HOW INTERVIEWERS' GOALS RELATE TO THE QUESTIONS THEY ASK

Alexis Hirvo

A Thesis

Submitted to the Graduate College of Bowling Green
State University in partial fulfillment of
the requirements for the degree of

MASTER OF ARTS

April 2022

Committee:

Scott Highhouse, Advisor

Margaret Brooks

Jari Willing

ABSTRACT

Scott Highhouse, Advisor

This study examines how likely interviewers are to choose each of three question types (i.e., traditional, behavioral, brainteaser) based on their goal for the interview (i.e., assess competence of the applicant, assess applicant's fit, asses the applicant's tolerance for stress). Participants were randomly assigned to one of three conditions in an online experiment where they imagined they were creating an interview question bank based on an assigned interviewing goal (i.e., assessing competence, assessing fit, assessing stress tolerance). Results were only partially consistent with hypotheses. Traditional questions were chosen by those assigned either a goal for assessing competence or for assessing fit significantly more frequently than those assigned a goal for assessing stress tolerance. Brainteaser questions were chosen by those assigned the goal for assessing stress tolerance significantly more frequently than those assigned a goal for assessing competence or assessing fit.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Dr. Scott Highhouse, for his guidance, feedback, and support throughout this process. I also want to acknowledge my committee members, Dr. Margaret Brooks and Dr. Jari Willing for their feedback and encouragement. I am exceptionally grateful for their numerous contributions not only to my thesis, but to my development as a researcher.

I would also like to acknowledge my amazing parents for their encouragement, love, and constant support throughout this process. And to my brother who always has a listening ear, encouraging word, and goofy video ready to keep me motivated. I would be remiss if I did not also recognize my grandpa, aunts, uncles, and cousins for their constant support and interest in my endeavors. I could not have achieved this without my family's unwavering confidence in me.

And to my amazing friends who have helped me throughout this process, I am exceedingly grateful. Your encouragement, support, movie nights, and countless hours of conversation gave me the strength to keep going. So, to Melissa, Kate, Marie, Karli, and Mary Stuart, thank you. The journey of graduate school would not be the same without you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Interviews have been studied in industrial-organizational (IO) psychology for over a century (Buckley, Norris, & Wiese, 2000; Posthuma, Morgeson, & Campion, 2002; Zickar & Gibby, 2007). As a staple of selection, interviewing is a cost-effective component of the selection process that is familiar to both applicants and interviewers (Dipboye, Macan, & Shahani-Denning, 2012; Gatewood, Feild, & Barrick, 2016). There are two categories of interviews primarily examined in the literature: structured and unstructured. While the structure of an interview can range on a spectrum from entirely unstructured to entirely structured with numerous levels of semi-structured lying between the two, research has largely focused on the two extremes (Campion, Palmer, & Campion, 1997). Generally, unstructured interviews lack consistency and are easily biased by extraneous factors (e.g., appearance, inappropriate questions) that impact an interviewer's evaluation (Dipboye, 2004). Alternatively, structured interviews are more consistent across applicants (e.g., asking the same questions, rating answers on a clearly defined scale) and limit irrelevant influences (Schmidt & Hunter, 1998). In general, considerable research suggests that unstructured interviews lack reliability and validity in selection (Conway, Jako, & Goodman, 1995; Huffcutt & Arthur, 1994; Schmitt, 1976; Ulrich & Trumbo, 1965; Wagner, 1949). Nevertheless, unstructured interviews remain widely utilized (König, Jöri, & Knüsel, 2011; Lievens & De Paepe, 2004; Lodato, Highhouse, & Brooks, 2011; Orpen, 1985).

Despite being extensively studied and used as a selection tool, interviews vary drastically in format, formality, and question type (Rynes, Colbert, & Brown, 2002). Many have called for the standardization of interviews (e.g., Chapman & Zweig, 2005), in hopes of maximizing reliability and job relatedness (Conway et al., 1995). When interviewers are confronted with

these findings, however, they often retain less reliable practices (e.g., unstructured interviews) believing that unstructured interviews are more effective at assessing "fit" (e.g., Klehe, 2004). With this in mind, we recognize that interviewers may not adhere to the ideal standards of interviews established in the literature and may instead be interested in choosing questions and interview styles that align with their interview goals. The interviewer's goal for the interview may receive higher priority than job-relatedness, for instance, when selecting interview questions. Therefore, this study examines if the types of questions interviewers prefer (behavioral, traditional, or brain-teaser) are related to their goals for the interview (assess competence, fit, or tolerance for stress).

Interviews

Unstructured interviews, though flawed, engender better applicant and interviewer reactions than structured ones (Hausknecht, Day, & Thomas, 2004). Rapport building is a major concern for both interviewers and applicants (Barrick et al., 2012; Chapman & Zweig, 2005). Rapport building makes the interview feel more casual (Barrick, Swider, & Stewart, 2010), putting both the applicant and interviewer at ease (Chapman & Zweig, 2005; Farago, Zide, & Shahani-Denning, 2013).

Interviewers frequently believe that they can trust their intuition in decisions (Highhouse, 2008; McConnell, 1999; Miles & Sadler-Smith, 2014) because unstructured interviews give them a better "feel" for personality and fit (Barrick et al., 2012). This assumption leads many interviewers to utilize faulty approaches in their selection process even when informed of the potential threats to validity (Barrick et al., 2012). Even trained interviewers who structure other aspects of their interview will ignore recommendations to reduce rapport building (Chapman & Zweig, 2005).

Structured interviews, on the other hand, are more reliable and predictive of applicant success (Campion, Campion, & Hudson, 1994; Huffcutt & Youngcourt, 2007). Interviewers are given (or create) a specified list of questions to ask every applicant in a set order with an anchored rating scale for consistent scoring (Campion, Pursell, & Brown, 1988). Follow up questions or prods are not allowed (Lievens & De Paepe, 2004) and interviewers are expected to evaluate each applicant in an "unbiased" way (though bias is never fully eliminated in human judgment; Bar, Neta, & Linz, 2006; Barrick et al., 2010). Interviewers are generally hesitant to utilize completely structured interviews because of the apparent rigidity (Dipboye, 1997) and possible impact on applicant reactions (Harris, 2000). To resolve these concerns, researchers push for a way to combine unstructured with structured interviews (Fylan, 2005).

Combining aspects of structured with unstructured interviews is known as *semi-structured* interviewing (Fylan, 2005). Generally, some structure is better than none, so implementing a semi-structured framework is an improvement over unstructured interviews (Kallio, Pietilä, Johnson, & Kangasniemi, 2016). There are also ways to train interviewers on the "unstructured" aspects to increase reliability (Swider, Barrick, & Harris, 2016). Training on structured rapport building (Gatewood et al., 2016), initial impression awareness (Swider et al., 2016) and asking relevant follow up questions (Chapman & Zweig, 2005) all increase the reliability and validity of the interview.

Types of Interview Questions

Highhouse, Nye, and Zhang (2019) distinguished three categories of questions used in interviews: traditional, behavioral, and brainteaser. Traditional interview questions are unstructured, unstandardized, and provide minimal job-related information about the applicant (e.g., "What do you look for in a job?"; Highhouse et al., 2019; Huffcutt & Arthur, 1994).

Alternatively, behavioral questions are standardized, based on job analyses, and relevant to the job (e.g., “Did you ever postpone making a decision? Why?”; Highhouse et al., 2019; Janz, 1982; Latham, Saari, Pursell, & Campion, 1980). Over time, a new category emerged, now labeled *brain teaser* questions. Brain teaser questions distress job applicants because they are logical in nature, appear inappropriate in the interview setting, and lack a correct answer (e.g., “Estimate how many windows are in New York.”; Poundstone, 2003).

Traditional Interview Questions

Traditional questions, also known as unstructured questions, are the most frequently used type of question asked in an interview (Campion et al., 1988). These questions rarely assess job-related constructs, are not based on a valid job analysis, and are not reliable (Barrick, Shaffer, & DeGrassi, 2009). However, traditional questions are usually preferred because interviewers overestimate the accuracy of their intuition (Highhouse, 2008). One of the biggest issues with traditional questions is how difficult it is to consistently score them because there is no consistent expected response (e.g., “Tell us why you want to work for us.”). With no clear scoring, interviewers rely on their initial impressions, gut feelings, and assumptions about the applicant (Gatewood et al., 2016). The easiest improvement scholars have recommended for over 70 years is to enforce structure through a variety of means including dimensional ratings (Landy, 1976; Yonge, 1956), note taking (Barrett, Svetlik, & Prien, 1967), behaviorally-anchored-rating-scales (Zedeck, Tziner, & Middlestadt, 1983), patterned question structure (McMurry, 1947), and situational interviewing (Latham et al., 1980). These researchers also recommend using a more job-related and consistent question style known as *behavioral* interview questions.

Behavioral Interview Questions

Behavioral interview questions have been studied at length as a more predictive alternative to unstructured interviews for over 70 years (e.g., Hakel, 1989; Van Iddekinge, Raymark, Roth, & Payne, 2006; Wagner, 1949). Two of the initial behaviorally-based models for interviews were situational interviews (Latham et al., 1980) and patterned behavior description interview (PBDI; Janz, 1982). Both of these models contain job-related questions and behaviorally-anchored rating scales (Latham et al., 1980; Janz, 1982). Situational interviews are based on the premise that intentions are the best predictor of future behaviors (Latham et al., 1980; Latham & Sue-Chan, 1996) whereas PBDI is based on the premise that the best predictor of future behavior is past behavior (Latham & Sue-Chan, 1996; Janz, 1982). Because behavioral description interviews are based on critical incidents at the current job, dimensions can be identified as either describing maximum or typical performance and altered appropriately when evaluating the past performance of the applicant (Janz, Hellervik, & Gilmore, 1986). Most behavioral interviews thus focus on the typical performance facets (e.g., organizational skills, punctuality, work ethic) and omit maximum performance dimensions (e.g., technical skills and knowledge; Gatewood et al., 2016).

Behavioral questions can be further broken down into two categories: Past-oriented and situational. *Past-oriented* behavioral questions are generally designed to elicit stories from applicants about previous actions or experiences (e.g., “Tell me about a time when you failed”; Bangerter, Corvalan, & Cavin, 2014). These questions can assess either maximum performance (e.g., technical skills, job knowledge, handling extreme situations) or typical performance (e.g., organization, working with others, distractibility; Gatewood et al., 2016). *Situational* (or hypothetical) behavioral questions are generally designed to assess problem solving abilities or

ideal performance (e.g., “Imagine you were faced with the following situation… what would you do?”; Ellis, West, Ryan, & DeShon, 2002). These questions are largely future oriented (Taylor & Small, 2002) and elicit responses dictated by impression management strategies (Ellis et al., 2002). Both past-oriented and situational behavioral questions are scored separately for each question and dimension throughout the interview (Gatewood et al., 2016).

