UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

THE RELENTLESS PURSUIT OF CONSTRUCT VALIDITY: WHY DO EMPLOYMENT INTERVIEWS PREDICT JOB PERFORMANCE?

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

MARIA RIAZ HAMDANI
Norman, Oklahoma
2011

THE RELENTLESS PURSUIT OF CONSTRUCT VALIDITY: WHY DO EMPLOYMENT INTERVIEWS PREDICT JOB PERFORMANCE?

A DISSERTATION APPROVED FOR THE
MICHAEL F. PRICE COLLEGE OF BUSINESS

BY

_____
Dr. Michael R. Buckley, Chair

_____
Dr. Mark C. Bolino

_____
Dr. Bret H. Bradley

_____
Dr. Eric A. Day

_____
Dr. Robert A. Terry

ACKNOWLEDGEMENTS

I am indebted to my committee chair Dr. Michael R. Buckley for his generous support, inspiration and mentorship throughout the Ph.D. program. His constant encouragement empowered me to fulfill my commitments as a doctoral candidate. I see him as a true mentor who is sincerely interested in his students' betterment and advises them on all aspects of becoming an academician. I am also grateful to my committee member. I would also like to thank the rest of my committee members; Dr. Mark C. Bolino for being a great role model and for his willingness to give feedback and extend support; Dr. Bret H. Bradley for giving me the opportunity to work on several research projects with him and for offering detailed feedback on my dissertation work; Dr. Eric A. Day for offering an awesome seminar on Assessment Centers and inspiring many key ideas in my dissertation work; Dr. Robert A Terry, for his insightful classes on statistical methods and for helping me grapple with the data analysis issues throughout my stay at OU. I am honored to have an excellent group of scholars as my dissertation committee members. I would also like to acknowledge many other faculty members, in the Price College and in the Department of Psychology, who contributed to my learning and who made my stay meaningful at OU.

My special thanks to Cynthia Lopez, Director J.C. Penney Leadership Program for agreeing to participate in this research. Despite her busy schedule, she was always willing to offer resources, time and valuable input for my research work. I am also thankful to all the students who volunteered to participate in my

study as research subjects. I am earnestly thankful to the fellow graduate students from the management and the psychology department who took out time to facilitate my data collection activities. I am especially thankful to my colleague and friend, Sorin Valcea, who openhandedly supported me during the intensive data collection and coding tasks despite his own dissertation work commitments.

I also want to acknowledge the support and kindness of my family; when faced with challenges, their love and friendship helped me keep a perspective. Last but not the least, as I cover another milestone in my life, I realize that the deepest gratitude is deserved by the entity that creates and sustains us; gives us the opportunity, willingness and ability to appreciate, experience and explore the amazing construction around us.

TABLE OF CONTENTS

LIST OF TABLES

Essay 2 Tables

Essay 3 Tables

LIST OF FIGURES

DISSERTATION INTRODUCTION

Employment interviews are one of the most frequently researched areas in the human resources management research (Buckley, Norris, & Wiese, 2000). Over the last eighty years, the employment interview research has gone through several phases. Fifty years ago, researchers seemed to have little hope for the criterion-related validity of employment interviews (Ulrich & Trumbo, 1965; Wagner, 1949). However, in the 90s several methodological advancements furnished evidence to support the criterion validity of employment interviews (Huffcutt & Arthur, 1994; McDaniel, Whetzel, Schmidt, & Maurer, 1994). Structuring interviews has been the major antecedent of increasing validity of interviews since then. The patterned behavior description interviews (Janz, 1982) and the situational interviews (Latham, Saari, Pursell, & Campion, 1980) have become the two most widely used techniques of conducting structured employment interviews. With all these advancements, any doubts regarding the criterion validity of structured employment interviews have been nullified. This development has freed researchers to turn their attention toward a newer set of issues in this line of research.

One new question attracting considerable attention from researchers is "why do employment interviews predict performance?" On the surface it seems to be a straightforward question with an easy answer. However, research so far has provided equivocal results regarding the construct validity of employment

interviews. My three-essay dissertation focuses on resolving those issues that plague the construct validity of employment interviews. In the first essay, I review the existing research that is concerned with the constructs underlying employment interviews. I examine various issues that emerge during the stages of interview design, administration, assessment, and analysis. I pay special attention to the theoretical issues that hinder construct validity evidence. More specifically, I highlight how grounding employment interviews in the theory of job performance (Campbell, McCloy, Oppler, & Sager, 1993) might positively influence construct validation efforts. I also stress the need to view an employment interview as a predictor method (rather than as a predictor construct) and emphasize the unique construct validation challenges that any predictor method (e.g., assessment center) faces. The purpose of the first study is to lay out a comprehensive framework that guides future research.

Many researchers have shown that different types of interview questions (e.g., behavioral questions, situational questions, and general questions) tap different constructs (e.g., Conway & Peneno, 1999). My second essay explores the construct validity of past behavior description interviews (PBDI). To date several studies have explored the constructs underlying PBDI questions. Social skills, experience, motivation, and intelligence have been proposed as some of the constructs that explain why PBDI questions predict performance (O'Leary, 2004). However, there seems to be little consensus on the specific constructs that are tapped by the different interview types. PBDIs are designed on the premise that a

pattern of past behaviors is a best predictor of future behaviors. The structuring efforts are aimed at ensuring that interview responses only assess the quality of past behaviors, and that no other biases (e.g., likeability of a candidate, personal biases of an interviewer, and impression management by a candidate) influence the evaluation of a candidate. If PBDI questions only measure the quality of past behaviors then the assessment of candidates based on oral interviews and the assessment based on written responses should be the same. I challenge this assertion and argue that an interview is primarily a social interaction process. Thus, even highly structured interviews, such as PBDI, measure additional constructs such as extraversion, emotional stability, self-esteem, and self-efficacy.

The third study explores the construct validity and incremental validity of general interview questions. Despite research evidence that supports the use of PBDI- and SI-type questions, managers continue using general questions in actual job interviews. In this study, I propose how to make the general interview questions more effective. I propose that researchers can increase the validity of general questions when they have an *a priori* understanding of underlying constructs, they ground the interview design in a relevant theory, and they make the intended dimensions transparent to the candidates. I argue that general type of questions can assess certain constructs such as values and personality, and thus add incremental validity to the interview process. In addition, by incorporating general questions in the interview design, interviews achieve the flexibility desired by practitioners and increase their comfort level with the interview process.

REFERENCES

Buckley, M. R., Norris, A. C., & Wiese, D. S. (2000). A brief history of the selection interview: may the next 100 years be more fruitful. *Journal of Management History, 6*(3), 113-126.

Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1993). A theory of performance. In N. Schmitt, W. C. Borman & Associates (Eds.), *Personnel selection in organizations* (pp. 35-70). San Francisco, CA: Jossey-Bass Publishers.

Conway, J. M., & Peneno, G. M. (1999). Comparing structured interview question types: Construct validity and applicant reactions. *Journal of Business and Psychology, 13*(4), 485-506.

Huffcutt, A. I., & Arthur, W. (1994). Hunter and Hunter (1984) revisited: Interview validity for entry-level jobs. *Journal of Applied Psychology, 79*(2), 184-190.

Janz, T. (1982). Initial comparisons of patterned behavior description interviews versus unstructured interviews. *Journal of Applied Psychology, 67*(5), 577-580.

Latham, G. P., Saari, L. M., Pursell, E. D., & Campion, M. A. (1980). The situational interview. *Journal of Applied Psychology, 65*(4), 422-427.

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

O'Leary, R. S. (2004). *A construct validity comparison of alternative structured interview scoring methodologies.* Ph.D., Auburn University, Auburn. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2004-99016-292&site=ehost-live

Ulrich, L., & Trumbo, D. (1965). The selection interview since 1949. *Psychological Bulletin, 63*(2), 100-116.

Wagner, R. (1949). The employment interview: A critical review. *Personnel Psychology, 2*, 17-46.

ESSAY 1: THE RELENTLESS PURSUIT OF CONSTRUCT VALIDITY: WHY DO EMPLOYMENT INTERVIEWS PREDICT JOB PERFORMANCE?

Abstract

Why do employment interviews predict job performance? The construct validity of employment interviews is the biggest challenge faced by employment interview researchers. In this study, I discuss in detail various theoretical and methodological issues which have an influence upon the construct validity of employment interviews. These issues emerge at all stages of employment interview including design, administration, assessment, and analysis. So far, structuring of employment interviews has been the primary driver of construct validation. I argue that for future endeavors theoretical grounding of employment interviews in theory of job performance should be the main driver for establishing the construct validity of interviews. I provide a road-map explaining how to bring theoretical rigor for advancing future construct validity endeavors.

The relentless pursuit of construct validity: Why employment interviews predict job performance?

Employment interviews have been around for almost as long as people have had to work for others. As such, the employment interview is one of the oldest and most frequently investigated areas in human resources management research (Buckley, Norris, & Wiese, 2000). As considerable support for criterion-related validity has already been demonstrated (Huffcutt & Arthur, 1994; McDaniel, Whetzel, Schmidt, & Maurer, 1994), many scholars consider the construct validity of employment interviews to be the next big puzzle to be solved (Buckley & Russell, 1999; Campion, Palmer, & Campion, 1997; Cortina, Goldstein, Payne, Davison, & Gilliland, 2000; Huffcutt, Roth, Conway, & Stone, 2001a; Macan, 2009; O'Leary, 2004). However, others have advised against confronting this puzzle by alluding to similarly intended and somewhat futile efforts with respect to assessment center research. Many consider this to be a vacuous endeavor (c.f. Harris, 1999) suggesting that the interview is a versatile instrument and can measure any construct one wishes to investigate (Dipboye, 1992).

