and behavior description interview validity. *International Journal of Selection and Assessment*, 12, 262–273.
- Huffcutt AI, Conway JM, Roth PL, Stone NJ. (2001). Identification and meta-analytic assessment of psychological constructs measured in employment interviews. *Journal of Applied Psychology*, 86, 897–913.
- *Huffcutt AI, Culbertson SS. (2010). Interviews. In Zedeck S (Ed.), *APA handbook of industrial and organizational psychology* (pp. 185–203). Washington, DC: American Psychological Association.
- *Huffcutt AI, Roth PL. (1998). Racial group differences in employment interview evaluations. *Journal of Applied Psychology*, 83, 179–189.
- Huffcutt AI, Roth PL, McDaniel MA. (1996). A meta-analytic investigation of cognitive ability in employment interview evaluations: Moderating characteristics and implications for incremental validity. *Journal of Applied Psychology*, 81, 459–473.
- *Huffcutt AI, Weekley JA, Wiesner WH, DeGroot TG, Jones C. (2001). Comparison of situational and behavior description interview questions for higher-level positions. *PERSONNEL PSYCHOLOGY*, 54, 619–644.
- *Huffcutt AI, Woehr DJ. (1999). Further analysis of employment interview validity: A quantitative evaluation of interviewer-related structuring methods. *Journal of Organizational Behavior*, 20, 549–560.
- *Huffcutt AI, Youngcourt SS. (2007). Employment interviews. In Whetzel DL, Wheaton GR (Eds.), *Applied measurement: Industrial psychology in human resource management* (pp. 181–199). London: Psychology Press.

- Hunter JE, Hunter RF. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96*, 72–98.
- Hurtz GM, Donovan JJ. (2000). Personality and job performance: The Big Five revisited. *Journal of Applied Psychology*, *85*, 869–879.
- Ilies R, Wagner DT, Morgeson FP. (2007). Explaining affective linkages in teams: Individual differences in susceptibility to contagion and individualism-collectivism. *Journal of Applied Psychology*, *92*, 1140–1148.
- Janz JT. (1982). Initial comparisons of patterned behavior description interviews versus unstructured interviews. *Journal of Applied Psychology*, *67*, 577–580.
- *Jelf GS. (1999). A narrative review of post-1989 employment interview research. *Journal of Business and Psychology*, *14*, 25–58.
- Judge TA, Martocchio JJ, Thoresen CJ. (1997). Five-factor model of personality and employee absence. *Journal of Applied Psychology*, *82*, 745–755.
- *Kataoka HC, Latham GP, Whyte GP. (1997). The relative resistance of the situational, patterned behavior, and conventional structured interviews to anchoring effects. *Human Performance*, *10*, 47–63.
- *Kennedy RB, Campbell S, Essary M, Harper B, Herring S, Shonesy L, Valcana G, Whitley WW. (2007). Utilizing the interview in the selection process: A continuing study, 2002–2007. *International Journal of the Academic Business World*, *1*, 31–36.
- *Klehe U, König CJ, Richter GM, Kleinmann M, Melchers KG. (2008). Transparency in structured interviews: Consequences for construct and criterion-related validity. *Human Performance*, *21*, 107–137.
- *Klehe U, Latham G. (2005). The predictive and incremental validity of the situational and patterned behavior description interviews for teamplaying behavior. *International Journal of Selection and Assessment*, *13*, 108–115.
- *Klehe U, Latham G. (2006). What would you do – really or ideally? Constructs underlying the behavior description interview and the situational interview in predicting typical versus maximum performance. *Human Performance*, *19*, 357–382.
- *^{cd}Kleinmann M, Klehe U. (2011). Selling oneself: Construct and criterion-related validity of impression management in structured interviews. *Human Performance*, *24*, 29–46.
- Kohn LS, Dipboye RL. (1998). The effects of interview structure on recruiting outcomes. *Journal of Applied Social Psychology*, *28*, 821–843.
- *König CJ, Melchers KG, Kleinmann M, Richter GM, Klehe U. (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*, 283–292.
- *Krajewski HT, Goffin RD, McCarthy JM, Rothstein MG, Johnston N. (2006). Comparing the validity of structured interviews for managerial-level employees: Should we look to the past or focus on the future? *Journal of Occupational and Organizational Psychology*, *79*, 411–432.
- Kristof-Brown AL, Barrick MR, Franke M. (2002). Applicant impression management: Dispositional influences and consequences for recruiter perceptions of fit and similarity. *Journal of Management*, *28*, 27–46.