Brain Teaser Interview Questions

Brain teaser questions were initially popularized by Microsoft in the 1990s (Wright, Sablynski, Manson, & Oshiro, 2012) as an unexpected and challenging question asked in the midst of a job interview (Highhouse et al., 2019; Honer, Wright, & Sablynski, 2006; Wright et al., 2012). These “odd-ball” questions (e.g., “How many windows are in New York City?”) are known for distressing job applicants because they are unexpected and require both job-irrelevant logic and creative thinking during an already stressful interview. Generally, brain teaser questions are criticized for lacking validity and serving no real purpose in selection (Mitchell et al., 2019), although no systematic attempts have been made to establish validity (Highhouse et al., 2019). Even though these questions lack a clear scoring key and are generally unreliable (Honer et al., 2006), they are assumed to measure critical thinking, creativity, intuition, flexibility, and ability to reason (Bock, 2015; Munk & Oliver, 1997). It is unknown, however, how many interviewers use brain teasers or believe they provide important information in their selection decisions (Wright et al., 2012).

Brain teasers can be further broken down into three sub-categories: Nonsense, problem-solving, and puzzle-like (Honer et al., 2006; Poundstone, 2012). *Nonsense* brainteasers are thought to assess quick creativity and flexibility (e.g., “If Hollywood made a movie about your life, who would you like to see play the lead role as you?”; Highhouse et al., 2019). They are

difficult to score and commonly used as a proxy for personality (Kluger & Rothstein, 1993). *Problem-Solving* brainteasers are thought to assess intuition and ability to reason (e.g., "Calculate the angle of two clock pointers when time is 11:50"; Highhouse et al., 2019). Unlike nonsense questions, problem-solving questions have a correct answer that could be solved through a combination of mental models and perspective changes (Krauss & Wang, 2003).

Finally, *Puzzle-Like* brainteasers are thought to assess ability to reason, critical thinking, and creativity (e.g., "How would you determine the weight of a commercial airplane without a scale?"; Highhouse et al., 2019). These questions are significantly correlated with cognitive ability and applicant reaction measures of procedural justice and perception of performance (Honer et al., 2006). These items are harder to score than problem-solving questions because there is no "correct" answer, but are easier to score than nonsense questions because there is a general expected type of response that can be evaluated.

With these different types of questions in mind, it is apparent that traditional, behavioral, and brainteaser questions are each geared to evoke different types of responses. Because all three are frequently used in a variety of interview settings, it is reasonable to expect that the interviewer's selection goals would influence the types of questions they prefer to ask in the interview.

Goals for an Interview

Just as there are different types of interviews (i.e., recruitment, screening, selection), interviewers have different goals for their interviews (Dipboye et al., 2012). This can be based on a variety of factors including the values of the organization (Cable & Judge, 1997) or personal preferences (Ramsay, Gallois, & Callan, 1997). Although many goals for an interview have been studied, three primary goals consistently arise: assessing competence, assessing fit, and assessing

tolerance for stress (Sutton & Watson, 2013; Swider, Zimmerman, & Barrick, 2015; Wright et al., 2012). With the existence of multiple goals, it is reasonable to expect that there are individual differences in the relative importance of interview goals, and that these differences relate with preferences for different types of interview questions.

Assess Competence

One obvious goal for interviewers is to assess applicant job competence (Carroll & McCrackin, 1998). Assessing competence is an important aspect of selection because it encompasses job-related knowledge, skills, and experiences (Sutton & Watson, 2013). Job-related competence allows interviewers to identify who already possesses the abilities required for the position (Huffcutt, 2011) and what training may be required if they are selected (Carretta, 2011). If competence is the primary goal in an interview, the interviewer needs to ask questions that are capable of assessing past performance and abilities. Thus, I predict a competency goal will relate to a preference for behavioral questions (i.e., past-oriented and situational). This is because behavioral questions are focused on job-related experiences, skills, and abilities. If an interviewer is interested in assessing an applicant's job competence, I predict they will prefer behavioral interview questions that provide more direct insight into job-related competence.

> *Hypothesis 1*: Interviewers who assign more importance to assessing applicant competence will indicate a higher likelihood of using behavioral questions than those who assign less importance to assessing applicant competence.

Assess Fit

Another common goal for an interview is to assess "fit" of the applicant with the job or the organization (Haptonstahl, 1998). Fit can be assessed in an interview formally through

interview questions on personality or preferences, and informally by allowing more time for applicant questions at the end of the interview (Swider et al., 2015). Literature has focused on different perceptions of fit, with most research focusing on perceptions of fit from the applicant's perspective (e.g., Kristof, 1996) or the interviewer's perspective (e.g., Rynes & Gerhart, 1990).

The term "fit" is a fuzzy concept that is applied to a variety of theories and has been criticized for its unclear definitions and numerous subtypes (e.g., person-environment, person-job, person-occupation; Schneider, 2001). Harrison (2007) argued that this multi-definitional use of fit does not provide much value to the literature. However, fit can still be useful when defined as a complex, person-environment construct rather than the basis for larger overarching theories (Harrison, 2007). With this perspective, fit can be seen as the compatibility of all the applicant's attributes with the job's environment, made up of unique values and attributes itself (Harrison, 2007). Although fit has many uses and definitions, this study focuses on an interviewer's intuitive sense of the compatibility between the applicant and the overall job environment.

Although fit is commonly considered an important factor for applicants when deciding to accept a job (Elfenbein & O'Reilly, 2007), assessing fit in an interview as a primary goal is not entirely reliable because it is highly dependent on personality and other influential factors (Werbel & Johnson, 2001). Rynes and Gerhart (1990) found subjective qualifications like interpersonal skills, goal orientation, and physical attractiveness contribute to interviewer's assessment of fit whereas objective qualifications like GPA, years of experience, and extra-curricular offices did not contribute to perceptions of fit.

When interviewing an applicant, the interviewer is taking in a lot of information, so fit can be used as a heuristic to achieve faster evaluations of the applicant (Vleugels, De Cooman, Verbruggen, & Solinger, 2018). These assumptions can be biased by a variety of aspects about

the applicant that may or may not be job-relevant. For instance, if an applicant's temperament reminds the interviewer of an old colleague, the interviewer may prescribe characteristics of that colleague to the applicant because they appear to act similarly. Depending on the similarly traits of the colleague, the association could be either beneficial (e.g., they are a team player) or harmful (e.g., they are unproductive) to the applicant. Therefore, the interviewer could claim the applicant is a bad fit for the organization simply because they are similar to an old colleague who was a bad fit. Because these assessments of fit are unreliable, they are likely aligned with, and utilized as a goal in, traditional (unstructured) interviews.

> *Hypothesis 2*: Interviewers who assign more importance to assessing applicant fit will indicate a higher likelihood of using traditional questions than those who assign less importance to assessing applicant fit.

Assess Tolerance for Stress

A third potential goal for an interview is to assess the applicant's tolerance for stress. This type of interview is typically utilized for high-pressure jobs (Wright et al., 2012) with the intention of throwing applicants off from what they would expect in an interview (Honer et al., 2006). This allows the interviewer to see how applicants handle stress, uncertainty, and "think on their feet" (Bock, 2015). Historically, stress interviews were used to assess police officers or military personnel to examine how they behave under intense stress while bounded by time (Freeman, Manson, Katzoff, & Pathman, 1942). Officers who handled the stressful interview well possessed qualities of "bold intelligent action in the face of confusion and stress" (Freeman et al., 1942, p. 427). By handling a stress interview well, the applicants proved they had the ability to remain calm and find a productive solution to the problem, a job relevant skill. Stress interviews are criticized, however, as an excuse for interviewers to wield power over applicants

in a callous way (Highhouse et al., 2019). Stress interviews are only successful if the applicant is highly motivated to receive a job offer (Freeman et al., 1942), making it seem even more cruel to take advantage of the power differential in an interview.

Although stress interviews have drawbacks, they have expanded in frequency because interviewers today want to ensure the applicant does not act out when under pressure (Chen, Lee, Huang, & Ko, 2019). So, these interviews could be used to demonstrate the applicant's ability to regulate emotions under a stressful situation (Chen et al., 2019). Employees' ability to stay calm under stress is increasingly important in a variety of organizations, so it is more frequently being assessed as part of the interview process (Chen et al., 2019; Posthuma et al., 2002). Similarly, brainteaser questions focus on throwing off the applicant to examine his or her ability to tolerate stress in an already stressful situation (Highhouse et al., 2019). Both stress interviews and brainteaser questions are seen negatively from the applicant's perspective as a cruel use of power that signifies the interviewer's indifference toward the applicant (Highhouse et al., 2019; Wright et al., 2012). Because brainteasers and tolerance for stress both focus on throwing off the applicant, I expect interviewers with the goal to assess tolerance for stress to most frequently use brainteasers.

> *Hypothesis 3:* Interviewers who assign more importance to assessing applicant stress tolerance will indicate a higher likelihood of using brainteaser questions than those who assign less importance to assessing applicant stress.

The purpose of this study is to examine the broader mechanisms that influence interviewer decision making when choosing interview questions. Based on previous research, I expect to find that an interviewer's primary goal for the interview will be associated with the questions they choose to ask. To assess this, I conducted a study on how lay people choose

questions to ask in a hypothetical interview setting to see if there are any effects when assigning interview goals.

The present study is an experimental manipulation of interview goals using a sample of lay people. In this study, I provided the participant with a scenario in which they were an interviewer preparing for an interview. They were given a goal for their hypothetical interview (i.e., assess fit, assess tolerance for stress, or assess competence) and asked to indicate their likelihood of asking different questions in that hypothetical interview.

METHOD

Participants

Data were collected using Amazon's Mechanical Turk (www.mturk.com) through Cloud Research (www.cloudresearch.com). MTurk is an online marketplace where people can receive small amounts of compensation for completing short tasks. MTurk participants are significantly more diverse than the average sample of American college students, and they tend to be slightly more diverse than the typical internet sample (Buhrmester, Kwang, & Gosling, 2011; Moss & Litman, 2020). Cloud Research (formerly known as Turk Prime) is a toolkit used alongside MTurk to improve the quality of data through extensive participant panels, study batching to reduce cost, and automatically excluding low-quality participants (Edelman, Rosenzweig, & Moss, 2020). The study was advertised to potential participants as a general examination of interview question preferences. Those who volunteered to participate were then instructed to use their assigned goal to guide their responses to the survey. Only those who were over 18 years of age and living and working in the United States were allowed to participate in the study. Participants were compensated $0.75 for conscientiously completing the survey.

Data were collected for 452 adults, 53 of whom were removed due to a failed manipulation check, leaving a final sample of 399 participants. Of the 53 who failed the manipulation check, 13 (25%) were from the "Fit" condition, 26 (49%) from the "Competence" condition, and 14 (26%) from the "Stress" condition. The majority (62%) of failed manipulation checks incorrectly selected "fit" as their intended goal, followed by "competence" as their incorrect intended goal (36%), with only one person (2%) incorrectly choosing stress tolerance as their intended goal. Sixty percent of participants removed for failing the manipulation check finished the survey faster than average of 6 minutes (*SD* = 4 minutes; 35% finished in under 3

minutes, 25% finished in 3-5.99 minutes) while the remaining 40% of removed participants finished the survey slower than average (18% finished in 6-8.99 minutes, 22% finished in 9+ minutes). Those removed from the sample did not significantly differ from the retained sample in any demographic variables.