I believe that the pursuit of construct validity is neither elusive nor vacuous. I strengthen and buttress my arguments by reviewing the relevant literature and answering the following important questions: Why is it essential to have construct validity evidence when we have sufficient criterion validity evidence? What have we learned so far regarding the construct validity issue from the current employment interview research? What has been done to address

construct-validity related issues, and what are the next logical steps in the research process?

By confronting these questions, I aim to provide a road-map for advancing future construct validation endeavors. Proposing this research agenda is essential and timely for employment interview research. Almost all the major employment interview reviews published in the last ten years have enlisted construct validation as the research agenda for future employment interview studies (e.g., Macan, 2009; Ployhart, 2006; Posthuma, Morgeson, & Campion, 2002). Thus, my first study reviews a significant research area and provides guidelines on construct validation issues. These issues are of broad interest to researchers in the areas of employment interview and personnel selection.

A central aspect of my approach is the treatment of factors that may influence the construct validity of an interview. For instance, in the last twenty years, several primary studies for assessing the construct validity of employment interviews have provided equivocal results. Some studies find support for construct validity (e.g., Klehe & Latham, 2006; Motowidlo et al., 1992; Van Iddekinge, Raymark, & Roth, 2005) while many others find little evidence of construct validity (e.g., Conway & Peneno, 1999; Menkes, 2002; Van Iddekinge, Raymark, Eidson, & Attenweiler, 2004). This ambiguity points towards the need to highlight important factors (or moderators) that influence the construct validity of interviews. In this respect, my review of primary studies of construct validity goes well beyond

meta-analytical efforts (e.g., Huffcutt, et al., 2001a; Salgado & Moscoso, 2002) which generally aim to indirectly identify underlying constructs in interviews.

I have organized my discussion around the key stages of the interviewing process including: design and development, administration, assessment of responses, and analysis. This organization will facilitate the identification of the nature of construct validity challenges present in different stages of the interviewing process and aid in the construction of an agenda for future research. Towards the end, I synthesize these findings to provide a comprehensive framework for future construct validation efforts. The primary objective of this analysis is to highlight key theoretical and methodological challenges and provide concrete suggestions for moving forward with this program of research.

Employment Interview is a Predictor Method

The first step toward construct validity of employment interviews is to understand the nature of the tool itself. An employment interview is a predictor method which is designed to provide information on a wide array of predictor constructs. Arthur and Villado (2008) cautioned researchers against the practice of confounding predictor methods with predictor constructs. "Predictor constructs may include or take the form of psychological constructs and variables, such as general mental ability, conscientiousness, psychomotor ability, and perceptual speed. They can also take the form of situational or job-content-based behaviors, such as word processing or troubleshooting an F-16 jet engine. In contrast, predictor methods may take the form of interviews, paper-and-pencil tests, and

computer-administered, video-based, or simulation-based modes of assessment" (Arthur & Villado, 2008, p. 436).

It is common in personnel selection literature to fail to distinguish between a predictor method and a predictor construct. For instance, researchers often compare the criterion-related validity of employment interviews (a predictor method) with that of cognitive ability (a predictor construct) (e.g., Campion, Campion, & Hudson, 1994; Cortina, et al., 2000). This lack of distinction between predictor methods and predictor constructs creates confusion by minimizing those unique issues that are present in the construct validity of predictor methods. Roth and his colleagues (2005) explained in detail the complexities involved in construct validation of employment interviews. They argued that unlike other predictors in applied psychology (e.g., IQ tests) employment interviews do not measure "clean" constructs, and are not designed to be "construct-centered." Instead employment interviews are designed to be "job-centered", and each identified dimension of work behavior taps into constructs which are intertwined with each other in complex ways (Roth, et al., 2005).

Unlike predictor constructs, predictor methods have much more complex psychometric properties. For instance, core self-evaluation is a predictor construct measured through a 12-item scale (Judge, Erez, Bono, & Thoresen, 2003). The construct validity of a core self-evaluation scale requires evaluating whether the twelve items in the scale are providing information on a specific psychological characteristic – core self-evaluation – known a priori to the researcher. On the other

hand, in employment interviews researchers often do not have a priori knowledge of the constructs being measured by the interview (c.f. Krajewski, Goffin, McCarthy, Rothstein, & Johnston, 2006). Being a predictor method, an employment interview can potentially measure almost any job-related construct.

These issues make the task of construct validation of employment interviews challenging. In other words, in each new study researchers have to identify what different constructs might be tapped by the interview dimensions. This requires re-exploring the complex predictor-criterion link every time one moves from one job to another (Klimoski, 1993). In addition, once the interview dimensions or constructs are designed, one needs to ensure that interview administration and response assessment is done effectively for accurately measuring the intended dimensions or constructs. Thus, construct validity of an employment interview requires developing a framework that envelopes all the processes involved in conducting an employment interview. My further discussion is organized around these key stages of the employment interview process.

Four Stages of Construct Validity

Construct validity concerns testing the theory behind a test or a measure. More specifically, it means testing whether the tool is measuring what we think it is measuring. Hence, the construct validity of employment interview means understanding why an employment interview predicts performance. According to Cronbach (1990) construct validation includes the following three elements: (a) suggesting which constructs account for variation in test scores. For example, if we

are validating an interview score, we need to propose why some applicants achieve higher ratings and others receive lower ratings; (b) deriving hypotheses from the theory relating to the construct. For example, explain how motivation (a construct hypothesized to be measured by employment interview) leads to high levels of job performance; (c) testing the hypotheses empirically.

An employment interview can be divided in the following three stages: design and development of an interview, actual administration of an interview, and assessment of the candidates. I will add a fourth and a final stage of "analysis" for discussing how construct validity analysis is carried out. Each stage has a different focus and a distinct set of factors that influence construct validity. In the following section, I will discuss these issues in more detail.

Stage 1: Design and Development of an Employment Interview.

The focus of the design and development stage is to know "what should be measured in an interview?" and "how it should be measured?" The former question concerns developing a theory of job performance, whereas, the latter question revolves around designing interview elements (such as type of questions, content of questions, number of interview questions, and number of interview sessions) that meet the psychometric standards.

Theory development. The central task in designing a selection interview is to understand what needs to be measured in a selection interview. This means understanding the task requirements of the focal job. More specifically, the focus is to define the criterion space and explicate the predictor-criterion link. This is

generally done by conducting some type of a job analysis such as critical incident analysis (Flanagan, 1954). A job analysis helps in identifying key work behaviors, which in turn leads to the definition of critical performance dimensions for the focal position.

Overall, in the last twenty years researchers have been successful in increasing the criterion-related validity of selection interviews by making the interviews job relevant and by introducing the notion of structure (O'Leary, 2004). However, structure and job-relevance do not necessarily ensure theoretical rigor, they merely ensure uniformity. At present, after the identification of desired work behaviors through a job analysis, no effort is made to link these job behaviors with the extant literature in order to define the criterion space and explicate predictor-criterion link. This is somewhat congruent with the assessment center research where casual definition of constructs without focusing much on theory, and lack of attention towards psychometric standards of construct definition have resulted in proliferation of weakly defined constructs (Arthur & Day, 2011). A poor definition of constructs and a lack of theoretical explication of predictor-criterion link at the design stage pose major challenges in ascertaining the construct validity of employment interviews.

An employment interview is a predictor method and can measure any job-related construct. It is designed to assess variables that can predict which candidates will perform well at a given job. Hence, in employment interviews job performance constitutes the criterion space. In order to introduce theoretical rigor in

the definition of criterion space it is important to view this space from a lens of job performance. Over the past years, several job performance frameworks have been introduced including: person-job-organization fit framework (Kristof-Brown, Zimmerman, & Johnson, 2005), task and contextual performance (Borman & Motowidlo, 1993) and knowledge-skill-ability framework (Campbell, McCloy, Oppler, & Sager, 1993). These theoretical frameworks provide basic framework within which components of selection interview can be designed for a systematic investigation.

Each job performance framework underscores different aspects of work performance. The person-job-organization fit theory emphasizes the notion of compatibility between a job candidate and the employer and/or the job. The job candidate can be compared and matched with the broader organization and its culture, the immediate working group, the supervisor or the job at hand. Thus, this framework offers flexibility in viewing the suitability of candidates from variety of perspectives. This flexibility is especially important for organization wide selection. When considering a wider scope, selection in an organization involves certain themes that are common across various job families and job levels (e.g., values and attitudes espoused by an organizational culture) and many other characteristics that are unique for a given job. The fit framework is flexible enough to capture the varying demands of a job in an organization. For example, for an entry level jobs it might be more important to explore the person-job fit, whereas, for a senior level position person-group or person-organization fit may become

more important. However, in order to define a good fit this framework demands considerable introspection from an organization that should go beyond understanding individual jobs. Thus, a simple job analysis might not be sufficient to fully utilize this framework; an organization might need to explore its culture, climate and values across different levels.

Another framework concerning job performance pertains to the definition of performance. Borman and Motowidlo (1993) argue that it is not only the task performance that defines job performance but also the contextual performance. This particular framework helps view job performance from a broader aspect. Contextual performance is an important aspect of job performance and should be considered while making selection decisions. There are three important components of contextual performance – organizational citizenship behavior, prosocial organizational behavior, and soldier effectiveness (Borman & Motowidlo, 1993). Task performance is directly related to the activities that directly or indirectly contribute to the technical core, whereas, contextual performance is related to activities that increase the organizational effectiveness by supporting the broader social and psychological environment within which technical core exists (Borman & Motowidlo, 1993).