- *Kutcher EJ, Bragger JD. (2004). Selection interviews of overweight job applicants: Can structure reduce the bias? *Journal of Applied Social Psychology*, *34*, 1993–2022.
- Latham GP. (1989). The reliability, validity, and practicality of the situational interview. In Eder RW, Ferris GR (Eds.), *The employment interview: Theory, research, and practice* (pp. 169–182). Newbury Park, CA: Sage.
- Latham GP, Saari LM, Pursell ED, Campion MA. (1980). The situational interview. *Journal of Applied Psychology*, *65*, 422–427.

- Latham GP, Sue-Chan C. (1999). A meta-analysis of the situational interview: An enumerative review of reasons for its validity. *Canadian Psychology*, 40, 56–67.
- *Levashina J, Campion MA. (2006). A model of faking likelihood in the employment interview. *International Journal of Selection and Assessment*, 14, 299–316.
- *Levashina J, Campion MA. (2007). Measuring faking in the employment interview: Development and validation of an interview faking behavior scale. *Journal of Applied Psychology*, 92, 1638–1656.
- Levashina J, Morgeson FP, Campion MA. (2012). Tell me some more: Exploring how item verifiability and cognitive ability influence responses to biodata questions in a high-stakes selection context. *PERSONNEL PSYCHOLOGY*, 65, 359–383.
- *Lievens F, De Paep A. (2004). An empirical investigation of interviewer-related factors that discourage the use of high structure interviews. *Journal of Organizational Behavior*, 25, 29–46.
- *Lievens F, Harris MM, Van Keer E, Bisqueret C. (2003). Predicting cross-cultural training performance: The validity of personality, cognitive ability, and dimensions measured by an assessment center and a behavior description interview. *Journal of Applied Psychology*, 88, 476–489.
- *^cLievens F, Peeters H. (2008). Interviewers' sensitivity to impression management tactics in structured interviews. *European Journal of Psychological Assessment*, 24, 174–180.
- ^aLin T, Dobbins GH, Farh J. (1992). A field study of race and age similarity effects on interview ratings in conventional and situational interviews. *Journal of Applied Psychology*, 77, 363–371.
- Locke EA, Latham GP. (1990). *A theory of goal setting and task performance*. Englewood Cliffs, NJ: Prentice-Hall.
- Macan T. (2009). The employment interview: A review of current studies and directions for future research. *Human Resource Management Review*, 19, 203–218.
- *MacHatton MT, Van Dyke T, Steiner R. (1997). Selection and retention of managers in the US restaurant sector. *International Journal of Contemporary Hospitality Management*, 9, 155–160.
- ^aMancuso RP. (1990, June). *Civil service interviews in the city of Los Angeles: An examination of relevant data*. Paper presented at the International Personnel Management Association Assessment Council Annual Conference, San Diego, CA.
- Marchese MC, Muchinsky PM. (1993). The validity of the employment interview: A meta-analysis. *International Journal of Selection and Assessment*, 1, 18–26.
- *Maurer SD. (2002). A practitioner-based analysis of interviewer job expertise and scale format as contextual factors in situational interviews. *PERSONNEL PSYCHOLOGY*, 55, 307–327.
- *Maurer SD. (2006). Using situational interviews to assess engineering applicant fit to work group, job, and organizational requirements. *Engineering Management Journal*, 18, 27–35.
- ^bMaurer SD, Lee TW. (2000). Accuracy of the situational interview in rating multiple job candidates. *Journal of Business and Psychology*, 15, 73–96.
- *Maurer T, Solamon J, Troxel D. (1998). Relationship of coaching with performance in situational employment interviews. *Journal of Applied Psychology*, 83, 128–136.
- *Maurer TJ, Solamon JM. (2006). The science and practice of a structured employment interview coaching program. *PERSONNEL PSYCHOLOGY*, 59, 433–456.
- *Maurer TJ, Solamon JM, Lippstreu M. (2008). How does coaching interviewees affect the validity of a structured interview? *Journal of Organizational Behavior*, 29, 355–371.

- Mayfield EC. (1964). The selection interview: A re-evaluation of published research. *PERSONNEL PSYCHOLOGY*, *17*, 239–260.
- *^bMcCarthy J, Van Iddekinge CH, Campion MA. (2010). Are highly structured job interviews resistant to demographic similarity effects? *PERSONNEL PSYCHOLOGY*, *63*, 325–359.