The remaining respondents were primarily white (76%), but gender was evenly represented (50% female) and ages ranged widely (M = 41 years of age, SD = 12 years). These descriptives are typical of MTurk samples (Moss & Litman, 2020). Participants had an average tenure of 7 years at their current organization. Interviewing experience varied across participants such that 39% had never conducted an interview, 36% had conducted a handful of job interviews, 20% had conducted quite a few job interviews, and 5% had conducted an enormous amount of job interviews over their career. Organization size also varied across participants with 31% from organizations with 1-19 employees, 18% from organizations with 20-99 employees, 22% from organizations with 100-499 employees, 10% from organizations with 500-999 employees, 5% from organizations with 1,000-2,499 employees, 4% from organizations with 2,500-5,499 employees, and 12% from organizations with 5,500+ employees. Participants across conditions did not significantly differ on any demographics.

Design and Procedure

Pilot Study

A brief pilot study was conducted on the interview questions to see if they could be sorted appropriately into their item question category. The questions came from Highhouse et al. (2019), which list publicly-available interview questions (e.g., glassdoor.com) used for hiring at organizations including brainteaser questions used at organizations like Dell, Microsoft, and Bank of America. Highhouse and colleagues observed a high degree of internal consistency

among the brainteaser questions as a group (.91), despite the wide differences in content (e.g., choosing to be an animal on the carousel versus picking two celebrities to be one's parents). Of the 49 items from Highhouse et al. (2019), 45 were initially selected for this study. See Appendix A for the list of questions.

Five graduate student subject matter experts (SMEs) sorted each interview question into the category they believed the item most closely aligned with (i.e., behavioral, traditional, or brainteaser). If they thought an item could be considered in multiple categories, they were instructed to sort it into the category they thought it most closely resembled. Definitions and examples for each category were provided so that all SMEs had the same frame of reference.

Once all SMEs sorted the 45 interview questions, I examined what proportion of SMEs properly sorted each question. Thirty-eight items (84%) were sorted into the proper category with complete agreement from SMEs. The remaining seven items were individually assessed for retention. Any item that was not properly sorted by at least 60% of SMEs was removed. Four items were removed from the final set of interview questions: One brainteaser question ("Have you ever stolen a pen from work?"), two behavioral questions ("If you could start your career over again, what would you do differently?"; "Tell me about a time when you misjudged a person"), and one traditional question ("If you could be anyone else, who would it be?"). This left a total of 41 interview questions to be used in the final survey (See Appendix A).

Main Study

The main study assessed how interviewer goals impact what questions lay people ask in a hypothetical interview. Participants were instructed to imagine that they were responsible for hiring at a Fortune 500 company and told to evaluate a set of interview questions for

management-level applicants (See Appendix B). In this study, I provided them with one of the three goals for the interview (i.e., assess fit, competence, or tolerance for stress). Participants then rated how likely they were to include each question in their pool of interview questions. After responding to individual-difference measures, participants were compensated for their time.

Measures

Interview Questions

To assess how likely participants are to ask different types of questions in the interview process, I presented them with the 14 brainteaser (e.g., "How would you explain what a chair is to an alien?"), 13 behavioral (e.g., "What would you do if you disagree with a coworker?"), and 14 traditional questions (e.g., "Are you a good listener?") remaining after the pilot study. Of the 49 items from Highhouse et al. (2019), 45 were initially selected for this study, with 41 retained after the pilot study (See Table 1).

TABLE 1
Interview Questions

Traditional	Behavioral	Brainteaser
Why should we hire you?	What would you do if you disagree with a coworker?	If you were to get rid of one state in the country, what would it be and why?
Do you consider yourself a leader?	What do you do when your schedule is interrupted? Give an example of how you handled it.	If Hollywood made a movie about your life, who would you like to see play the lead role as you?
What do you look for in a job?	Describe a situation when you took a risk professionally. What was the outcome?	Calculate the angle of two clock pointers when time is 11:50.
Can you work under pressure and deal with deadlines?	Tell me about a creative solution you developed for a challenging situation or problem.	How would you explain what a chair is to an alien?
What do you know about our organization?	What is the most stressful situation you have handled and what was the outcome?	If you could be any animal on a carousel, what would you pick and why?
Are you a good listener?	Describe a decision you made that was a failure. What happened and why?	How would you determine the weight of a commercial airplane without a scale?
What led you to this point in your life?	If you could start your career over again, what would you differently?*	Why are manhole covers round?
How long would it take you to make a meaningful contribution to our firm?	Tell me about a time that you participated in a team. What was your role?	What is your favorite song? Perform it for us now.
If you could be anyone else, who would it be?*	Tell me about a time when you failed.	Estimate how many windows are in New York.

What do you love?	Give an example of a goal you reached and tell me how you achieved it.	Pick two celebrities to be your parents.
How do you get along with older coworkers?	Have you handled a difficult situation? How?	Have you ever stolen a pen from work?*
Do you check voicemail and email while on vacation?	Give an example of an occasion when you used logic to solve a problem.	How many quarters do you need to reach the height of the Empire State Building?
How successful do you think you've been so far?	Did you ever postpone making a decision? Why?	Name three previous Nobel Prize winners.
Tell us why you want to work for us.	Tell me about a time when you were faced with conflicting priorities. How did you determine the top priority?	What do you think about when you are alone in your car?
What salary do you think you deserve?	Tell me about a time when you misjudged a person.*	What songs best describe your work ethic?

* Item was removed after Subject Matter Expert (SME) review because less than 60% of SMEs sorted it into the correct category

Participants were instructed to imagine that they work for a Fortune 500 company and are responsible for creating a pool of general interview questions to ask all management-level applicants that assess qualifications from an assigned goal (i.e., assess fit, competence, tolerance for stress). Specifically, they were told "Imagine that you work for a Fortune 500 company. You are constructing a pool of interview questions to be asked of all supervisory or management-level applicants. To help you create the question pool, your supervisor provides a specific goal for you to focus on. Your primary goal is to select questions that will determine [**whether the person is a good fit for the organization/the applicant's overall job competence/the person's ability to handle stress**]. Consider what you need to ask the applicant to determine if they are a good overall fit. **Please indicate the likelihood that you would select each of the items below to be included in the set of general interview questions. Remember, your goal is to assess [fit/competence/stress tolerance]**". Participants then rated how likely they are to include each question in their pool of interview questions on a scale of 1 (*would never include*) to 5 (*would definitely include*). Responses were averaged across question-type to create the participant's average brainteaser, behavioral, and traditional question score (See Table 2 for correlations and Cronbach's alpha).

Table 2

Correlation coefficients for interview questions

	1.	2.	3.
1. Traditional	(.88)		
2. Behavioral	.52*	(.87)	
3. Brainteaser	.01	.20*	(.90)

Note: Cronbach's alpha listed in the diagonal. * indicates $p < .001$

A manipulation check was included to ensure participants understood their provided interview goal. Specifically, they were asked "Which of the following best describes **your**

provided goal for creating the pool of interview questions?" and could select "Assess applicant fit", "Assess applicant competence", or "Assess applicant stress tolerance". Participants were instructed to select the statement that most closely resembled their provided goal (See Appendix C).

Demographics

At the conclusion of the survey, I collected relevant demographic information including age, race, gender, tenure at current job, interviewing experience, and size of organization (See Appendix C).

RESULTS

The hypotheses and exclusion criteria are preregistered on AsPredicted.org (https://aspredicted.org/Y3Z_YWQ). Before running the ANOVA to test my hypotheses, I ran an exploratory factor analysis on the 41 interview questions to see if the items would sort into their correct category. Because these items were chosen from three question categories (i.e., behavioral, brainteaser, traditional), I forced the analysis to have three factors but still allowed the items to freely associate. All 14 brainteaser questions loaded cleanly on Factor 1 and predicted 15% of the variance. Thirteen traditional questions and 2 behavioral questions ("Tell me about a time that you participated in a team. What was your role?" and "Give an example of a goal you reached and tell me how you achieved it.") loaded onto Factor 2, predicting 13% of the variance. Finally, the remaining 11 behavioral questions and 1 traditional question ("Can you work under pressure and deal with deadlines?") loaded onto Factor 3, predicting 11% of the variance (See Table 3).

Table 3

Exploratory Factor Analysis

		Factor			
		1	2	3	Unique
B1	What would you do if you disagree with a coworker?			.47	.67
B2	What do you do when your schedule is interrupted? Give an example of how you handled it.			.71	.52
B3	Describe a situation when you took a risk professionally. What was the outcome?			.67	.54
B4	Tell me about a creative solution you developed for a challenging situation or problem.			.54	.65
B5	What is the most stressful situation you have handled and what was the outcome?			.70	.54
B6	Describe a decision you made that was a failure. What happened and why?			.60	.61
B8	Tell me about a time that you participated in a team. What was your role?		.51		.62

	ctd.	1	2	3	Unique
B9	Tell me about a time when you failed.			.59	.62
B10	Give an example of a goal you reached and tell me how you achieved it.		.48		.63
B11	Have you handled a difficult situation? How?			.69	.50
B12	Give an example of an occasion when you used logic to solve a problem.			.36	.73
B13	Did you ever postpone making a decision? Why?			.50	.70
B14	Tell me about a time when you were faced with conflicting priorities. How did you determine the top priority?			.71	.60
R1	If you were to get rid of one state in the country, what would it be and why?	.69			.53
R2	If Hollywood made a movie about your life, who would you like to see play the lead role as you?	.62			.58
R3	Calculate the angle of two clock pointers when time is 11:50.	.62			.61
R4	How would you explain what a chair is to an alien?	.59			.65
R5	If you could be any animal on a carousel, what would you pick and why?	.62			.58
R6	How would you determine the weight of a commercial airplane without a scale?	.62			.60
R7	Why are manhole covers round?	.70			.50
R8	What is your favorite song? Perform it for us now.	.64			.60
R9	Estimate how many windows are in New York.	.61			.60
R10	Pick two celebrities to be your parents.	.64			.57
R12	How many quarters do you need to reach the height of the Empire State Building?	.72			.47
R13	Name three previous Nobel Prize winners.	.65			.56
R14	What do you think about when you are alone in your car?	.62			.58
R15	What songs best describe your work ethic?	.55			.64
T1	Why should we hire you?		.65		.58
T2	Do you consider yourself a leader?		.59		.58
T3	What do you look for in a job?		.69		.53
T4	Can you work under pressure and deal with deadlines?			.32	.75
T5	What do you know about our organization?		.73		.50
T6	Are you a good listener?		.68		.52
T7	What led you to this point in your life?		.54		.64
T8	How long would it take you to make a meaningful contribution to our firm?		.42		.66
T10	What do you love?		.37		.71
T11	How do you get along with older coworkers?		.51		.68
T12	Do you check voicemail and email while on vacation?		.32		.75
T13	How successful do you think you've been so far?		.49		.64
T14	Tell us why you want to work for us.		.72		.50
T15	What salary do you think you deserve?		.50		.70

Note: B = Behavioral question, R = Brainteaser question, T = Traditional question

This exploratory factor analysis shows 93% of items naturally loaded onto the correct factor for their item type. As a final check, I entered all 41 items into a Confirmatory Factor Analysis (CFA) and assigned each item to its originally assigned group (behavioral, brainteaser, or traditional). In the CFA model, all items significantly loaded onto their factor at $p < .001$ and the overall model had okay fit CFI = .77, TLI = .76, SRMR = .09, RMSEA = .07, $\chi^2(776) = 2249$, $p < .001$. From both of these factor analyses, I concluded the items functioned as intended within their question type.