There is yet another job performance framework proposed by Campbell et al. (1993). According to this framework, the criterion space consists of various components, each with its own set of antecedents (Campbell, et al., 1993). Campbell and his colleagues (1993) defined job performance as a set of behaviors

that are under the complete control of an individual and hence distinct from the outcomes associated with job performance (i.e., efficiency and effectiveness) which are not under a complete control of job incumbent.

Campbell et al. (1993) viewed job performance to have eight major components: Job specific task proficiency (i.e., performance of technical tasks such as, designing a bridge or preparing cash-flow statements); non job-specific task proficiency (i.e., performance on tasks which are common to all jobs in an organization such as writing project proposals in a consulting firm); written and oral communication proficiency; demonstrating effort; maintaining self-discipline; facilitating team performance through supportive behaviors; supervision or leadership; and general management and administration. Campbell et al. (1993) argued that individual differences on any of the performance components (e.g., leadership behavior) are a function of knowledge, skills, and motivation. Thus, knowledge, skills and motivation are the proximal determinants of job performance (hereafter referred to as determinants). Each proximal determinant (e.g., interpersonal skills) of performance in turn has its own set of distal antecedents (e.g., ability, personality, aptitudes, education, training, values, beliefs, and needs. There is a vast body of research that informs how distal antecedents (hereafter referred to as antecedents) account for a variance in a particular determinant of a job performance.

All of the frameworks discussed above emphasize different aspects of job performance and can supplement each other. The notion of person-job-organization

fit enables us to view the selection process in terms of organizational needs and how individuals can satisfy these requirements. The distinction between task performance and contextual performance supplements the discussion on person-organization fit as well as helps expand the criterion domain. The framework proposed by Campbell et al. (1993) emphasizes the predictor-criterion link and helps in identifying what is being measured in the interview. In this sense, Campbell et al.'s (1993) theory of job performance serves as a primary framework that can be used to map any other theory of job performance into a predictor-criterion space.

Overall, the employment interview research can benefit by grounding interview design in a theory of performance. At present, little effort has been made to integrate the work dimensions that are identified through a job analysis, to any theoretical framework. Each study identifies and selects performance dimensions (e.g., handling irate customers or motivating sub-standard performers) on a basis of a job analysis. The dimensions selected in this manner are often job specific, do not necessarily conform to any standard knowledge-skill-ability inventory, and contribute towards construct proliferation. For instance, some of the seventeen work dimensions identified by Campion et al. (1994) include: initiative, teamwork, resolving conflict, commitment to improvement, work ethic, safety orientation, accepting responsibility, growth orientation. In this study, the subsequent development of questions and design of assessment keys were based solely on the results of job analysis with no theoretical grounding. The resulting constructs

proliferation and a lack of a common theoretical ground hinders the task of construct validity of selection interviews.

As a first step, theoretical grounding of interviews requires linking the work dimensions with broader performance components that differentiate the criterion space. Bartram (2005) argued for a "criterion-centric" approach of job performance that can help in defining the criterion-space according to some model of job performance. According to Bartram (2005), a criterion-centric approach enables defining a priori hypotheses of job performance and facilitates finding clearer and stronger empirical evidence for the relationships between predictors and job performance components. Therefore, the success of construct validation of employment interviews to a large extent depends on differentiating the criterion space and defining one-to-one relationships between components and predictors of performance.

The second step in theoretical grounding of interviews depends on understanding which aspects of performance are being measured. An employment interview is a predictor method and can potentially measure any construct in the nomological network of individual job performance including performance components (e.g., helping behavior or financial planning), determinants of performance (knowledge of labor laws, interpersonal skills, motivation to help others) and antecedents of performance (e.g., conscientiousness, supervisory experience, and altruism). In employment interview research, it is often not clear which part of the antecedent-determinant-performance link is being measured.

Studies generally state that a job analysis resulted in identifying the critical components of jobs. Some studies do not even mention the dimensions being measured in the interview (e.g., Barrick, Patton, & Haugland, 2000; Janz, 1982) and others merely provide the list of dimensions without mentioning whether these dimensions are performance components, determinants or antecedents of performance. Many studies have a list of dimensions that are a mix of components, determinants and antecedents (e.g., Krajewski, et al., 2006; Schuler, 1989; Van Iddekinge, et al., 2004). Similarly, there are studies that solely focus on measuring performance behaviors in the employment interview (e.g., Huffcutt, Weekley, Wiesner, Groot, & Jones, 2001b; Latham & Skarlicki, 1995).

The issue of not explicating which particular aspect of antecedent-determinant-performance link is being tapped in the employment interview impedes construct validity efforts. The complexity of construct validity efforts is a direct function of what constructs an interview is designed to measure. For instance, an interview that directly taps into the antecedents of performance, such as personality (e.g., Van Iddekinge, et al., 2005), is by definition simpler than an interview that taps into performance components or a mix of different aspects of performance (e.g., Conway & Peneno, 1999). Interviews that tap into more complex mixes of factors or the right-hand side components of the job performance equation have a wider network of constructs to investigate. For example, the construct validity of an interview that is designed to measure leadership behaviors (a performance component) will have to assess both determinants of performance

(e.g., interpersonal skills) and antecedents of performance (e.g., assertiveness) as possible explanation of why different candidates scored differently on the interview. On the other hand, an interview that is designed to measure conscientiousness (an antecedent of performance) is simpler and cleaner in terms of the nature of underlying constructs that explain variance in the interview score.

Development of interview elements. The second important component at the design stage is development of various interview elements (e.g., interview questions, number of interview sessions, and a scheme for rating interview responses). A job analysis is commonly an important source of developing these elements. The development of interview elements faces various theoretical as well as methodological challenges. There are two types of interview questions that are widely used in the structured employment interviews: patterned behavior description interviews (PBDI or BDI) and situational interviews (SI). Theoretical rationale and empirical evidence suggests that different types of interview questions measure different constructs (Taylor & Small, 2002). For example, the situational interviews are grounded in goal-setting theory which extends the premise that intentions are the best predictor of future behavior (Latham, Saari, Pursell, & Campion, 1980). Similarly, the patterned behavior description interviews (PBDI or BDI) are based on the rationale that a pattern of past behaviors is the best predictor of future behaviors (Janz, 1982). Empirical evidence suggests that different types of questions provide information on different criterion measures (e.g., Klehe & Latham, 2006) and tap into different predictor constructs (e.g.,

Huffcutt, et al., 2001b). However, the results are ambivalent with no clear answers with regard to which types of interview questions should be designed for which types of constructs.

Selection of the interview content is another major aspect in defining the interview elements. In the majority of the studies, the content of interview questions is derived from the job relevant information gained through a job analysis. For instance, during a critical incident analysis, examples of successful and unsuccessful events are collected from the job incumbents. These examples are often used to write various interview questions (e.g., Campion, et al., 1994; Latham & Skarlicki, 1995). The rigor of job analysis ensures that the interview questions are job relevant. However, these efforts are not sufficient to ensure that the interview questions are designed on the principles that assure measurement of theoretical constructs. Overall, there is a lack of effort to link the identified dimensions with relevant theoretical constructs.

This state of disconnect from the theory is also evident from the fact that none of the primary studies that explore the construct validity of employment interviews provide a full list of interview questions. Many studies provide few examples of the questions (e.g., Klehe, Konig, Richter, Kleinmann, & Melchers, 2008; Van Iddekinge, et al., 2005), and some do not even provide any example of the questions (e.g., Allen, Facteau, & Facteau, 2004; Van Iddekinge, et al., 2004). Efforts to validate interview questions pale in comparison to efforts to validate new measurement scales in other areas of organizational behavior. The measurement

scales used in other areas of organizational behavior research are often validated by following rigorous scientific procedures (Hinkin, 1998) and the development of these scales often involve dedicated studies (e.g., Bolino & Turnley, 1999; an extensive series of studies for developing an impression management scale). These scale-development studies provide evidence on all aspects of the construct validity including reliability, content validity, convergent validity, discriminant validity, and criterion validity. In addition, once the scale is validated, it continues to be used in the subsequent studies, thereby accumulating empirical evidence for the internal consistency and the proposed factor structure.

On the contrary, an employment interview is a predictor method whose content necessarily changes from job to job. It can probably be argued that it is over ambitious and unrealistic to draw comparisons between the psychometric properties of interview questions and psychological scales. However, one can also argue that the inherent context-specific nature of interview research should compel researchers to enforce higher standards of scale definition. There is a need for clear guidelines on developing theory-driven interview questions. The disconnect between the contents of the interview questions and theory and measurement principles is one of the key factors that weaken the construct validity of employment interviews.

Van Iddekinge et al. (2004) failed to find construct validity evidence for an interview study designed to select customer service managers. This study identified key work dimensions on the basis of job analysis, but did not provide sufficient

details (e.g., total number of dimensions, definition of these dimensions) on these identified dimensions in the methods section, and did not provide any example of the interview questions. The researchers grouped the dimensions into three main categories (stress management, interpersonal skills, and conscientiousness) at the later analysis stage using a non-quantitative method. No mention was made of how these work dimensions were related to the extant literature, and if any relevant theory was used in designing the interview questions. This indicates a lack of focus with regard to linking interview questions with the extant literature and relevant theory.

However, a year later, when some of these authors designed another study to specifically tap into the three main constructs identified in the earlier study for a customer services position, they found encouraging support for the construct validity of interviews (Van Iddekinge, et al., 2005). In the later study, it was known a priori that the interview questions were tapping three specific personality constructs. Though not clearly stated, it is evident from the comparison of the two studies that in the later study, the authors relied more on theory to design the interview questions. Moreover, by repeating a study involving the same job position and the same constructs, the authors were probably able to draw on the experience of the previous study, bringing improvement in the psychometric properties of the interview questions. This illustrates the importance of theoretical rigor at the design stage in improving the construct validity of employment interviews.