- McDaniel MA, Whetzel DL, Schmidt FL, Maurer SD. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology*, *79*, 599–616.
- *^cMcFarland LA, Ryan AM, Kriska SD. (2003). Impression management use and effectiveness across assessment methods. *Journal of Management*, *29*, 641–661.
- *^bMcFarland LA, Ryan AM, Sacco JM, Kriska SD. (2004). Examination of structured interview ratings across time: The effects of applicant race, rater race, and panel composition. *Journal of Management*, *30*, 435–452.
- McGrath JE, Hollingshead AB. (1993). Putting the group back in group support systems: Some theoretical issues about dynamic processes in groups with technological enhancements. In Jessup LM, Valacich JS (Eds.), *Group support systems: New perspectives* (pp. 78–96). New York: Macmillan.
- *^cMcKay PF, Davis J. (2008). Traditional selection methods as resistance to diversity in organizations. In Thomas K (Ed.), *Diversity resistance in organizations* (pp. 151–174). New York: Taylor Francis Group/ Erlbaum.
- *^cMelchers KG, Lienhardt N, Aarburg MV, 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*, 53–87.
- *^cMiceli NS, Harvey M, Buckley MR. (2001). Potential discrimination in structured employment interviews. *Employee Responsibilities and Rights Journal*, *13*, 15–38.
- *^cMolgaard LK, Lewis RE. (2008). The use of a structured interview guide in veterinary college admissions. *Journal of Veterinary Medical Education*, *35*, 460–465.
- Moore MG, Kearsley G. (1996). *Distance education: A systems view*. Belmont, CA: Wadsworth.
- Morgeson FP, Campion MA, Dipboye RL, Hollenbeck JR, Murphy K, Schmitt N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *PERSONNEL PSYCHOLOGY*, *60*, 683–729.
- *^cMorgeson FP, Reider MH, Campion MA. (2005). Selecting individuals in team settings: The importance of social skills, personality characteristics, and teamwork knowledge. *PERSONNEL PSYCHOLOGY*, *58*, 583–611.
- Moscoco S. (2000). Selection interview: A review of validity evidence, adverse impact and applicant reactions. *International Journal of Selection and Assessment*, *8*, 237–247.
- *^cMoscoco S, Salgado JF. (2001). Psychometric properties of a structured behavioral interview to hire private security personnel. *Journal of Business and Psychology*, *16*, 51–59.
- Motowidlo SJ. (1999). Asking about past behavior versus hypothetical behavior. In Eder RW, Ferris GR (Eds.), *The employment interview handbook* (pp. 179–190). Thousand Oaks, CA: Sage.
- *^aMotowidlo SJ, Carter GW, Dunnette MD, Tippins N, Werner S, Burnett JR, Vaughan MJ. (1992). Studies of the structured behavioral interview. *Journal of Applied Psychology*, *70*, 774–776.
- Oh I, Wang G, Mount MK. (2011). Validity of observer ratings of the five-factor model of personality traits: A meta-analysis. *Journal of Applied Psychology*, *96*, 762–773.

- *Oliphant GC, Hansen K, Oliphant BJ. (2008). A review of a telephone-administered behavior-based interview technique. *Business Communication Quarterly*, 71, 383–386.
- *Palmer JK, Loveland JM. (2008). The influence of group discussion on performance judgments: Rating accuracy, contrast effects, and halo. *The Journal of Psychology*, 142, 117–130.
- *Pawlowski DR, Hollwitz JH. (2000). Work values, cognitive strategies, and applicant reactions in a structured pre-employment interview for ethical integrity. *The Journal of Business Communication*, 37, 58–76.
- *^{cd}Peeters H, Lievens F. (2006). Verbal and nonverbal impression management tactics in behavior description and situational interviews. *International Journal of Selection and Assessment*, 14, 206–222.
- Personnel Decisions International. (1997). *Selecting for success: Selection interview training. Participant's manual. Section 3: Opening and structuring the interview*. Houston, TX: Author.
- Ployhart RE. (2006). Staffing in the 21st century: New challenges and strategic opportunities. *Journal of Management*, 32, 868–897.
- Posthuma RA, Morgeson FP, Campion MA. (2002). Beyond employment interview validity: A comprehensive narrative review of recent research and trends over time. *PERSONNEL PSYCHOLOGY*, 55, 1–81.