Next, I examined the descriptive statistics for the interview questions (See Table 4) in each condition and across the total sample to see if there were any general differences between groups before running the ANOVA. In total, behavioral questions were rated as the most likely to be used in the interview ($M = 3.93$, $SD = .65$), followed by traditional questions ($M = 3.12$, $SD = .76$), with brainteaser questions as the least likely to be used ($M = 1.53$, $SD = .62$; See Figure 1). I also ran paired sample t-tests to see if these differences were significant when examined in pairs (See Table 5). Behavioral questions were significant more likely to be used in an interview than traditional questions, $t(398) = 23.1$, $p < .001$, $d = 1.16$, or brainteaser questions, $t(398) = 53.6$, $p < .001$, $d = 2.69$. Additionally, traditional questions were significantly more likely to be used in an interview than brainteaser questions, $t(398) = 36.1$, $p < .001$, $d = 1.81$. This supports previous findings of the appeal of behavioral and traditional questions over brainteaser questions (Hausknecht et al., 2004; Honer et al., 2006).

Table 4

Preference for interview questions across 3 experimental conditions

	Assess Competence (n = 122)	Assess Fit (n = 142)	Assess Tolerance for Stress (n = 135)	Total (n = 399)

Variable	M	SD	M	SD	M	SD	M	SD
Traditional	3.18	.74	3.37	.61	2.82	.82	3.12	.76
Behavioral	3.98	.69	3.92	.60	3.90	.67	3.93	.65
Brainteaser	1.51	.61	1.42	.54	1.69	.67	1.53	.62

Figure 1

Mean Rating of Question Type by Assigned Goal

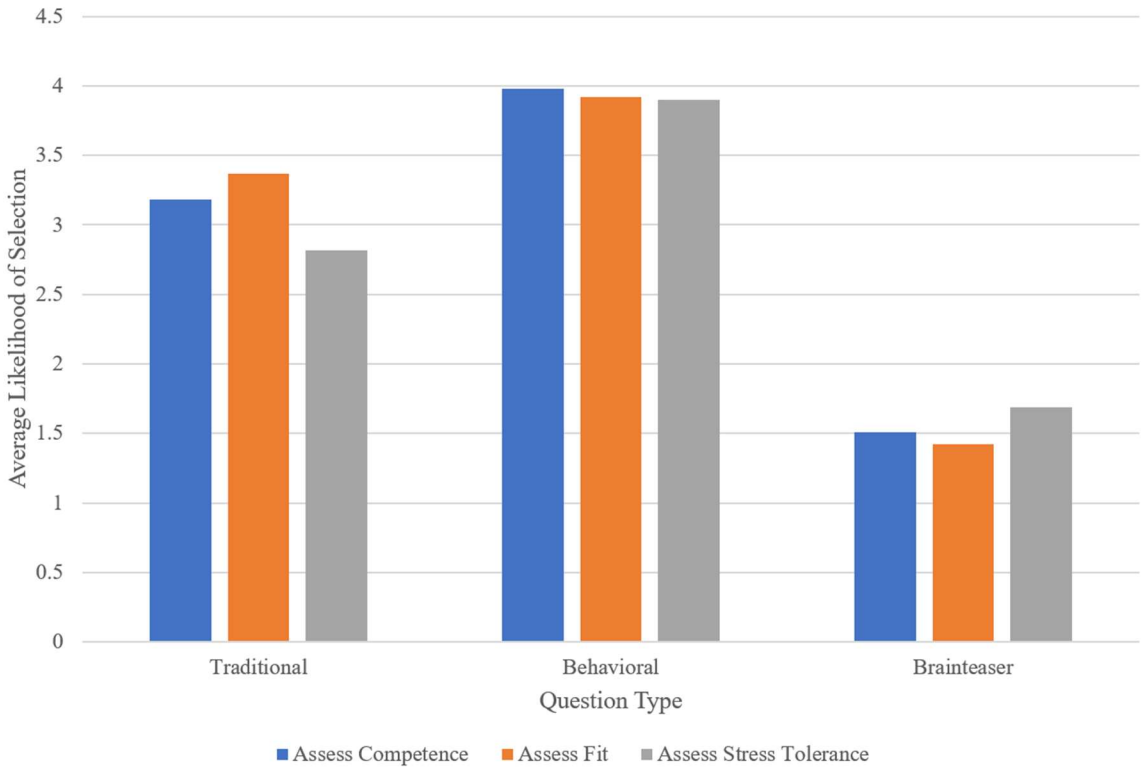

Table 5

Paired-sample t-tests for interview questions across 3 experimental conditions

	t	*df*	*p*	Cohen's d
Behavioral-Traditional	23.1	398	< .001	1.16
Behavioral-Brainteaser	53.6	398	< .001	2.69
Traditional-Brainteaser	36.1	398	< .001	1.81

To test Hypothesis 1-3, I conducted a One-Way ANOVA on the three goals. Hypothesis 1 posited that those with a competency goal would assign a higher likelihood to use behavioral

questions than those presented with alternative goals. The One-Way ANOVA showed that there was an insignificant main effect of goal for behavioral questions, $F(2, 396) = .61$, $p = .54$, $\eta^2 = .003$ (See Table 6), with Tukey's post-hoc tests revealing those assigned the goal for competence ($M = 3.98$, $SD = .69$) were not significantly more likely to use behavioral questions than those assigned the goal for fit ($M = 3.92$, $SD = .60$, $d = -.09$) and those assigned the goal for stress tolerance ($M = 3.90$, $SD = .67$, $d = .14$). Despite finding a slightly higher preference for behavioral questions from those with a goal for competence (See Figure 1), it was not significant; thus, Hypothesis 1 was not supported.

Table 6

ANOVA table of interview questions by assigned interview goal

		Sum of Squares	df	Mean Square	F	Sig.
Traditional	Between Groups	20.95	2	10.47	19.95	< .001
	Within Groups	207.87	396	.53		
	Total	228.82	398			
Behavioral	Between Groups	.52	2	.26	.61	.54
	Within Groups	167.64	396	.42		
	Total	168.15	398			
Brainteaser	Between Groups	5.07	2	2.54	6.90	.001
	Within Groups	145.54	396	.37		
	Total	150.61	398			

Note. $N = 399$.

Hypothesis 2 posited that those with a fit goal would assign a higher likelihood to use traditional questions than those presented with alternative goals. The One-Way ANOVA showed that there was a significant main effect of goal for traditional questions, $F(2, 396) = 19.95$, $p < .001$, $\eta^2 = .092$ (See Table 6), with Tukey's post-hoc tests revealing those assigned a goal for fit ($M = 3.37$, $SD = .61$) were significantly more likely to choose traditional questions than those assigned the goal for stress tolerance ($M = 2.82$, $SD = .82$, $d = .75$; See Figure 1). However, those assigned a fit goal were not significantly more likely to choose traditional questions than those

assigned the goal for competence (M = 3.18, SD = .74, d = .26). Therefore, Hypothesis 2 was partially supported. Consistent with the hypothesis, those assigned a fit goal were significantly more likely to choose traditional questions than those assigned a stress tolerance goal. Inconsistent with the hypothesis, those assigned a fit goal were equally as likely to choose traditional questions as those assigned a competence goal.

Hypothesis 3 posited those with a goal to assess tolerance for stress would assign a higher likelihood to use brainteaser questions. The One-Way ANOVA for brainteaser questions showed that there was a significant main effect of goal, $F(2, 396) = 6.90$, $p < .001$, $\eta^2 = .034$, (See Table 6), with Tukey's post-hoc tests revealing those assigned the goal for stress tolerance (M = 1.69, SD = .67) were significantly more likely to choose brainteaser questions than those assigned the goal for competence (M = 1.51, SD = .61, d = -.28) and fit (M = 1.42, SD = .54, d = -.44). Those with the goal for competence or fit did not significantly differ in their preference for brainteaser questions (See Figure 1). Therefore, Hypothesis 3 was supported.

Exploratory Analyses

Because literature on interviewer goals is still relatively novel, I performed three exploratory regressions on the average behavioral, brainteaser, and traditional scores. The data collected contained demographic information on age, race, gender, tenure at current job, interviewing experience, and size of organization. I used the three HR-related variables (tenure, size of organization, and interviewing experience) in a standardized forward linear regression to see if any job-related variables predicted item preference above and beyond the experimentally assigned goal. I used a forward linear regression specifically because it empirically examines which variable predicts the outcome the most without regard to theory. Given that I had no clear predictions about any of the variables, a forward regression method was most appropriate.

In all three regressions, the interview question type (behavioral, brainteaser, traditional) was the dependent variable, the assigned goal (fit, competence, stress) was entered in Step 1, and the HR-variables (tenure, organization size, and interviewing experience) were entered in Step 2. If any of the HR-variables significantly predicted the interview question type, that variable was retained in the final regression model. Any HR-variable that did not add significance was automatically dropped from the model (See Bendel & Afifi, 1977; Thayer, 2002).

When testing for behavioral interview question preferences, the initial model with only the assigned goal was insignificant, $R^2 = .00$, $F(1,393) = .13$, $p = .72$. After the HR-variables were added in Step 2, Interviewing Experience significantly changed the F-value and explained an incremental 1% of the variance in behavioral interview preference, $\Delta R^2 = .01$, $\Delta F(1,392) = 4.09$, $p < .05$, though the entire model was ultimately still insignificant, $R^2 = .01$, $F(2,392) = 2.11$, $p = .12$ (See Table 7).

Table 7

Exploratory Standardized Forward Regression Coefficients for interview questions

	Behavioral				Brainteaser		Traditional			
	Step 1		Step 2		Step 1		Step 1		Step 2	
Variable	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*
Goal	-.02	.04	-.02	.04	.18	.04	-.29	.04	-.30	.04
Int. Exp.			.10	.04						
Tenure									.10	.01
R^2	.00		.01		.03*		.09*		.10*	
Adjusted R^2	.00		.01		.03		.09		.10	
ΔR^2	.00		.01		.03		.09		.01	

Note. $N = 399$. * indicates value is statistically significant at $p < .001$.

When testing for brainteaser interview question preferences, the initial model with only the assigned goal was significant, $R^2 = .03$, $F(1,393) = 13.22$, $p < .001$. After the HR-variables

were added in Step 2, no variable predicted brainteaser interview question preference above and beyond the assigned goal. Thus, only the goal predicted brainteaser preference (See Table 7).

When testing for traditional interview question preferences, the initial model with only the assigned goal was significant, $R^2 = .09$, $F(1,393) = 37.29$, $p < .001$. After the HR variables were added in Step 2, Tenure significantly changed the F-value and explained an incremental 1% of the variance in traditional interview preference, $\Delta R^2 = .01$, $\Delta F(1,392) = 4.72$, $p < .05$, and the entire model was still significant, $R^2 = .10$, $F(2,392) = 21.18$, $p < .001$ (See Table 7).