Overall, at the design stage of the interview, at present, rigorous job analysis is done to assure job relevance of the interview. Moreover, considerable efforts are made to ensure standardization of and structure in the interview (Huffcutt & Arthur, 1994). However, structure and job relevance are not enough to ensure construct validity of job interviews. At the design stage, theoretical grounding of interview is the third element that is currently missing but necessary to ensure construct validity of interviews.

Stage 2: Interview Administration.

The actual administration of an interview is the second stage that influences the construct validity of employment interviews. At this stage an interviewer and a job candidate interact with each other to exchange information. Thus, at this stage the primary focus is on the data collection. Construct validity at this stage is influenced by various issues that pertain to the judgment and evaluation of individuals during a social setting. At this stage various contextual, psychological and cognitive processes influence the construct validity.

In a structured interview, considerable effort is expended to ensure that all candidates are evaluated in a uniform environment. This distinguishes employment interviews from other predictor methods such as assessment centers. In assessment center research, the exercise effects are known to impact construct validity (Arthur & Day, 2011). Each exercise (e.g., in-basket exercise or leaderless group discussion) poses unique situational demands and elicits different set of behaviors. However, in a given structured employment interview all candidates are evaluated

in a similar situation (i.e., face-to-face questions and answers asked in a consistent manner). In this sense, structured employment interviews face limited construct validity challenges as compared to assessment centers. Therefore, instead of concentrating on contextual factors, I will focus more toward the psychological and cognitive factors that influence the construct validity of the interviews at the administration stage.

An interviewer's cognitive and psychological processes influence the collection and storage of information during the interview process. There is a vast body of decision-making literature (e.g., Tversky & Kahneman, 1974, 1981) and performance evaluation literature (e.g., Bernardin & Buckley, 1981; Cardy & Dobbins, 1986; De Nisi, Cafferty, & Meglino, 1984; Ferris, Judge, Rowland, & Fitzgibbons, 1994) that addresses these issues. At this stage, untrained interviewers, lack of observation and proper note keeping can deleteriously impact the validity of even properly designed interviews (Macan, 2009). There is a need to systematically investigate the influence of these factors on the interviewer's judgment and observation of the candidate during the interview administration.

For example, in terms of cognitive capacity, in the assessment center literature it has been shown that the number of dimensions assessors is asked to observe, record, and subsequently rate, and the conceptual distinctiveness of these dimensions influences the construct validity (Woehr & Arthur, 2003). In employment interviews, researchers have been gainsaying the optimal number of constructs that can be measured in interviews (e.g., Ulrich & Trumbo, 1965; Van

Iddekinge, et al., 2004). However, there is no systematic assessment of these issues. Moreover, the legal defensibility of interviews (Campion, Pursell, & Brown, 1988) while exploring these issues adds to the challenges associated with construct validation. An interview which is designed to tap fewer constructs might show superior psychometric properties but assessing limited number of constructs may pose challenges in terms of legal defensibility and perceived process fairness.

Another pertinent issue at this stage is how much control researchers have over the integrity of the actual interview process. In field studies, researchers probably have more control at the design stage. At the design stage, managers primarily provide information for job analysis and aid in question development. However, in the subsequent stages, managers are actively involved in actually conducting the interviews and assessing the candidates. In general, a systematic and scientific administration and assessment of interviews implies greater investment of time and cognitive resources. Scientific rigor also takes away considerable flexibility from the managers. For example, asking the interview questions in the same order without further probing – a typical requirement of a highly structured interview – imposes structure on the interview process as well as on the managers.

Almost all interview studies claim that managers were trained before the actual interviews. However, there is no systematic evidence that shows that as a result of training, these managers rated the candidates more scientifically (i.e., developed a common frame of reference) and did not rely on their idiosyncratic biases. Moreover, while getting field data for employment interviews, it is difficult

to convince managers of the utility of the rigorous scientific process. Often managers perceive their own systems to be working well and feel no need to fix something that is not broken.

At the candidate's end, faking and impression management have been discussed as major factors that influence the validity of employment interviews (Barrick, Shaffer, & DeGrassi, 2009). More research is needed to investigate which constructs are more prone to faking. For instance, communication skills are difficult to fake but traits like conscientiousness and helping behavior can be faked more easily. Researchers also need to investigate which aspects of impression management (e.g., verbal, non-verbal) influence which types of constructs.

Another important factor that may influence the candidate's understanding of the interview questions is the transparency of dimensions. In assessment center research it has been shown that making the dimensions transparent improves construct validity (Lievens, 1998). In employment interview research, Klehe et al. (2008) investigated the impact of transparency of dimensions on the validity of interviews. The study found that transparency of dimensions improves the performance of candidates, increases the construct validity, and does not influence the criterion-validity of interviews. Based on these results, Klehe et al. (2008) stated that transparency of interview dimensions should be considered a facet of interview structure.

In this regard, another important question is to investigate whether transparency of dimensions can increase the validity of general interview questions.

The general interview questions have been shown to tap different set of constructs (e.g., Conway & Peneno, 1999), and many managers continue using these type of questions. One challenge in incorporating general questions in interview research is their open-ended nature which makes these question less structured. However, there is a need to explore if transparency can bestow more structure upon these questions thereby positively influencing their construct validity.

Overall, at the interview administration stage, various issues need to be systematically examined for improving construct validity. In this regard, assessment center research can be taken as a general starting point. Many reviews and meta-analyses are available in assessment center research (e.g., Arthur & Day, 2011; Lievens, 1998; Sagie & Magnezy, 1997; Schleicher, Day, Mayes, & Riggio, 2002; Woehr & Arthur, 2003) that inform researchers about factors that influence the construct validity of predictor methods during administration and assessment stages.

Stage 3: Assessment of Responses.

The third stage involves the assessment of the information obtained by the interviewer. At this stage, the construct validity issues are primarily methodological in nature. Some of the aspects that influence the rating process are related to the cognitive capacity of information processing. More specifically, the construct validity at this stage drives from accurately assessing the stored information. One such aspect is related to the interrelation of constructs being measured. It is difficult for interviewers to accurately rate questions that tap into dimensions that have

considerable conceptual and construct space overlap (e.g.,"helping" and "relationships" in Conway & Peneno, 1999). This conceptual overlap limits the accurate assessment of these dimensions. The conceptual overlap has also been highlighted as a threat to construct validity in assessment center research (Arthur & Day, 2011).

Another critical issue is the timing of rating the responses. A response score can either be assigned by an in-interview method or a post-interview method. In the in-interview method, the scores are assigned during the interview immediately after each response is given by the candidate (e.g., Huffcutt, et al., 2001b). In the post-interview method, the scores are assigned after the interview is completed (e.g., Van Iddekinge, et al., 2005). The proponents of the in-interview approach consider immediate rating of interview responses a superior rating method as it reduces the formation of an overall interviewee effect which might result in halo effect. However, the proponents of the post-interview approach argue that rating during the interview not only affects the quality of assessment by adding cognitive burden but also biases (e.g., priming) the interviewers as they rate the candidates before asking the next question.

In the post-interview rating approach several alternatives are present to rate the candidates. Research has shown that the specific methodology (e.g., across dimensions-within question, within dimension-across questions, and across dimensions-across questions) used to rate the interview responses influence the construct validity (O'Leary, 2004). O'Leary (2004) found that response method

influenced the pattern of convergent and discriminant validity. The across dimension-across question method provided the most encouraging results and a confirmatory factor analysis showed that across question-across dimension scoring method accounted for the maximum trait variance and minimum methods variance compared to the other rating schemes. Thus, the across dimensions – across questions methodology produced relatively superior construct validity evidence.

We find similar results in assessment center research. Woehr and Arthur (2003) found that the within dimension-across exercises approach has more construct validity as compared to across dimensions-within exercise approach. The latter approach primarily had exercise effects. However, it is important to bear in mind the practical difficulties of various rating approaches. First, generally interviews are designed in a manner that each interview question just taps one dimension. This eliminates the option of across dimension-across question rating option – the one for which most encouraging results for construct validity have been found (O'Leary, 2004). Second, even if this rating scheme is possible, from a practical point of view, managers probably would prefer to rate candidates in the most efficient manner. For managers the easiest option is to rate candidates either through the in-interview method, or use the across dimension-within question option; as in this option one reads each question response just once and quickly rate all the dimensions reflected through the response.

However, as mentioned earlier, most studies are designed to tap one dimension per question. Thus, in the majority of the situations, the post-interview

approach involves reading each question response and evaluating the underlying dimension. In such an interview, a more useful comparison is between the validity of in-interview evaluation and post-interview evaluation. There is a need for studies that examine how these two response assessment strategies might influence the construct validity of interviews.

Stage 4: Construct Validity Analysis.

There are various methods of testing construct validity in the employment interview studies including multitrait-multimethod (MTMM) matrix, testing correlation with parallel tests of various predictor constructs, and factor analysis of the interview score. The multitrait-multimethod matrix (MTMM) approach is the most commonly used method to assess the construct validity of interviews. The MTMM matrix approach involves an integrated examination of the internal consistency, convergence of multiple measures of the same construct, and distinctiveness of measures of different constructs (Bagozzi & Edwards, 1998). Some studies of the employment interview have found evidence for construct validity by demonstrating convergent and discriminant validities (e.g., Motowidlo, et al., 1992) and others haves found no support for construct validity (e.g., Van Iddekinge, et al., 2004)

In the case of MTMM matrix, convergent validity is inferred by finding evidence for significant *mono trait-hetro method* (MtHm) correlations. The discriminant validity is established by showing that both *hetro trait-hetro method* (HtHm) and *hetro trait-mono method* (HtMm) correlations are smaller than the

MtHm correlations. In employment interview studies, the term hetro-method is often taken to mean different things in different studies. In most studies (e.g., Latham & Skarlicki, 1995; Schuler & Funke, 1989), hetro-method refers to the use of different types of interview questions (e.g., situational vs behavioral questions) which were asked in the same interview session (i.e., multimode interview), whereas in a fewer studies (e.g., Van Iddekinge, et al., 2004), hetro-method means different interview sessions having either the same types of questions or different types of questions (i.e., multiphase interviews).