- *Prewett AJ. (1994). Racial composition of selection interview panels and effects on candidates' performance: A field study. Doctoral dissertation, Auburn University. *Dissertation Abstracts International*, 55/03-B, 1215.
- *^cProost K, Schreurs B, De Witte K, Derous E. (2010). Ingratiation and self-promotion in the selection interview: The effects of using single tactics or a combination of tactics on interviewer judgments. *Journal of Applied Social Psychology*, 40, 2155–2169.
- *Pulakos ED, Schmitt N. (1995). Experience-based and situational interview questions: studies of validity. *PERSONNEL PSYCHOLOGY*, 48, 289–308.
- *Pulakos ED, Schmitt N, Whitney D, Smith M. (1996). Individual differences in interviewer ratings: The impact of standardization, consensus discussion, and sampling error on the validity of a structured interview. *PERSONNEL PSYCHOLOGY*, 49, 85–102.
- *Reilly NP, Bocketti SP, Maser SA, Wennet CL. (2006). Benchmarks affect perceptions of prior disability in a structured interview. *Journal of Business and Psychology*, 20, 489–500.
- Reilly RR, Chao GT. (1982). Validity and fairness of some alternative selection procedures. *PERSONNEL PSYCHOLOGY*, 35, 1–62.
- *^cRiggio RE, Throckmorton B. (1988). The relative effects of verbal and nonverbal behavior, appearance, and social skills in evaluations made in hiring interviews. *Journal of Applied Social Psychology*, 18, 331–348.
- *Roth PL, Van Iddekinge CH, Huffcutt AI, Eidson CE Jr., Bobko P. (2002). Corrections for range restrictions in structured interview ethnic group differences: The values may be larger than researchers thought. *Journal of Applied Psychology*, 87, 369–376.
- *Roth PL, Van Iddekinge CH, Huffcutt AI, Eidson CE Jr., Schmit MJ. (2005). Personality saturation in structured interviews. *International Journal of Selection and Assessment*, 13, 261–273.
- *Rynes SL, Barber AE, Varma GH. (2000). Research on the employment interview: Usefulness for practice and recommendations for future research. In Cooper CL, Locke EA (Eds.), *Industrial and organizational psychology: Linking theory with practice* (pp. 250–277). Oxford: Blackwell.

- *bSacco JM, Scheu CR, Ryan AM, Schmitt N. (2003). An investigation of race and sex similarity effects in interviews: A multilevel approach to relational demography. *Journal of Applied Psychology*, 88, 852–865.
- *Salgado JF, Moscoso S. (2002). Comprehensive meta-analysis of the construct validity of the employment interview. *European Journal of Work and Organizational Psychology*, 11, 299–324.
- Sanchez ME, Morchio G. (1992). Probing “Don’t know” answers: Effects on survey estimates and variable relationships. *The Public Opinion Quarterly*, 56, 454–474.
- Schlenker BR. (1980). *Impression management*. Monterey, CA: Brooks/Cole.
- *Schmidt FL, Rader M. (1999). Exploring the boundary conditions for interview validity: Meta-analytic validity findings for a new interview type. *PERSONNEL PSYCHOLOGY*, 52, 445–462.
- *Schmidt FL, Zimmerman RD. (2004). A counterintuitive hypothesis about employment interview validity and some supporting evidence. *Journal of Applied Psychology*, 89, 553–561.
- Schmidt WV, Conaway RN. (1999). *Results-oriented interviewing: Principles, practices, and procedures*. Needham Heights, MA: Allyn and Bacon.
- Schmitt N. (1976). Social and situational determinants of interview decisions: Implications for the employment interview. *PERSONNEL PSYCHOLOGY*, 29, 79–101.
- Schmitt N, Kuncie C. (2002). The effects of required elaboration of answers to biodata questions. *PERSONNEL PSYCHOLOGY*, 55, 569–587.
- Schwab DP, Heneman HG. (1969). Relationship between interview structure and interviewer reliability in an employment situation. *Journal of Applied Psychology*, 53, 214–217.
- Sedikides C, Herbst KC, Hardin DP, Dardis GJ. (2002). Accountability as a deterrent to self-enhancement: The search for mechanisms. *Journal of Personality and Social Psychology*, 83, 592–605.
- *Simola SK, Taggar S, Smith GW. (2007). The employment selection interview: Disparity among research-based recommendations, current practices and what matters to human rights tribunals. *Canadian Journal of Administrative Sciences*, 24, 30–44.