DISCUSSION

Very little research has been done on interviewer goals and how they could potentially influence which questions are asked in an interview. Although previous research has identified that interviewers might have different goals prior to conducting the interview, this study connected their goals to the questions they are likely to ask in an interview setting. Given how little research has been done on the mechanism of interviewer goals, this study focused on three goals in particular (i.e., to assess applicant fit, to assess applicant competence, and to assess applicant tolerance for stress) and how they relate to three types of interview questions (i.e., behavioral, traditional, and brainteaser questions).

In the current study, I examined interviewer goals as a predictor of behavioral, brainteaser, and traditional interview question preference. Overall, there was a strong preference for behavioral interview questions independent of assigned interviewer goals. Those who were assigned the goal to assess competence indicated the highest likelihood of choosing behavioral interview questions, as predicted; however, the difference was insignificantly higher than the other two conditions. Therefore, there were slightly different preferences for behavioral questions based on the assigned goal condition with the competence condition having the strongest preference of the three. Although this difference was not significant, it still shows differential preference for the questions in the predicted direction. Future research should investigate this directional trending further to see if it is truly insignificant. The preference for behavioral questions supports previous findings that behavioral questions are seen by interviewees as fair (Gilliland, 1993), frequently used (Simola et al., 2007), and job-relevant (Campion et al., 1997). Behavioral questions tap into job-related abilities or experiences which would explain why people overwhelmingly indicated a preference for these questions above all

else. In the exploratory analyses, people with more interviewing experience were significantly more likely to use behavioral questions, indicating an additional factor into the preference for behavioral questions that should be explored further.

Those assigned the goal to assess fit were significantly more likely to choose traditional questions than those with the stress-tolerance goal, as predicted, and equally as likely as those with the competence goal (not predicted). Overall, those assigned the goal for fit had the highest likelihood of choosing traditional questions of the three conditions, indicating differential preference between the groups in the predicted direction. Much like the behavioral questions, traditional questions are frequently perceived as job-relevant (Hausknecht et al., 2004) and appropriate to ask in an interview setting (Dipboye et al., 2012). This would explain why participants gravitate toward traditional questions in a similar manner they chose behavioral questions. In the exploratory analyses, people with more tenure at their organization were significantly more likely to use traditional questions, indicating an additional factor into the preference for traditional questions.

Finally, when participants were assigned the goal for stress tolerance, they were significantly more likely to choose brainteaser questions than those assigned fit or competence goals, as predicted. This finding again indicates differential question preferences between the three conditions in the predicted direction. Brainteaser questions are frequently perceived as irrelevant to the job (Wright et al., 2012), distressing (Munk & Oliver, 1997), and inappropriate to ask in an interview setting (Mitchell et al., 2019). This would explain why participants in the tolerance for stress condition were most likely to choose the brainteaser questions in this experiment. It is also important to note that exploratory analyses revealed no additional demographic variables predicted brainteaser question use over and above the assigned goal.

Taken together, the findings of the present study provide initial support for the influence of interviewer goals when selecting interview questions. Across all interview questions, the assigned goal predicted different preferences for interview questions for all groups, finding significant results for traditional and brainteaser questions in particular. By using an experimental manipulation, we know the assigned goal was what changed question preference and should be studied further.

Theoretical and Practical Implications

This study supported numerous aspects of the current literature on selection interviews and paves the way for future research to further explore the mechanism of interviewer goals. First of all, the present study may suggest an unspoken understanding people have that traditional questions are not as stressful as behavioral questions. Because traditional questions (e.g., "Can you work under pressure and deal with deadlines?") can sound less formal than behavioral questions (e.g., "Have you handled a difficult situation? How?"), they may appear to be less stressful. If the added structure inherent in behavioral questions sound more formal to the participant, that structure could indicate an added level of stress. Perhaps the underlying stress of behavioral questions explains why applicants experience more negative reactions after behavioral interviews than traditional (Hausknecht et al., 2004). The assigned stress goal in this study reveals an interesting mechanism of inherent stress found in behavioral interview questions that should continue to be studied moving forward. Although stress goals are found less frequently in applied settings than the other goals examined in this study, this experimental goal points to something inherently stressful about behavioral interview questions.

Another important implication of the present research is how similarly goals for fit and goals for competence appear to function. Across all three question types, the groups assigned

goals for fit and competence never significantly differed. Some people could have individually interpreted the fit goal as a competence goal or vice versa since they are both commonly assessed in interviews. This could even be because people can hold multiple goals going into an interview (e.g., mainly want to assess competence but also want to make sure the applicant will fit in on the team) so the goals function similarly. Although people may colloquially use "fit" and "competence" interchangeably, more attempts should be made to conceptually distinguish the goals moving forward. The manipulation used in the current study states the goal plainly without explaining what that goal really means (e.g., "Your goal is to assess the applicant's fit"). Future attempts to study interviewer goals should state what the goals mean more clearly to strengthen the manipulation even more (e.g., "Your goal is to assess the applicant's fit. Fit can include how the applicant would fit into their job, fit with the team, or fit with the organization"). A strengthened manipulation would reveal if the goal for fit and goal for competence actually function similarly or if they were simply misinterpreted in the present study.

If the goals do function similarly as seen in this study, it brings some optimism to the implementation of more structure in interviews. If goals for fit and competence overlap, it could open a new avenue to encourage people to use more structure by appealing to their personal goals for the interview (e.g., if you want someone who fits in with your team, this behavioral question is the best way to assess that).

As an initial study combining interviewer goals and question preference, the present study highlighted differences in how likely people are to ask certain questions in an interview. This can be extended in future studies to see what items specifically people choose to include in a limited item bank. By limiting the number of items someone can include to accomplish their interview goal, it is possible the effect interviewer goal has on question choice could change. In

the present study, we examined how likely people are on average to include interview questions in an item bank with no specified number of items to ultimately choose at the end. If this study were replicated, it would be interesting to see what would happen if participants were presented with the items they would consider including in their item bank and asked to only choose 10 as their final set of questions. This would require participants to choose only the top items they want to include and could encourage them to choose items that closely align with their assigned goal. The present study allows participants to rate as many items positively as they want, so theoretically their final item bank could have 40 items in it. If we require participants to make difficult choices and only choose a limited number of items, we could see some interesting results. This could be an important stepping stone into future research to explore when and why these goals function differently.

Given the experimental nature of this study, there are a few practical implications that can be taken away. First, the main practical implication of the present study is for practitioners to simply be aware that interviewer goals can influence the questions they ultimately ask in an interview. If the organization does not have a designated system for structuring their selection interviews or choosing interview questions, they could see inconsistent interview question choice across interviewers based on their individual goals for the interview. Organizations could improve this by focusing specifically on the goals of the interviewers by requiring them to pre-identify their personal goals and training them on the influence of goals. This could inform interviewers on the impact their goals could have. Additionally, organizations could provide frame of reference training to help interviewers rate responses more consistently and reliably. By providing all of this added structure, organizations would help reduce the impact goals have on

interview question choice and train their interviewers to provide more reliable and consistent interviews.

With this in mind, practitioners concerned about the influence of interviewer goals could consider training hiring mangers on how to standardize question choice across interviewers to limit the influence of individual preference on the interview. Although this training may resemble previous trainings on structured or semi-structured interviews, it would primarily differ in the framing. It would still highlight the importance of using reliable and valid interview questions and enforcing as much structure as possible, but it would do so by explaining the influence of individual goals on question choice. By teaching hiring managers how individual interviewers differ in what questions they personally want to ask based on their goal for the interview, it could encourage managers to create a standard set of questions across interviewers if they haven't already done so. It could also reveal why it is important to enforce more structure from the onset of the interview. If the interviewer's goal influences what they look for in the interview (and thus what questions they ask), practitioners should consider enforcing more structure to keep interviews as consistent as possible.

The final practical recommendation I propose is to frame why people should ask more structured questions differently. Rather than telling people to use behavioral questions because they are more reliable, valid, or predictive of performance, practitioners could try framing question choice based on interviewer goals. For example, imagine someone is hesitant to use behavioral items in their interview because they believe their gut instinct is a good indication of the applicant's fit with the organization. Instead of encouraging them to use the more structured behavioral questions because they are a better predictor, practitioners could encourage them to use a behavioral question that specifically assesses organizational fit. The item would then

appease one aspect of the interviewer's concerns by assessing organizational fit and would allow for more reliable and valid interpretations of fit. While I do not propose this framing as the ultimate way to get people to use more structure in their interviews, I think it is a tactic worth exploring. In understanding the influence of individual goals, practitioners could appeal to that goal specifically when encouraging interviewers to use more structure in their interviews.

Future Research Directions

From this study, four additional research directions logically emerge: (1) how hiring managers' goals influence their interview question choice, (2) how hiring manager individual differences relate to their goals, (3) how organizations can influence interview goals, and (4) how organizational and interviewer goals interact. Regarding the first potential research direction, this study showed through experimental manipulation how interview question choice was influenced by an assigned goal for the interview. However, it would be interesting to expand this further into an applied HR setting to see if actual hiring managers choose interview questions that align with their personal goal for the interview. They could complete the same task to indicate the likelihood they would choose each interview question for a general interview pool, however, they would not be given a specific goal to guide their decisions. Instead, they would use their own personal goal to choose the questions. This new research direction would pair well with the findings of the present study because it shows how the same mechanic functions in an applied HR setting. It could also investigate what happens in applied settings when hiring managers have multiple personal goals (e.g., mainly focused on competence but also interested in fit). Future research should continue to consider how interviewer goals impact the questions they choose in both an experimental and applied setting.

The next research direction logically follows. Once we know what goals currently exist in HR managers, we should examine what individual differences predict interviewers' goal choice. Considering the previous research on goal setting in organizations (e.g., Judge & Ilies, 2002), I would expect differences in personality traits (e.g., Big 5, HEXACO) to predict different interview goal preferences. For example, someone high on extraversion might have the goal to assess fit in an interview because they prefer to just get to know the candidate in the interview setting. Alternatively, someone high in conscientiousness might prefer to assess competence in the interview to increase their likelihood of choosing a successful candidate. Studying individual differences is a logical step as previous research has found people high in dark triad traits perceive brainteaser questions as more appropriate in interview settings (Highhouse et al., 2019). Future research should consider personality, dark triad traits, and other individual differences (e.g., intelligence, risk aversion) when assessing interviewer goals and the questions they ultimately choose to ask.

Future research should also seek to understand how organizations influence interview goals and how organizational goals interact with or influence individual hiring manager goals. Previous research has primarily focused on selection interviews as an interaction between the applicant and the interviewer. In this research stream, the interviewer is considered synonymous with the organization because they serve as a representative for the organization while in a selection interview setting with the applicant. However, in shifting the focus of the research away from the applicant and toward interviewer, a new interesting piece emerges. Organizations as an entity can disseminate company goals for their employees which could influence the goal of the interviewer. For example, if a fast-paced technology company is focused on hiring employees who can handle any situation, they might highlight the need to hire people who can

remain calm under pressure or stress. This could then influence the interviewer to select more questions that will assess how the candidate can tolerate stress. The role of the organization should not be ignored moving forward.