In a multiphase interview, there are two separate interview sessions. The use of separate sessions makes the assessment of responses less susceptible to an interviewer effect. This is evident from a closer examination of the correlation matrices in the construct validity studies. In a multimode interview, all validities (i.e., MtHm, HtMm, and HtHm) are nearly equal (e.g., Conway & Peneno, 1999), suggesting a strong effect of the interviewer. Finally, there is another operationalization of hetro-method, in which different raters are regarded as different methods. In one study, behavioral interviews were conducted in two sessions, each with two interviewers (Van Iddekinge, et al., 2004). In this study, the researchers operationalized hetro-method by considering four raters as four different methods (Van Iddekinge, et al., 2004), thereby confounding the effects of different interview sessions with the effects of different raters. Thus, in the future there is a need to systematically examine the effects of various operationalization of construct validity on the employment interviews.

Another interesting issue at this stage is the nature of dimensions itself. Schuler (1989) argued that MTMM approach does not provide evidence for convergent and discriminant validity because of a lack of distinction between various interviews dimensions and an insufficient number of questions per construct. According to Kerlinger and Lee (1999), in order to establish discriminant validity one has to show positive, negative, and zero correlations. However, various interview dimensions are often correlated. The lack of conceptual distinction is further exacerbated by using multimodal interviews where interview effects dominate the interview scores. However, when construct validity is explored by comparing the interview scores with constructs measured through other methods (e.g., personality tests) more encouraging results are found (e.g., Van Iddekinge, et al., 2005) for convergent and discriminant validities.

In order to establish construct validity, some researchers also examine the factor structure of the interview scores (Klehe, et al., 2008; Schuler, 1989; Van Iddekinge, et al., 2004). The confirmatory factor analysis conducted by Van Iddekinge and colleagues (2004) showed that interviewee effects explain almost half the variance in an interview assessment, interview sessions explain around twenty-five percent of the variance, whereas, an interviewer factor explains almost twenty percent of the variance in the data. The traits or interview dimensions explained the least amount of variance – nine percent. Thus, for making inferences about construct validity, there is a need to systematically review different

operationalization of construct validity, and closely examine the results found by combining various analytical techniques.

Agenda for Future Research on Construct Validity of Interviews

In the preceding discussion, I examined the employment interview research in a systematic manner to identify and classify the issues related to the construct validity. I integrated the construct validity research in the four key stages and identified theoretical and methodological challenges that plague this research. In this section, my aim is to draw a road map for future by proposing the following steps.

Construct Validity of a Predictor Method

Framing employment interviews as a predictor method is one of the critical steps in integrating future construct validation efforts. The versatility and flexibility of employment interviews has been discussed in the past (e.g., Binning, LeBreton, & Adorno, 1999; Harris, 1999; Schuler, 1989). What is needed is to explicate how it influences the construct validity process. One of the main challenges in assessing construct validity of predictor methods is introducing theoretical rigor at the design and development stage of the tool. An employment interview is a flexible and context specific tool that can tap into a wide array of constructs (Binning, et al., 1999). This flexibility requires clear explication of what is being measured and how it will be measured.

Integration of interview design with the theory of job performance

Flexibility and versatility of employment interviews is a double-edged sword. The fact that interviews can be designed to measure almost any construct poses considerable challenges in integrating the research on some theoretical basis. One way to overcome this challenge is to ground interviews in a theory of job performance. An employment interview is a selection tool that is primarily used to identify the most suitable candidate for performing the job. This essentially means that employment interview is a tool to assess future job performance.

Various work dimensions identified through a job analysis should be grouped under higher-level job components. In other words, an integration of interview dimensions with a job performance theory would require researchers to state whether the interview is measuring samples of performance components (e.g., helping behavior, leadership behavior), determinants of job performance (i.e., knowledge, skills, and motivation), antecedents of performance (e.g., personality, values, or education), or some combination of these aspects. At present, very few employment interview studies have used an a priori definition of underlying constructs (Krajewski, et al., 2006). The clear explication of "what is being measured" will enable researchers to know in advance which component of antecedent-determinant-performance link is being measured.

This integration with the theory of performance may also help in identifying whether interviews are more suitable to measure performance components,

proximal determinants, or distal antecedents – thereby answering the call of researchers (e.g., Macan, 2009) to examine what should be measured by employment interviews. The nature of performance aspects also influences which types of interview questions are more suitable or effective for the validity evidence. For example, PBDI questions measure past behaviors (Janz, 1982) and are more suitable for interviews designed to measure performance components. Similarly, research has shown that general question tap into personality and attitudes (e.g., Conway and Peneno, 1999) and thus, might be more suitable for measuring antecedents of performance.

Another important task in the theoretical grounding of interviews is to explore how intended interview dimensions relate with the available research. For example, if an employment interview is measuring helping behavior, researchers have a vast body of literature to benefit from. Past research has shown that dispositional factors (Organ, 1990) and the values of employees predict helping behavior (Deckop, Mangel, & Cirka, 1999). The availability of meta-analyses and literature reviews provide another convenient way to understand the nomological network of a focal construct (e.g., an extensive review of OCB: Podsakoff, MacKenzie, Paine, & Bachrach, 2000). In other words, theoretical grounding of interviews makes clear the predictor-criterion link of an interview at the onset of the interview process.

The Design of Interview Content

The integration of interview development with the relevant theory requires designing and developing interview questions, and rating responses on a theoretical basis. Researchers need to go beyond job analysis and feedback of subject-matter-experts (i.e., managers and job incumbents). The current practice of seeking the input of job incumbents and managers during a job analysis certainly ensures the job relevance of an interview design, but is not sufficient to introduce a theoretical rigor. There is a need to refer to the research available on the intended constructs while designing interview questions and response keys. For example, research has identified five major dimensions of teamwork behavior (Stevens & Campion, 1994, 1999). The researcher designing interview questions and response keys for teamwork can use these dimensions as an additional source of information. The teamwork rating instrument developed by Stevens and Campion (1999) can aid in designing interview questions and rating responses.

To ensure theoretical rigor at the design stage researchers will also need to adopt some norms for publishing in employment interview research. In order to build upon the previous research and to add theoretical and methodological rigor, it is important to share a complete list of interview questions and rating responses in an appendix. This information is necessary in order to understand what is being measured and how it is being measured. In addition, sharing questions and response keys may also help in developing a pool of questions and responses which are both valid and theory driven.

The limited space in journals might not allow publishing the complete list of questions and assessment keys. In addition, some may argue that the publication of response assessment keys may negatively impact the validity of these questions by making them accessible for job candidates. However, with advent in technology and availability of online knowledge sharing platforms (such as *Measurement Toolchest* – an online library of references for existing scales – maintained by the Research Methods Division of the Academy of Management), a suitable method can be designed to share the vital information regarding interview response assessment keys for research purposes. Construct validation efforts will never be complete until we understand the complete nature of the measurement instrument. Thus, as employment interview researchers we need to assume higher standards of designing interview questions and response keys, and also make this information easily accessible to other researchers.

Exploring Context Specific Construct Validity

Researchers need to both integrate the employment interview research with the theory of job performance, and assess the construct validity of employment interviews in a context-specific manner. Specifying a context means specifying the criterion domain as well as the nature of the job. Consider the performance dimension of communication behavior in two different jobs – a human resource director for a manufacturing firm and a senior marketing coordinator for an architectural consulting firm. For the position of human resource director, an effective communication behavior involves a specific set of performance

determinants (e.g., knowledge of labor laws, understanding of various stakeholders, and know-how of organizational social network) and antecedents (e.g., tactfulness, and assertiveness). Similarly, for a senior marketing coordinator position the effective communication behavior requires an entirely different set of determinants (e.g., marketing and customer service skills, technical writing skills, Photoshop and Acrobat skills) and antecedents (e.g., on-the-spot thinking).

A job analysis process at the start of the employment interviews is designed to capture these job-specific differences. However, many construct validity studies ignore the context and aim to explore broader constructs that underlie any interview (e.g., Campion, et al., 1994; Latham & Skarlicki, 1995; Pulakos & Schmitt, 1995; Roth, et al., 2005). Such approaches under-emphasize job-specific requirements by exploring common construct underlying the overall interview score. This inevitably results in equivocal results. For example, some studies show that situational interviews tap into intelligence (Sue-Chan & Latham, 2004) and have validity (Latham & Skarlicki, 1995), while other studies find that situational interviews do not correlate with intelligence (Conway & Peneno, 1999) and have minimal validity (Pulakos & Schmitt, 1995). Perhaps this equivocation is due to context specific issues.

These equivocal findings have encouraged researchers to delve into possible job-specific moderators of validity. Some researchers have found support for job complexity as a moderator of construct validity for situational interviews. For instance, Huffcutt and his colleagues (2001b) conducted two studies to compare

PBDI and SI questions. They found that the two types of interview questions, though designed to measure the same constructs did not show convergent validity for the higher-level jobs. They found some evidence that suggests that for higher-level jobs PBDI questions might be more relevant as they tap into a critical determinant of performance at higher-level jobs – social skills (measured through extraversion). However, in this study, intelligence, age, tenure and other Big-five personality traits were not related to either SI or PBDI.