- *Smith VW, Ebmeier H. (2009). The development and field testing of a school psychologist employment interview instrument. *Journal of Applied School Psychology*, 25, 342–363.
- *Stevens CK. (2000). Structure interviews to hire the best people. In Locke EA (Ed.), *The Blackwell handbook of principles of organizational behavior* (pp. 29–40). Oxford: Blackwell.
- Stevens CK, Kristof AL. (1995). Making the right impression: A field study of applicant impression management during job interviews. *Journal of Applied Psychology*, 80, 587–606.
- *Sue-Chan C, Latham GP. (2004). The situational interview as a predictor of academic and team performance: A study of the mediating effects of cognitive ability and emotional intelligence. *International Journal of Selection and Assessment*, 12, 312–320.
- *cSwider BW, Barrick MR, Harris TB, Stoverink AC. (2011). Managing and creating an image in the interview: The role of interviewee initial impressions. *Journal of Applied Psychology*, 96, 1275–1288.
- *Taylor PJ, Small B. (2002). Asking applicants what they would do versus what they did do: A meta-analytic comparison of situational and past behavior employment interview questions. *Journal of Occupational and Organizational Psychology*, 75, 277–294.

- *Terpstra DE, Rozell EJ. (1997). Why some potentially effective staffing practices are seldom used. *Public Personnel Management*, 26, 483–495.
- *Townsend RJ, Bacigalupi SC, Blackman MC. (2007). The accuracy of lay integrity assessments in simulated employment interviews. *Journal of Research in Personality*, 41, 540–557.
- *Tross SA, Maurer TJ. (2008). The effect of coaching interviewees on subsequent interview performance in structured experience-based interviews. *Journal of Occupational and Organizational Psychology*, 81, 589–605.
- *^cTsai W, Chen C, Chiu S. (2005). Exploring boundaries of the effects of applicant impression management tactics in job interviews. *Journal of Management*, 31, 108–125.
- Turnley WH, Bolino MC. (2001). Achieving desired images while avoiding undesired images: Exploring the role of self-monitoring in impression management. *Journal of Applied Psychology*, 86, 351–360.
- Ulrich L, Trumbo D. (1965). The selection interview since 1949. *Psychological Bulletin*, 63, 100–116.
- *US Office of Personnel Management. (2008). *Structured interviews: A practical guide*. Retrieved from <http://www.feddesk.com/federalguides.htm>.
- *van der Zee KI, Bakker AB, Bakker P. (2002). Why are structured interviews so rarely used in personnel selection? *Journal of Applied Psychology*, 87, 176–184.
- *^dVan Iddekinge CH, McFarland LA, Raymark PH. (2007). Antecedents of impression management use and effectiveness in a structured interview. *Journal of Management*, 33, 752–773.
- *Van Iddekinge CH, Raymark PH, Eidson CE Jr., Attenweiler WJ. (2004). What do structured selection interviews really measure? The construct validity of behavior description interviews. *Human Performance*, 17, 71–93.
- *Van Iddekinge CH, Raymark PH, Roth PL. (2005). Assessing personality with a structured employment interview: Construct-related validity and susceptibility to response inflation. *Journal of Applied Psychology*, 90, 536–552.
- *Van Iddekinge CH, Raymark PH, Roth PL, Payne HS. (2006). Comparing the psychometric characteristics of ratings of face-to-face and videotaped structured interviews. *International Journal of Selection and Assessment*, 14, 347–359.
- *Van Iddekinge CH, Sager CE, Burnfield JL, Heffner TS. (2006). The variability of criterion-related validity estimates among interviewers and interview panels. *International Journal of Selection and Assessment*, 14, 193–205.
- Wagner R. (1949). The employment interview: A critical summary. *PERSONNEL PSYCHOLOGY*, 2, 17–46.
- Wiesner WH, Cronshaw SF. (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.
- *Williamson LG, Campion JE, Malos SB, Roehling MV, Campion MA. (1997). Employment interview on trial: Linking interview structure with litigation outcomes. *Journal of Applied Psychology*, 82, 900–912.
- Wright OR. (1969). Summary of research on the selection interview since 1964. *PERSONNEL PSYCHOLOGY*, 22, 391–413.
- Wright PM, Lichtenfels PA, Pursell ED. (1989). The structured interview: Additional studies and meta-analysis. *Journal of Occupational Psychology*, 62, 191–199.