The last future research direction that logically arises from the present study is to examine the interaction between interviewer goals and organizational goals. Because the role of the organization in the interviewer goal setting process has not been studied as widely, it would be interesting to see how organizational goals interact with personal interviewer goals. Take the previous example of the fast-paced technology company focused on hiring employees high in tolerance for stress. In this scenario, how would the interviewer's own personal goals influence their question choice? If their goal aligns with the organization's goal, they might be more inclined to select a large proportion of questions to assess stress tolerance. On the other hand, if their own personal goal is to assess if the candidate is the most competent for the job, they could experience internal conflict when selecting questions. They have to weigh out their own goals compared to the organization's goals and decide which will take precedence in the interview. I think this dynamic would be fascinating to study further to understand what mechanics lie at the center of this balancing act.

Limitations

Although the findings of this study contribute to the academic literature, it is important to note some limitations. First, data was collected using Amazon's Mechanical Turk (MTurk), which has previously received criticism surrounding data quality concerns. Participants recruited through MTurk have been criticized for lacking full demographic representation, having a high degree of careless responding, and lacking motivation to participate meaningfully in the study. There are concerns over who is participating in MTurk and if they provide data that can be

significantly interpreted by researchers. However, research has found that MTurk results are generalizable (Buhrmester, Talaifar, & Gosling, 2018; Landers & Behrend, 2015; Shadish, Cook, & Campbell, 2002), replicable (Berinsky et al., 2012; Mason & Suri, 2012), and valid (Horton et al., 2011). Additionally, MTurk participants tend to be more diverse than the typical internet sample (Buhrmester et al., 2011). In the present study, I used Cloud Research, an Amazon Web Service, and a manipulation check to increase participant quality. While there are some concerns about using MTurk, there is support for conducting experimental manipulations on the platform. Horton et al. (2011) found experiments conducted on MTurk were as internally and externally valid as other kinds of experiments (including laboratory and field experiments). With this in mind, I believe MTurk was an appropriate service for the present experimental study.

Another potential limitation of the study was the incentive given to participants. People would not gain anything in particular by participating in the study outside of becoming more aware of different interview questions. Participants on MTurk were also paid at a fair rate for the platform ($0.75), but that pay could have been too little to motivate participants to respond conscientiously. Beyond the base pay, I could have provided a bonus, additional payment, or potential to win a larger raffle for conscientious responding. By motivating the participants more, I could have seen fewer people fail the manipulation check and may have had better data in general.

Another limitation of this study is the manipulation itself. Although I included a manipulation check to ensure participants understood what their assigned goal was and removed anyone who failed to identify their correct goal, the number of failures points to a potential issue with the manipulation. One way to strengthen the manipulation is to expand upon what the goals mean by providing more than just a statement of the goal. For instance, instead of simply telling

a participant their goal is to "Assess candidate fit", the stronger manipulation could expand on that by clearly defining what fit means. This could be expanded with additional explanation through a sentence (e.g., "Fit can include how the applicant would fit into their job, fit with the team, or fit with the organization.") or could be presented in a chart below the scenario that highlights important elements of that goal (e.g., "Consider how the applicant will fit in their position. Will they fit in with their coworkers? Imagine how they will fit in with the team. Do they fit into the larger organization?"). Alternatively, the manipulation could be altered by expanding it even further to explain why that goal is particularly relevant to the participant. For example, after assigning the goal of fit, an additional paragraph could be added that explains the potential dangers of hiring someone who does not fit in with the organization. This could include anecdotes of people performing poorly and turning over quickly because they could not fit into the company or team culture. If there is interest in future studies to examine different levels of fit, general fit could be split into more goal conditions by dividing it into two separate goals: The goal for company fit and the goal for job fit. Additionally, after assigning the competence goal, the additional paragraph to strengthen the manipulation could include information about the importance of choosing someone who is fully prepared for the position and is most qualified to perform the job. This could include estimates of how much high performing individuals contribute to the organization's overall performance (e.g., productivity) with an emphasis on how much money is saved by hiring the most competent person.

Another check that could strengthen the manipulation is to include an open-ended manipulation check immediately after participants are assigned their goal. This open-ended response could simply ask participants to explain what their assigned condition means to them. Adding this step would improve the manipulation in a few ways. First, it would require

participants think about their condition more deeply by reflecting on the goal more explicitly. Second, it would provide extra clarification for the researcher by showing exactly how participants interpreted their assigned goal. This would serve as a good check for the researcher to ensure the manipulation worked and that participants interpreted their condition in a similar manner.

By strengthening and expanding the manipulation further, it could improve the study in many ways. First, I would expect fewer participants to fail the manipulation check. By clearly stating the goal and expanding on what that goal really means, I would expect fewer people to forget or confuse their assigned goal when responding to the manipulation check. In the present study, it is easy for a participant to misinterpret the goal for fit as a goal for competence or vice versa because there is no clear explanation of what the goal really means which could in turn cause a manipulation check failure. Second, I would expect a stronger manipulation to clarify the results of fit and competence specifically. As previously stated, these two functioned similarly in the data which could be due to the colloquial misinterpretation of what fit and competence mean in an interview setting. People in the present study may have misinterpreted one for the other or could just think they mean the same thing entirely. Without the stronger manipulation check, I cannot know if fit and competence goals actually function similarly or if my participants misinterpreted the meaning of these goals. Therefore, future studies focused on interviewer goals should define the goals more clearly in the manipulation.

Another potential limitation is the nature of the task in the study. As an initial study of interviewer goals, I wanted participants to indicate how likely they are to choose each item individually without regard to the size of their item bank or with any strict rules. This seemed to be an appropriate first step to understand these three goals, but it is possible this task was not

strong enough to differentiate all the goals. If the task itself were changed, we could see different results. For instance, participants could still begin by indicating how likely they are to choose each task as they did in the present study. However, we could add a new task following it that requires participants to definitively decide if they will include that item in their final item bank. That way, we have both their likelihood and a final set of items for each participant. Although I see ways to improve the task used in the present survey, I believe the present task was useful as an initial study on interviewer goals.

The study could be considered limited by its experimental nature. As explained in Highhouse (2009), experiments are frequently criticized for not being generalizable; however, they allow us to control more of the study (Gravetter & Forzano, 2016), infer a level of causality (Bleske-Rechek, Morrison, & Heidtke, 2015), and understand how variables relate to each other (Toomela, 2008). However, experiments are overwhelmingly critiqued for creating unrealistic situations (Bauman, McGraw, Bartels, & Warren, 2014), providing mundane realism (Highhouse, 2009), or failing to provide an explanation for the results (Gravetter & Forzano, 2016). However, in novel areas of research, an experiment is an ideal place to begin because it provides control and clear interpretation of the relations between variables (Highhouse, 2009; Stone-Romero, 2011). Because limited previous research exists that explicitly examines how interviewer goals relate to the interview questions they choose to ask, an experiment was a useful first step in understanding the larger topic at hand and was appropriate in the present study.

The final limitation is the potential lack of generalizability to real-world choices by hiring managers. An online convenience sample was ideal for an experimental scenario to better understand how variables relate, but it does not tell us how these findings transfer to an actual hiring setting. As previously outlined, studying these goals and questions in HR professionals is

an important next step for the research. Although online experiments are generalizable to the larger population (Mullinix, Leeper, Druckman, & Freese, 2016), using a non-experimental approach with real hiring managers could reveal additional information about the underlying mechanics of this study's findings.

Conclusion

Selection interviews have been a staple of IO Psychology for well over a century (Buckley et al., 2000; Posthuma et al., 2002; Zickar & Gibby, 2007) and will continue to be as long as they are widely utilized in the selection process. Throughout that time, researchers have explored how best to encourage hiring managers to use more structured processes and questions in their interviews with mixed results (Conway et al., 1995; Huffcutt & Arthur, 1994; König et al., 2011; Lievens & De Paepe, 2004; Lodato et al., 2011; Orpen, 1985; Schmitt, 1976; Ulrich & Trumbo, 1965; Wagner, 1949). This study approached this long-standing question from a new angle by examining how interviewer goals can ultimately influence the questions they choose to ask in an interview. Results indicate some support for the relation between interviewer goals and the questions they choose to ask, suggesting there is potential for continuation in this area of research. By more thoroughly understanding the role of interviewer goals, this experiment points toward an under-studied mechanism that could explain why hiring managers are hesitant to use fully structured interviews and instead favor semi-structured or unstructured interviews based on their own personal goals for the interview. Future research should continue to study the influence of interviewer goals by addressing the limitations of the current study.

REFERENCES

Bangerter, A., Corvalan, P., Cavin, C. (2014). Storytelling in the selection interview? How applicants respond to past behavior questions. *Journal of Business and Psychology, 29,* 593-604. doi: 10.1007/s10869-014-9350-0

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion, 6*, 269–278. doi: 10.1037/1528-3542.6.2.269

Barrett, G. V., Svetlik, B., Prien, E. P. (1967). Validity of the job-concept interview in an industrial setting. *Journal of Applied Psychology, 51*, 233-235. doi: 10.1037/h0024682

Barrick, M. R., Dustin, S. L., Giluk, T. L., Stewart, G. L., Shaffer, J. A., & Swider, B. W. (2012). Candidate characteristics driving initial impressions during rapport building: Implications for employment interview validity. *Journal of Occupational and Organizational Psychology, 85*, 330-352. doi: 10.1111/j.2044-8325.2011.02036.x

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

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

Bauman, C. W., McGraw, A. P., Bartels, D. M., & Warren, C. (2014). Revisiting external validity: Concerns about trolley problems and other sacrificial dilemmas in moral psychology. *Social and Personality Psychology Compass, 8*(9), 536-554. doi: 10.1111/spc3.12131

Bendel, R. B., & Afifi, A. A. (1977). Comparison of stopping rules in forward "stepwise" regression. *Journal of the American Statistical Association, 72*(357), 46-53.

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis, 20*, 351- 368. doi: 10.1093/pan/mpr057

Berkowitz, L., & Donnerstein, E. (1982). External validity is more than skin deep: Some answers to criticisms of laboratory experiments. *American Psychologist, 37*(3), 245-257. doi: 10.1037/0003-066X.37.3.245

Bleske-Rechek, A., Morrison, K. M., & Heidtke, L. D. (2015). Causal inference from descriptions of experimental and non-experimental research: Public understanding of correlation-versus-causation. *The Journal of General Psychology, 142*(1), 48-70. doi: 10.1080/ 00221309.2014.977216

Bock, L. (2015). *Work Rules!: Insights from inside Google that will transform how you live and lead.* New York, NY: Grand Central Publishing.

Buckley, M. R., Norris, C. A., & Wiese, D. S. (2000). A brief history of the selection interview: May the next 100 years be more fruitful. *Journal of Management History, 6*, 113–126. doi: 10.1108/EUM0000000005329

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data?. *Perspectives on Psychological Science, 6*, 3-5. doi: 10.1037/14805-009

Buhrmester, M. D., Talaifar, S., & Gosling, S. D. (2018). An evaluation of Amazon's Mechanical Turk, its rapid rise, and its effective use. *Perspectives on Psychological Science, 13,* 149-154. doi: 10.1177/1745691617706516

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

Campion, M. A., Campion, J. E., & Hudson, J. P. Jr. (1994). Structured interviewing: A note on incremental validity and alternative question types. *Journal of Applied Psychology, 79*, 998-1002. doi: 10.1037/0021-9010.79.6.998

Campion, M. A., Palmer, D. K., & Campion, J. E. (1997). A review of structure in the selection interview. *Personnel Psychology, 50*(3), 655-702.