Similarly, Krajewski et al., (2006) found no relationship between intelligence and scores of SI questions for higher-level executive jobs. They found that PBDI scores were related to intelligence and outperformed the scores of SI questions in predicting performance. They explained that for higher-level jobs, SI might be tapping into some other constructs such as work values instead of intelligence. Some other researchers have also highlighted that for higher-level jobs executives may weigh every contingency of the situation while answering a situational question (Pulakos & Schmitt, 1995). In sum, these studies point towards the need for context-specific exploration of construct validity of employment interviews.

Utility and Validity

The utility of an interview is another concept that is important while discussing validity concerns. What is the practical value of measuring various constructs through a selection interview? Utility of a selection method is a direct function of the validity of a selection method, number of workers hired per year

through a given method, standard deviation of job performance, selection ratio, and the cost of administering a selection method (Schmidt & Hunter, 1998; Schmidt, Mack, & Hunter, 1984). Utility can either be measured in terms of dollar value of increased output or percentage of increase in the output. If bringing theoretical rigor in the interview process can aid in increasing the validity of a selection interview then this will certainly translate into enhancing the practical value or utility of the selection method. In addition, construct validity efforts should also guide us in understanding what constructs are more economically measured through a selection interview and what constructs can be more easily measured through other selection methods. For example, consider the utility of an employment interview for assessing personality of the job candidates. In this case, the cost of administering the interview should include the amount of resources invested in various activities such as designing the interview questions, conducting job analysis and reviewing the literature for preparing response assessment keys and training the interviewers for evaluating the responses. An alternative method to assess personality is administering a validated self-report test of personality (e.g., Goldberg et al., 2006). The comparison of both methods in terms of validity and utility should ultimately guide the decision of which selection method is the most suitable for a given construct.

Conclusion

The basic driver of construct validity so far has been structure in the interview process. The structure in the interview stems from standardization of

interview questions. Structure ensures that all candidates will be asked the same pre-determined questions in the same manner consistently. However, for future research much more work is needed to introduce theoretical rigor in the design and development of interview questions. More specifically, researchers should explicate the antecedent-determinant-performance link at the design stage. In order to move forward, greater effort is required to use extant literature for designing response assessment keys. It is also important to take measures that enable sharing detailed information regarding interview question and response assessment keys with other researchers.

The data collection in employee selection research poses special challenges. The ethical and practical concerns in the field studies take precedence over the design requirements of a systematic investigation. In addition, collecting an adequate sample size in field studies may result in longer time periods for completing a study. All these challenges probably raise concerns regarding the utility of construct validity studies for a researcher trying to make a career in the prevalent publish or perish paradigm. Thus, in addition to the theoretical and methodological initiatives suggested in this analysis, in order to move forward, the employment interview researchers would need to use various academic platforms to generate ideas that can help surmount the practical issues that decrease the utility of employment interview research.

REFERENCES

Allen, T. D., Facteau, J. D., & Facteau, C. L. (2004). Structured Interviewing for OCB: Construct Validity, Faking, and the Effects of Question Type. *Human Performance, 17*(1), 1-24.

Arthur, W., Jr., & Day, E. A. (2011). Assessment centers. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology, Vol 2: Selecting and developing members for the organization.* (pp. 205-235). Washington, DC: American Psychological Association.

Arthur, W., Jr., & Villado, A. J. (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection research and practice. *Journal of Applied Psychology, 93*(2), 435-442. doi: 10.1037/0021-9010.93.2.435

Bagozzi, R., & Edwards, J. (1998). A general approach for representing constructs in organizational research. *Organizational Research Methods, 1*(1), 45-87.

Barrick, M. R., Patton, G. K., & Haugland, S. N. (2000). Accuracy of interviewer judgments of job applicant personality traits. *Personnel Psychology, 53*(4), 925-951.

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

Bartram, D. (2005). The Great Eight Competencies: A Criterion-Centric Approach to Validation. *Journal of Applied Psychology, 90*(6), 1185-1203.

Bernardin, H. J., & Buckley, M. R. (1981). Strategies in Rater Training. *Academy of Management Review, 6*(2), 205-212.

Binning, J. F., LeBreton, J. M., & Adorno, A. J. (1999). Assessing personality. *The employment interview handbook*, 105–123.

Bolino, M. C., & Turnley, W. H. (1999). Measuring Impression Management in Organizations: A Scale Development Based on the Jones and Pittman Taxonomy. *Organizational Research Methods, 2*(2), 187-206.

Borman, W. C., & Motowidlo, S. J. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt, W. C. Borman

& Associates (Eds.), *Personnel selection in organizations* (pp. 71-98). San Francisco: Jossey-Bass.

Buckley, M. R., Norris, A. C., & Wiese, D. S. (2000). A brief history of the selection interview: may the next 100 years be more fruitful. *Journal of Management History, 6*(3), 113-126.

Buckley, M. R., & Russell, C. J. (1999). Validity evidence. In R. W. Eder & M. M. Harris (Eds.), *The employment interview handbook* (pp. 35-48). Thousand Oaks: California: Sage Publications.

Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1993). A theory of performance. In N. Schmitt, W. C. Borman & Associates (Eds.), *Personnel selection in organizations* (pp. 35-70). San Francisco, CA: Jossey-Bass Publishers.

Campion, M. A., Campion, J. E., & Hudson, J. P. (1994). Structured interviewing: A note on incremental validity and alternative question types. *Journal of Applied Psychology, 79*(6), 998-1102.

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*(1), 25-42.

Cardy, R. L., & Dobbins, G. H. (1986). Affect and appraisal accuracy: Liking as an integral dimension in evaluating performance. *Journal of Applied Psychology, 71*(4), 672-678.

Conway, J. M., & Peneno, G. M. (1999). Comparing structured interview question types: Construct validity and applicant reactions. *Journal of Business and Psychology, 13*(4), 485-506.

Cortina, J. M., Goldstein, N. B., Payne, S. C., Davison, H. K., & Gilliland, S. W. (2000). The incremental validity of interview scores over and above cognitive ability and conscientiousness scores. *Personnel Psychology, 53*(2), 325-351.

Cronbach, L. J. (1990). *Essentials of Psychological Testing* (5th ed.). New York: HarperCollins Publishers.

De Nisi, A. S., Cafferty, T. P., & Meglino, B. M. (1984). A cognitive view of the performance-appraisal process: A model and research propositions. *Organizational Behavior and Human Performance, 33*(3), 360-396.

Deckop, J. R., Mangel, R., & Cirka, C. (1999). Getting more than you pay for: Organizational citizenship behavior and pay-for-performance plans. *Academy of Management Journal, 42*, 420-428.

Dipboye, R. L. (1992). *Selection interviews: Process perspectives*. Cincinnati: OH: South-Western.

Ferris, G. R., Judge, T. A., Rowland, K. M., & Fitzgibbons, D. E. (1994). Subordinate influence and the performance evaluation process: Test of a model. *Organizational Behavior and Human Decision Processes, 58*(1), 101-135.

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51*(4), 327-358.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality, 40*(1), 84-96.

Harris, M. M. (1999). What is being measured? In R. W. Eder & M. M. Harris (Eds.), *The employment interview handbook* (pp. 143-157). Thousand Oaks: California: Sage Publications.

Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods, 1*, 104-121.

Huffcutt, A. I., & Arthur, W. (1994). Hunter and Hunter (1984) revisited: Interview validity for entry-level jobs. *Journal of Applied Psychology, 79*(2), 184-190.

Huffcutt, A. I., Roth, P. L., Conway, J. M., & Stone, N. J. (2001a). Identification and Meta-Analytic Assessment of Psychological Constructs Measured in Employment Interviews. *Journal of Applied Psychology, 86*(5), 897-913.

Huffcutt, A. I., Weekley, J. A., Wiesner, W. H., Groot, T. G., & Jones, C. (2001b). Comparison of situational and behavior description interview questions for higher-level positions. *Personnel Psychology, 54*(3), 619-644.

Janz, T. (1982). Initial comparisons of patterned behavior description interviews versus unstructured interviews. *Journal of Applied Psychology, 67*(5), 577-580.

Judge, T. A., Erez, A., Bono, J. E., & Thoresen, C. J. (2003). The core self-evaluations scale: development of a measure. *Personnel Psychology, 56*(2), 303-331.

Kerlinger, F., & Lee, H. (1999). Foundations of behavior research. Philadelphia: Harcourt College Publishers.

Klehe, U.-C., Konig, C. J., Richter, G. M., Kleinmann, M., & Melchers, K. G. (2008). Transparency in Structured Interviews: Consequences for Construct and Criterion-Related Validity. *Human Performance, 21*(2), 107-137.

Klehe, U.-C., & Latham, G. (2006). What Would You Do - Really or Ideally? Constructs Underlying the Behavior Description Interview and the Situational Interview in Predicting Typical Versus Maximum Performance. *Human Performance, 19*(4), 357-382.

Klimoski, R. J. (1993). Predictor constructs and their measurement. In N. Schmitt, W. C. Borman & Associates (Eds.), *Personnel selection in organizations* (pp. 99-134). San Francisco, CA: Jossey-Bass Publishers.

Krajewski, H. T., Goffin, R. D., McCarthy, J. M., Rothstein, M. G., & Johnston, N. (2006). Comparing the validity of structured interviews for managerial-level employees: Should we look to the past or focus on the future? *Journal of Occupational & Organizational Psychology, 79*(3), 411-432.

Kristof-Brown, A. L., Zimmerman, R. D., & Johnson, E. E. (2005). Consequences of individuals' fit at work: A meta-analysis of person-job, person-organization, person-group, and person-supervisor fit. *Personnel Psychology, 58*, 281-342.

Latham, G. P., Saari, L. M., Pursell, E. D., & Campion, M. A. (1980). The situational interview. *Journal of Applied Psychology, 65*(4), 422-427.

Latham, G. P., & Skarlicki, D. P. (1995). Criterion-Related Validity of the Situational and Patterned Behavior Description Interviews With Organizational Citizenship Behavior. *Human Performance, 8*(2), 67.

Lievens, F. (1998). Factors which improve the construct validity of assessment centers: A review. *International Journal of Selection and Assessment, 6*(141-152).

Macan, T. (2009). The employment interview: A review of current studies and directions for future research. *Human Resource Management Review, 19*(3), 203-218.

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

Menkes, J. M. (2002). *What do structured interviews actually measure? A construct validity study*. 63, ProQuest Information & Learning, US. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2002-95018-100&site=ehost-live

Motowidlo, S. J., Carter, G. W., Dunnette, M. D., Tippins, N., Werner, S., Burnett, J. R., & Vaughan, M. J. (1992). Studies of the Structured Behavioral Interview. *Journal of Applied Psychology, 77*(5), 571-587.

O'Leary, R. S. (2004). *A construct validity comparison of alternative structured interview scoring methodologies.* Ph.D., Auburn University, Auburn. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2004-99016-292&site=ehost-live

Organ, D. W. (1990). The motivational basis of organizational citizenship behavior. In B. M. S. L. L. Cummings (Ed.), *Research in Organizational Behavior* (Vol. 12). Greenwich: CT: JAI Press.

Ployhart, R. E. (2006). Staffing in the 21st Century: New Challenges and Strategic Opportunities. *Journal of Management, 32*(6), 868-897.

Podsakoff, P. M., MacKenzie, S. B., Paine, J. B., & Bachrach, D. G. (2000). Organizational citizenship behaviors: A critical review of the theoretical and empirical literature and suggestions for future research. *Journal of Management, 26*, 513-563

Posthuma, R. A., Morgeson, F. P., & Campion, M. A. (2002). Beyond employment interview validity: a comprehensive narrative review of recent research and trends over time. *Personnel Psychology, 55*(1), 1-81.

Pulakos, E. D., & Schmitt, N. (1995). Experience-based and situational interview questions: studies of validity. *Personnel Psychology, 48*(2), 289-308.

Roth, P. L., Van Iddekinge, C. H., Huffcutt, A. I., Eidson, C. E., Jr., & Schmit, M. J. (2005). Personality Saturation in Structured Interviews. *International Journal of Selection and Assessment, 13*(4), 261-273.

Sagie, A., & Magnezy, R. (1997). Assessor type, number of distinguishable dimension categories, and assessment centre construct validity. *Journal of Occupational & Organizational Psychology, 70*(1), 103-108.

Salgado, J. F., & Moscoso, S. (2002). Comprehensive meta-analysis of the construct validity of the employment interview. *European Journal of Work & Organizational Psychology, 11*(3), 299-324.

Schleicher, D. J., Day, D. V., Mayes, B. T., & Riggio, R. E. (2002). A new frame for frame-of-reference training: Enhancing the construct validity of assessment centers. *Journal of Applied Psychology, 87*(4), 735-746.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*(2), 262-274.

Schmidt, F. L., Mack, M. J., & Hunter, J. E. (1984). Selection utility in the occupation of U.S. Park Ranger for three modes of test use. *Journal of Applied Psychology, 69*, 490-497.

Schuler, H. (1989). Construct validity of a multimodal employment interview. In B. J. Fallon, H. P. Pfister & J. Brebner (Eds.), *Advances in industrial organizational psychology* (pp. 343-354). Amsterdam, North-Holland: Elsevier Science Publishers.

Schuler, H., & Funke, U. (1989). The interview as a multimodal procedure. In R. W. Eder & G. R. Ferris (Eds.), *The employment interview: Theory, research, and practice* (pp. 183-192). Newbury Park, CA: Sage Publications.

Stevens, M. J., & Campion, M. A. (1994). The Knowledge, Skill, and Ability Requirements for Teamwork: Implications for Human Resource Management. *Journal of Management, 20*(2), 503.

Stevens, M. J., & Campion, M. A. (1999). Staffing Work Teams: Development and Validation of a Selection Test for Teamwork Settings. *Journal of Management, 25*(2), 207-228.

Sue-Chan, C., & Latham, G. P. (2004). The Situational Interview as a Predictor of Academic and Team Performance: A Study of the Mediating Effects of

Cognitive Ability and Emotional Intelligence. *International Journal of Selection & Assessment, 12*(4), 312-320.

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 & Organizational Psychology, 75*(3), 277-294.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124-1131.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*, 453-458

Ulrich, L., & Trumbo, D. (1965). The selection interview since 1949. *Psychological Bulletin, 63*(2), 100-116.

Van Iddekinge, C. H., Raymark, P. H., Eidson, J. C. E., & Attenweiler, W. J. (2004). What Do Structured Selection Interviews Really Measure? The Construct Validity of Behavior Description Interviews. *Human Performance, 17*(1), 71-93.

Van Iddekinge, C. H., Raymark, P. H., & Roth, P. L. (2005). Assessing Personality With a Structured Employment Interview: Construct-Related Validity and Susceptibility to Response Inflation. *Journal of Applied Psychology, 90*(3), 536-552.

Woehr, D. J., & Arthur, W. (2003). The construct-related validity of assessment center ratings: A review and meta-analysis of the role of methodological factors. *Journal of Management, 29*(2), 231-258.

ESSAY 2: THE POTENT ROLE OF PERSONALITY IN A STRUCTURED BEHAVIORAL INTERVIEW: A FIELD EXPERIMENT

Abstract

This study is an attempt to explore the construct validity of patterned behavior description interviews (PBDI). PBDI are based upon the premise that a pattern of past behaviors is an effective predictor of future behaviors. The structuring efforts are aimed at ensuring that interview responses only assess the quality of past behaviors, and that no other biases influence the evaluation of a candidate. I challenge this assertion and argue that an interview is primarily a social interaction process. Thus, even highly structured interviews, such as PBDI, measure additional constructs such as personality of a candidate and confidence. The results of the study show that personality influences the candidates' performance in oral and written interview. Moreover, personality interacts with a type of the interview to predict future job performance.

Once when reflecting upon history, philosopher George Santayana stated that "Those who cannot remember the past are condemned to repeat it" (Santayana, 1905, p. 284). We look at this differently from a human resources management standpoint as we revisit the past hoping that candidates do repeat the past. We throw around aphorisms like "the best predictor of future behavior is past behavior." Our goal is to always predict which of a set of candidates will perform the best in the future based upon our insight into the past behaviors of a group of candidates. Accumulating this information can be accomplished in myriad different ways. One such procedure, the interview, has a long and illustrious history of predicting future behavior based upon past behavior. Numerous approaches to collecting interview information have been developed and one of these approaches serves as the focal point of this paper.

The patterned behavior description interviews (Janz, 1982) assess the suitability of a job candidate by asking questions that tap into patterns of past behaviors. The behavioral interviews are designed on the premise that past behaviors are the best predictor of future job behaviors. There is ample evidence supporting the criterion-related validity of PBDI in a wide-range of job settings (Campion, Campion, & Hudson, 1994; Harel, Arditi-Vogel, & Janz, 2003; Huffcutt, Weekley, Wiesner, Groot, & Jones, 2001b; Taylor & Small, 2002). A number of questions, though, still remain. Do structured patterned behavior description interviews (PBDI or BDI) predict performance by merely tapping into the past behaviors of a job candidate? To what extent structured behavior

interviews tap into personality-related constructs? These questions are essential to understand the construct validity of employment interviews.

Over the years, efforts have been made to improve structured patterned behavior interviews by introducing standardized assessment keys, asking a uniform set of questions from all the candidates, and by either limiting the information probing to a predetermined set of questions or eliminating it altogether (e.g., Huffcutt, et al., 2001b; Motowidlo et al., 1992). The structured PBDI questions are generally designed around the critical job dimensions identified through a job analysis. The objective of introducing structure in the behavioral interview is to ensure that the content of the applicant's answers contributes to the validity of interview scores.

Motowidlo and his colleagues (1992) conducted a series of interviews to investigate the validity of structured behavioral interviews. In one of the studies they asked trained raters to rate the audio tapes (in the absence of visual cues) and the written transcripts (in the absence of visual and oral cues) of actual interviews. The analysis of these ratings showed that the absence of visual and oral cues does not decrease the validity of interview questions. This provided support for the assertion that the content of the behavioral interview responses explains variance in the interview ratings. This finding is related to the broader issue of the construct validity of the employment interview.

Do structured behavioral interviews predict performance by merely tapping into the past behaviors of a job candidate? An employment interview is a social

interaction process, where both interviewer and the candidate interact with each other to exchange information. Thus, according to social cognition theories the interviewer and the candidate, both, engage in assimilation and judgment of social information. One objective of bringing structure and standardization in conducting the structured interviews – such as PBDI and situational interviews (SI) – is to eliminate systematic errors and reduce personal preferences and psychological biases of the interviewer. Motowidlo et al. (1992) reported that in the structured behavioral interviews the visual and audio cues do not contributed to the validity of the interview responses. Raters provided similar ratings to transcripts and audio tapes. These findings, however, provide no insight into how candidates may have performed differently under a social versus an asocial (i.e., written test) setting. Therefore, we cannot infer from this study that social processes do not impact the validity of the structured behavioral interviews.