Campion, M. A., Pursell, E. D., & Brown, B. K. (1988). Structured interviewing: Raising the psychometric properties of the employment interview. *Personnel Psychology, 41*, 25-42. doi: 10.1111/j.1744-6570.1988.tb00630.x

Carretta, T. R. (2011). Pilot Candidate Selection Method: Still an effective predictor of US Air Force pilot training performance. *Aviation Psychology and Applied Human Factors, 1*, 3–8. doi: 10.1027/2192-0923/a00002

Carroll, A., & McCrackin, J. (1998). The competent use of competency-based strategies for selection and development. *Performance Improvement Quarterly, 11*, 45-63. doi: 10.1111/j.1937-8327.1998.tb00099.x.

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

Chen, C-C., Lee, Y-H., Huang, T-C., Ko, S-F. (2019). Effects of stress interviews on selection/recruitment function of employment interviews. *Asia Pacific Journal of Human Resources, 57*, 40-56. doi: 10.1111/1744-7941.12170

Conway, J. M., Jako, R. A., & Goodman, D. F. (1995). A meta-analysis of interrater and internal consistency reliability of selection interviews. *Journal of Applied Psychology, 80,* 565-579. doi: 10.1037/0021-9010.80.5.565

Dana, J., Dawes, R., & Peterson, N. (2013). Belief in the unstructured interview: The persistence of an illusion. *Judgment and Decision Making, 8*(5), 512-520.

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

Dipboye, R. L. (1997). Structured selection interview publications: Why do they work? Why are they underutilized? In N. Anderson and P. Herriot (Eds.), *International Handbook of Selection and Assessment* (pp. 455-473). Chichester: John Wiley and Sons.

Dipboye, R. L. (2004). The selection/recruitment interview: Core processes and contexts. In A. Evers, N. Anderson, and O. Voskuijil (Eds.), *The Blackwell Handbook of Personnel Selection* (pp. 121-142). Oxford: Blackwell Publishers.

Dipboye, R. L., Macan, T., & Shahani-Denning, C. (2012). The selection interview from the interviewer and applicant perspectives: Can't have one without the other. In N. Schmitt (Ed.), *The Oxford Handbook of Personnel Assessment and Selection* (pp. 323-352). New York: Oxford University Press.

Edelman, J., Rosenzweig, C., & Moss, A. (2020, July 7). *How to conduct better research studies with the MTurk toolkit.* CloudResearch. https://www.cloudresearch.com/resources/research/how-mturk-toolkit-enhances-research-studies/

Elfenbein, H. A., & O'Reilly, C. A. III. (2007). Fitting in: The effects of relational demography and person-culture fit on group process and performance. *Group and Organization Management, 32,* 109-142. doi: 10.1177/1059601106286882

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

Farago, B., Zide, J. S., & Shahani-Denning, C. (2013). Selection interviews: Role of interviewer warmth, interview structure, and interview outcome in applicants' perceptions of organizations. *Consulting Psychology Journal: Practice and Research, 65*, 224-239. doi: 10.1037/a0034300

Freeman, G. L., Manson, G. E., Katzoff, E. T., & Pathman, J. H. (1942). The stress interview. *The Journal of Abnormal and Social Psychology, 37*, 427-447. doi: 10.1037/h0059025

Fylan, F. (2005). Semi-structured interviewing. *A handbook of research methods for clinical and health psychology, 5*, 65-78. doi: 10.1093/med:psych/9780198527565.003.0006

Gatewood, R. D., Feild, H. S., & Barrick, M. R. (2016). *Human Resource Selection* (8th ed.). Boston, MA: Cengage Learning.

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

Gravetter, F. J., & Forzano, L-A. B. (2016). *Research Methods for the Behavioral Sciences* (5th ed.). Stamford, CT: Cengage Learning.

Hakel, M. D. (1989). The state of employment interview theory and research. In R. W Eder & G. R. Ferris (Eds.), *The employment interview: Theory, research, and practice* (pp. 285-306). Newbury Park, CA: Sage Publications.

Haptonstahl, D. E. (1998). *Assessing applicant "fit" within the selection process: Which interviewer perceptions matter?* (Doctoral dissertation). Retrieved from LSU Historical Dissertations and Theses. (https://digitalcommons.lsu.edu/gradschool_disstheses/6626)

Harris, L. (2000). Procedural justice and perceptions of fairness in selection practice. *International Journal of Selection and Assessment, 8*, 48-156. doi: 10.1111/1468-2389.00143

Harrison, D. A. (2007). Pitching fits in applied psychological research: Making fit methods fit theory. In C. L. Ostroff & T. Judge (Eds.), *Perspectives on organizational fit* (pp. 389-416). Psychology Press.

Hausknecht, J. P., Day, D. V., & Thomas, S. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology, 57*, 639-683. doi: 10.1111/j.1744-6570.2004.00003.x

Highhouse, S. (2008). Stubborn reliance on intuition and subjectivity in employee selection. *Industrial and Organizational Psychology, 1*, 333-342. doi: 10.1111/j.1754-9434.2008.00058.x

Highhouse, S. (2009). Designing experiments that generalize. *Organizational Research Methods, 12*(3), 554-566.

Highhouse, S., Nye, C. D., & Zhang, D. C. (2019). Dark motives and elective use of brainteaser questions. *Applied Psychology: An International Review, 68,* 311-340. doi: 10.1111/apps.12163

Honer, J., Wright, C. W., & Sablynski, C. J. (2006). Puzzle interviews: What are they and what do they measure. *Applied HRM Research, 11*, 79–96. doi: 10.1111/j.1559-1816.2012.00963.x

Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics, 14*, 399-425. doi: 10.1007/s10683-011-9273-9

Huffcutt, A. I. (2011). An empirical review of the employment interview construct literature. *International Journal of Selection and Assessment, 19*, 62-81. doi: 10.1111/j.1468-2389.2010.00535.x

Huffcutt, A. I., & Arthur, W. (1994). Hunter and Hunter (1984) revisited: Interview validity for entry-level jobs. *Journal of Applied Psychology, 79*, 184–190. doi: 10.1037/0021-9010.79.2.184

Huffcutt, A. I., & Youngcourt, S. S. (2007). Employment interviews. In D. L. Whetzel and G. R. Wheaton (Eds.), *Applied Measurement: Industrial Psychology in Human Resource Management* (pp. 181-199). London: Psychology Press.

Janz, T. (1982). Initial comparisons of patterned behavior description interviews versus unstructured interviews. *Journal of Applied Psychology, 67,* 577-580. doi: 10.1037/0021-9010.67.5.577

Janz, T., Hellervik, L., & Gilmore, D. C. (1986). *Behavior description interviewing: New, accurate, cost effective*. Boston: Allyn and Bacon.

Judge, T. A., & Ilies, R. (2002). Relationship of personality to performance motivation: A meta-analytic review. *Journal of Applied Psychology, 87*(4), 797-807. doi: 10.1037//0021-9010.87.4.797

Kallio, H., Pietilä, A. M., Johnson, M., & Kangasniemi, M. (2016). Systematic methodological review: Developing a framework for a qualitative semi-structured interview guide. *Journal of Advanced Nursing, 72,* 2954-2964. doi: 10.1111/jan.13031

Klehe, U. C. (2004). Choosing how to choose: Institutional pressures affecting the adoption of personnel selection procedures. *International Journal of Selection and Assessment, 12*, 327-342. doi: 10.1111/j.0965-075X.2004.00288.x

Kluger, A. N., & Rothstein, H. R. (1993). The influence of selection test type on applicant reactions to employment testing. *Journal of Business and Psychology, 8*, 3–25. doi: 10.1007/BF02230391

König, C. J., Jöri, E., & Knüsel, P. (2011). The amazing diversity of thought: A qualitative study on how human resource practitioners perceive selection procedures. *Journal of Business and Psychology, 26*, 437-452. doi: 10.1007/s10869-010-9199-9

Krauss, S., & Wang, X. T. (2003). The psychology of the Monty Hall Problem: Discovering psychological mechanisms for solving a tenacious brain teaser. *Journal of Experimental Psychology, 132,* 3-22. doi: 10.1037/0096-3445.132.1.3

Kristof, A. L. (1996). Person-organization fit: An integrative review of its conceptualizations, measurement, and implications. *Personnel Psychology, 49,* 1-49. doi: 10.1111/j.1744-6570.1996.tb01790.x

Landers, R. N., & Behrend, T. S. (2015). An inconvenient truth: Arbitrary distinctions between organizational, Mechanical Turk, and other convenience samples. *Industrial and Organizational Psychology, 8*, 142-164. doi: 10.1017/iop.2015.13

Landy, F. L. (1976). The validity of the interview in police officer selection. *Journal of Applied Psychology, 61*, 193-198. doi: 10.1037/0021-9010.61.2.193

Latham, G. P., Saari, L. M., Pursell, E. D., & Campion, M. A. (1980). The situational interview. *Journal of Applied Psychology, 65*, 422-427. doi: 10.1037/0021-9010.65.4.422

Latham, G. P., & Sue-Chan, C. (1996). A legally defensible interview for selecting the best. In R. S. Barrett (Ed.), *Fair employment strategies in human resource management* (pp. 134–143). New York: Quorum Books

Lievens, F., & De Paepe, A. (2004). An empirical investigation of interviewer-related factors that discourage the use of high structure interviews. *Journal of Organizational Behavior, 25,* 29-46. doi: 10.1002/job.246

Lodato, M. A., Highhouse, S., & Brooks, M. E. (2011). Predicting professional preferences for intuition-based hiring. *Journal of Managerial Psychology, 26*, 352– 365. doi: 10.1108/02683941111138985

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods, 44*, 1-23. doi: 10.3758/s13428-011-0124-6

McConnell, C. R. (1999). A working manager's guide to effective and legal employee selection interviewing. *The Health Care Supervisor, 17*, 77-89.

McMurry, R. N. (1947). Validating the patterned interview. *Personnel, 23*, 263-272.

Miles, A., & Sadler-Smith, E. (2014). "With recruitment I always feel I need to listen to my gut": the role of intuition in employee selection. *Personnel Review, 43*, 606-627. doi: 10.1108/PR-04-2013-0065

Mitchell, T. D., Thoroughgood, C. N., Campion, M. A., Jacobs, R. R., Oki, T., Wright, C., & Zhang, D. C. (2019, April). Horse-sized duck? Pros and cons of odd-ball interview questions. *SIOP Annual Conference.* Panel presented at 2019 SIOP conference, Washington D.C.

Moss, A., & Litman, L. (2020, June 12). *Demographics of people on Amazon Mechanical Turk*. CloudResearch. https://www.cloudresearch.com/resources/blog/who-uses-amazon-mturk-2020-demographics/

Mullinix, K., Leeper, T., Druckman, J., & Freese, J. (2015). The Generalizability of Survey Experiments. *Journal of Experimental Political Science, 2*(2), 109-138. doi: 10.1017/XPS.2015.19

Munk, N., & Oliver, S. (1997). Think Fast! *Forbes*, *159*(6), 146-150.

Orpen, C. (1985). Patterned behavior description interviews versus unstructured interviews: A comparative validity study. *Journal of Applied Psychology, 70*, 774-776. doi: 10.1037/0021-9010.70.4.774

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*, 879–903. doi:10.1037/0021-9010.88.5.879

Posthuma, R. A., Morgeson, F. P., & Campion, M. A. (2002). Beyond employment interview validity: A comprehensive narrative review of recent research and trends over time. *Personnel Psychology, 55,* 1-82. doi: 10.1111/j.1744-6570.2002.tb00103.x

Poundstone, W. (2003). *How would you move Mount Fuji?: Microsoft's cult of the puzzle-how the world's smartest companies select the most creative thinkers*. Little, Brown and Company.

Poundstone, W. (2012). *Are you smart enough to work at Google?: Trick questions, zen-like riddles, insanely difficult puzzles, and other devious interviewing techniques you need to know to get a job anywhere in the new economy*. Little, Brown and Company.

Ramsay, S. Gallois, C., Callan, V. J. (1997). Social rules and attributions in the personnel selection interview. *Journal of Occupational and Organizational Psychology, 70*, 189-203. doi: 10.1111/j.2044-8325.1997.tb00642.x

Rynes, S. L., Colbert, A. E., & Brown, K. G. (2002). HR professionals' beliefs about effective human resource practices: Correspondence between research and practice. *Human Resource Management, 41*, 149-174. doi: 10.1002/hrm.10029

Rynes, S. L., & Gerhart, B. A. (1990). Interviewer assessments of applicant "fit": An exploratory investigation. *Personnel Psychology, 43,* 13-35. doi: 10.1111/j.1744-6570.1990.tb02004.x

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. Psychological Bulletin, 124, 262-274. doi: 10.1037/0033-2909.124.2.262

Schmitt, N. (1976). Social and situational determinants of interview decisions: Implications for the employment interview. *Personnel Psychology, 29*, 79-101. doi: 10.1111/j.1744-6570.1976.tb00404.x

Schneider, B. (2001). Fits about fit. *Applied Psychology: An International Review, 50*, 141–152. doi: 10.1111/1464-0597.00051

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Simola, S. K., Taggar, S., & Smith, G. W. (2007). The employment selection interview: disparity among research-based recommendations, current practices and what matters to Human

Rights Tribunals. *Canadian Journal of Administrative Sciences/Revue Canadienne des Sciences de l'Administration, 24*(1), 30-44.

Spector, P. E. (2006). Method variance in organizational research: Truth or urban legend? *Organizational Research Methods, 9*, 221-232. doi:10.1177/1094428105284955

Speer, A. B., Tenbrink, A. P., Schwendeman, M. G., & Wegmeyer, L. J. (2020). Individual differences in effective interview design: Factors affecting question choice. *International Journal of Selection and Assessment, 28*(3), 310-321.

Stone-Romero, E. F. (2011). Research strategies in industrial and organizational psychology: Nonexperimental, quasi-experimental, and randomized experimental research in special purpose and nonspecial purpose settings. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology, Vol. 1. Building and developing the organization* (pp. 37–72). American Psychological Association. doi: 10.1037/12169-002

Sutton, A., & Watson, S. (2013). Can competencies at selection predict performance and development needs? *Journal of Management Development, 32,* 1023-1035. doi: 10.1108/JMD-02-2012-0032

Swider, B. W., Barrick, M. R., & Harris, T. B. (2016). Initial impressions: What they are, what they are not, and how they influence structured interview outcomes. *Journal of Applied Psychology, 101,* 625-638. doi: 10.1037/apl0000077

Swider, B. W., Zimmerman, R. D., & Barrick, M. R. (2015). Searching for the right fit: Development of applicant person-organization fit perceptions during the recruitment process. *Journal of Applied Psychology, 100,* 880-893. doi: 10.1037/a0038357

Taylor, P. J., & Small, B. (2002). Asking applicants what they *would do* versus what they *did do*: A meta-analytic comparison of situational and past behaviour employment interview questions. *Journal of Occupational and Organizational Psychology, 75,* 277-294. doi: 10.1348/096317902320369712

Thayer, J. D. (2002, April 2). Stepwise regression as an exploratory data analysis procedure [Paper]. American Educational Research Association, New Orleans, L.A. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.946.5295&rep=rep1&type=pdf

Toomela, A. (2008). Variables in psychology: A critique of quantitative psychology. *Integrative Psychological and Behavioral Science, 42*(3), 245-265. doi: 10.1007/s12124-008-9059-6

Ulrich, L., & Trumbo, D. (1965). The selection interview since 1949. *Psychological Bulletin, 63,* 100-116. doi: 10.1037/h0021696

Van Iddekinge, C. H., Raymark, P. H., Roth, P. L., & Payne, H. S. (2006). Comparing the psychometric characteristics of ratings of face-to-face and videotaped structured interviews. *International Journal of Selection and Assessment, 14*, 347-359. doi: 10.1111/j.1468-2389.2006.00356.x

Vleugels, W., De Cooman, R., Verbruggen, M., & Solinger, O. (2018). Understanding dynamic change in perceptions of person-environment fit: An exploration of competing theoretical perspectives. *Journal of Organizational Behavior, 39*, 1066-1080. doi: 10.1002/job.2294

Wagner, R. (1949). The employment interview: A critical summary. *Personnel Psychology, 2,* 17-46. doi: 10.1111/j.1744-6570.1949.tb01669.x

Werbel, J. D., & Johnson, D. J. (2001). The use of person-group fit for employment selection: A missing link in person-environment fit. *Human Resource Management, 40,* 227-240. doi: 10.1002/hrm.1013

Wright, C. W., Sablynski, C. J., Manson, T. M., & Oshiro, S. (2012). Why are manhole covers round? A laboratory study of reactions to puzzle interviews. *Journal of Applied Social Psychology, 42*, 2834-2857. doi: 10.1111/j.1559-1816.2012.00963.x

Yonge, K. A. (1956). The value of the interview: An orientation and a pilot study. *Journal of Applied Psychology, 40*, 25-31. doi: 10.1037/h0045240

Zedeck, S., Tziner, A., Middlestadt, S.E. (1983). Interviewer validity and reliability: An individual analysis approach, *Personnel Psychology, 36*, 355-370. doi: 10.1111/j.1744-6570.1983.tb01443.x

Zhang, D. C. (2021). Horse-sized ducks or duck-sized horses? Oddball Personality Questions are likable (but useless) for organizational recruitment. *Journal of Business and Psychology*, *2021*, 1-19.

Zickar, M. J., & Gibby, R. E. (2007). Four persistent themes throughout the history of I-O psychology in the United States. In L. L. Koppes (Ed.), *Historical Perspective in Industrial and Organizational Psychology* (pp. 61-80). New York, NY: Psychology Press.

APPENDIX A. INTERVIEW QUESTIONS

Interview Questions (Highhouse et al., 2019)

Brainteaser Questions

1. If you were to get rid of one state in the country, what would it be and why?
2. If Hollywood made a movie about your life, who would you like to see play the lead role as you?
3. Calculate the angle of two clock pointers when time is 11:50.
4. How would you explain what a chair is to an alien?
5. If you could be any animal on a carousel, what would you pick and why?
6. How would you determine the weight of a commercial airplane without a scale?
7. Why are manhole covers round?
8. What is your favorite song? Perform it for us now.
9. Estimate how many windows are in New York.
10. Pick two celebrities to be your parents.
11. Have you ever stolen a pen from work?*
12. How many quarters do you need to reach the height of the Empire State Building?
13. Name three previous Nobel Prize winners.
14. What do you think about when you are alone in your car?
15. What songs best describe your work ethic?

Behavioral Questions

1. What would you do if you disagree with a coworker?
2. What do you do when your schedule is interrupted? Give an example of how you handled it.
3. Describe a situation when you took a risk professionally. What was the outcome?
4. Tell me about a creative solution you developed for a challenging situation or problem.
5. What is the most stressful situation you have handled and what was the outcome?
6. Describe a decision you made that was a failure. What happened and why?
7. If you could start your career over again, what would you differently?*
8. Tell me about a time that you participated in a team. What was your role?
9. Tell me about a time when you failed.
10. Give an example of a goal you reached and tell me how you achieved it.
11. Have you handled a difficult situation? How?
12. Give an example of an occasion when you used logic to solve a problem.
13. Did you ever postpone making a decision? Why?

14. Tell me about a time when you were faced with conflicting priorities. How did you determine the top priority?
15. Tell me about a time when you misjudged a person.*

Traditional Questions

1. Why should we hire you?
2. Do you consider yourself a leader?
3. What do you look for in a job?
4. Can you work under pressure and deal with deadlines?
5. What do you know about our organization?
6. Are you a good listener?
7. What led you to this point in your life?
8. How long would it take you to make a meaningful contribution to our firm?
9. If you could be anyone else, who would it be?*
10. What do you love?
11. How do you get along with older coworkers?
12. Do you check voicemail and email while on vacation?
13. How successful do you think you've been so far?
14. Tell us why you want to work for us.
15. What salary do you think you deserve?

Note: All items were presented in a randomized order.
* The item was removed after Subject Matter Expert (SME) review because less than 60% of SMEs sorted the item into the correct category
Participants responded on a scale from 1 (*would never include*) to 5 (*would definitely include*).

APPENDIX B. EXPERIMENTAL SCENARIOS

Experimental Scenario:

Assess Fit

"Imagine that you work for a Fortune 500 company. You are constructing a pool of interview questions to be asked of all supervisory or management-level applicants. To help you create the question pool, your supervisor provides a specific goal for you to focus on. Your primary goal is to select questions that will determine **whether the person is a good fit for the organization**. Consider what you need to ask the applicant to determine if they are a good overall fit. **Please indicate the likelihood that you would select each of the items below to be included in the set of general interview questions**.

Remember, your goal is to assess fit."

Assess Competence

"Imagine that you work for a Fortune 500 company. You are constructing a pool of interview questions to be asked of all supervisory or management-level applicants. To help you create the question pool, your supervisor provides a specific goal for you to focus on. Your primary goal is to select questions that will determine **the applicant's overall job competence**. Consider what you need to ask the applicant to determine if they are highly likely to perform well on the job. **Please indicate the likelihood that you would select each of the items below to be included in the set of general interview questions**.

Remember, your goal is to assess competence."

Assess Tolerance for Stress

"Imagine that you work for a Fortune 500 company. You are constructing a pool of interview questions to be asked of all supervisory or management-level applicants. To help you create the question pool, your supervisor provides a specific goal for you to focus on. Your primary goal is to select questions that will determine **the person's ability to handle stress**. Consider what you need to ask the applicant to determine if they are able to handle pressure on the job. **Please indicate the likelihood that you would select each of the items below to be included in the set of general interview questions.**

Remember, your goal is to assess stress tolerance."

APPENDIX C. MANIPULATION CHECK

Manipulation Check

Which of the following best describes **your provided goal** for creating the pool of interview questions?

1. Assess applicant fit
2. Assess applicant competence
3. Assess applicant stress tolerance

Interview Experience

Which of the following best describes your experience with conducting interviews?

1. I have never conducted an interview.
2. I have conducted a handful of job interviews over my career
3. I have conducted quite a few job interviews over my career.
4. I have conducted an enormous amount of job interviews over my career.