I contend that the employment interview is a social interaction process. Even if we eliminate interviewers' biases and reduce the influence of other situational variables (e.g., ambience, the number of interviewers on the panel and order of questions) through standardization and structure, the candidates may respond to these conditions differently depending upon individual differences. Thus, it is very possible that the responses elicited in a face-to-face interview still remain susceptible to the social interaction process. If this is the case, then the responses elicited from the candidates in an oral interview condition and those in the written interview condition should be different and should result in different

assessments of a candidate. The comparison of oral and written interview can address the question concerning whether structured behavioral interviews only draw upon past behaviors or also utilize the ability of the candidate to effectively narrate these behaviors in a face-to-face interview. My study will examine the aforementioned issues and clarify why structured behavioral interviews predict performance.

The similarity of responses in an oral and a written interview condition depends upon two necessary conditions. First, it requires that the candidates should provide similar responses in both conditions. In other words, the content or the essence of the responses in the social discourse and in the written interview condition should be the same. Second, the rater's assessment of the responses should be similar in both conditions. In other words, the second condition requires that provided the content is similar in the two interview conditions, than the context (e.g., absence or presence of visual and oral cues, ambience and noise) should not influence the rating of the responses. Regarding the latter factor, Motowidlo et al. (1992) showed that in a structured behavioral interview judgmental accuracy is related to amount of relevant behavioral content and that the absence of nonverbal cues does not impede this validity. However, in this study, the audio interviews and transcripts were related to responses elicited during the social interaction process. Thus, it leaves unanswered the question regarding the first condition i.e., whether the content of the responses should be the same in an oral and a written condition.

In a given structured interview, the external conditions such as the type of interviewer (e.g., a manager or a trained psychologist), interview environment, and implementation of interview (e.g., panel interview or one-to-one interview) are kept constant. Hence, I will only discuss the candidate-related factors that will influence the candidate's behavior and responses. In the selection interview literature, much work has been done in investigating the influence of interviewers' judgments and interview design and content on the validity of interviews, whereas, the influence of candidate-related factors has been largely ignored (c.f. Cook, Vance, & Spector, 2000). There is scant research that addresses how candidates' personality might influence the outcomes of structured employment interviews. In this way, the current study makes an important contribution by directly exploring the influence of personality on the construct validity of structured interviews.

The candidate-related factors are either relevant to job performance and, thus, are relevant to the validity criterion, or are irrelevant (either through criterion contamination or criterion deficiency) to job performance and, thus, negatively influence the validity of the interviews. If face-to-face and written interviews differ in terms of validity then we can argue that content alone is not sufficient for validity of interviews and that social interaction processes potently influence the validity of interviews despite structuring. The question is whether social interaction is necessary for the validity of structured interview or can equally valid behavioral judgments be made from information gained through an interview methodology that lacks social interaction.

The role of personality has been elucidated in interviews (e.g., Huffcutt, Roth, Conway, & Stone, 2001a; Salgado & Moscoso, 2002; Schuler & Funke, 1989). Dipboye and Gaugler (1993) presented an information-processing model of employment interviews and argued that a candidate's personality is a key factor influencing interview outcomes. Hufcutt et al. (2001a) meta-analyzed fifty interview studies and showed that personality explains criterion-validity of interview scores. Another study showed that the scores of structured behavioral interviews for higher-level managerial positions are significantly correlated with extraversion (0.26) and conscientiousness (0.22) (Krajewski, Goffin, McCarthy, Rothstein, & Johnston, 2006). Similarly, Huffcutt et al. (2001b) found that extraversion is correlated (0.3) with the structured PBDI scores of district managers. These studies show that the structured behavioral interviews utilize various personality dimensions. Therefore, I argue that performance on a structured behavioral interview is influenced by the personality characteristics of the candidate.

The influence of personality on the validity of the interview is partly due to the relationship between personality and job dimensions. For example, in Krajewski et al.'s (2001b) study, one of the interview dimensions was organizing and planning behavior, for which conscientiousness is a distal predictor. However, personality also influences interview validity due to candidate's ability to perform well in the interview. The interview is a social interaction and personality influences the way we interact with each other (Berry & Hansen, 2000; Mount,

Barrick, & Stewart, 1998). Thus, it is reasonable to assume that certain personality traits will influence candidates' coping with the interview situation, his or her level of inhibition during the interview, the quality of the information presented in the interview, and the impression management skills of the candidate. More specifically, the traits of extraversion and emotional stability will influence the content of responses provided in the structured interview.

Extraverted individuals are talkative and outgoing social beings who need external stimulation and who appreciate social contact with others (Goldberg et al., 2006). Thus, it is reasonable to argue that extraverts will be comfortable in the employment interview situation and will be stimulated to participate in the social discourse aspect of the interview. On the other hand, a person low on extraversion has an internal focus and is characterized as being shy and taciturn (Goldberg, et al., 2006). Thus, it is very likely that introverts fail to provide sufficient information in the structured behavioral interviews (where interviewers often do not ask probing questions). This may truncate the accurate sampling of past behaviors. On the other hand, these individuals might provide a more realistic sample of their past behaviors in a written test – hereafter referred to as patterned behavior description test (PBDT). Likewise, an extrovert individual might not be stimulated to perform well in an environment that lacks social interaction such as a paper-pencil test.

Thus, I argue that depending on the personality trait of extraversion, a candidate will score differently in the oral interview and the written tests. These differences will also impact the criterion-related validity of interview scores. As

extroverts are at their best in social situations, I argue that for extroverts a structured PBDI will provide maximum information and for introverts PBDT will be the optimal form of information acquisition. Thus, I propose the following hypotheses:

> *Hypothesis 1a: For an extrovert the PBDI score is significantly higher than the PBDT score.*
>
> *Hypothesis 1b: For an introvert the PBDT score is significantly higher than the PBDI score.*
>
> *Hypothesis 1c: The PBDI scores are positively related to future performance. Specifically, the relation between PBDI scores and performance will be more positive for extroverts than for introverts.*
>
> *Hypothesis 1d: The PBDT scores are positively related to future performance. Specifically, the relation between PBDT scores and performance will be more positive for introverts than for extroverts.*

The personality dimension of emotional stability captures the disposition of people to be nervous. The role of social anxiety is well-captured in the employment interview situations (e.g., Caldwell & Burger, 1998; Cook, et al., 2000). The neurotic job candidates should be more nervous in a face-to-face interview than in a non-interactive written test. It can be argued that a certain level of anxiety should be present in a paper-pencil evaluation. However, in a face-to-face interview, the social interface and apprehension of evaluation by the person sitting in front of a candidate exacerbates both anxiety and nervousness. Hence, a neurotic individual should be more apprehensive in a face-to-face interview than in a paper-pencil test format. Neurotic individuals may be more inhibited in the interview resulting in the inadequate presentation of their qualifications in an interview. Research has

demonstrated that experience can serve to overcome nervousness in social interactions (Allen, Hunter, & Donohue, 1989). Thus, the effects of emotional stability should be weaker for candidates who have more experience in performing in a selection interview.

Hypothesis 2a: For a candidate low in emotional stability, the PBDT score is significantly higher than the PBDI score.

Hypothesis 2b: The effect of emotional stability on interview scores will be lower for candidates with more interview experience.

Hypothesis 2c: The PBDI scores are positively related to future performance. Specifically, the relation between PBDI scores and performance will be more positive for individuals with higher emotional stability than for individuals with lower emotional stability.

Core self-evaluations can be viewed as a meta-personality trait (Judge, Locke, Durham, & Kluger, 1998) comprising the traits of emotional stability, self-efficacy, self-esteem, and locus of control. Some researchers have argued that candidate's self-esteem (Liden, Martin, & Parsons, 1993), self-efficacy and locus of control (Cook, et al., 2000) influence interviewer judgments. Earlier, I argued how neurotic candidates may perform differently in oral interviews. The construct of core self-evaluations provides a parsimonious way to investigate the impact of these traits on interview validity. Core self-evaluations have been shown to be related to various job-relevant constructs such as motivation, goal attainment, job performance, and job satisfaction (Erez & Judge, 2001; Judge, Bono, Erez, & Locke, 2005; Judge, et al., 1998; Kacmar, Collins, Harris, & Judge, 2009). However, the influences of this meta-trait have not been directly investigated in

employment interviews. This is surprising, as employment interviews represent a situation where many factors relevant to core self-evaluation (e.g., motivation to succeed, ability to perform, and achievement focus) are salient. Moreover, as mentioned earlier, the role of the individual constructs constituting core self-evaluation has been shown to be relevant in employment interviews.

The broad trait of core self-evaluation is related to motivation and performance (Erez & Judge, 2001). Overall, job candidates with higher core self-evaluations will have confidence, self-assurance, psychological resources, and determination to effectively communicate in the social interaction. These individuals will be able to present their qualifications effectively and convincingly without fear of evaluations. For example, Liden et al. 1993 showed through a laboratory experiment that individuals with high self-esteem (a key component of core self-evaluations) are not influenced by the indifferent behaviors of interviewer, whereas, candidates with low self-esteem are negatively influenced by the detached behaviors of an interviewer. It is possible that individuals with lower core self-evaluations are less confident and feel less secure in a face-to-face interview situation. I argue that individuals with lower core self-evaluations should perform relatively better in the written version of the interview where they will not be required to interact with an interviewer. This lack of social interaction liberates their psychological resources so that they can focus upon the content of the questions, thus resulting in relatively better performance. Consequently;

> *Hypothesis 3a: For a candidate with low core self-evaluations, the PBDT score is significantly higher than the PBDI score.*

Hypothesis 3b: For a candidate with high core self-evaluations, the PBDI score is significantly higher than the PBDT score.

Hypothesis 3c: The PBDI scores are positively related to future performance. Specifically, the relation between PBDI scores and performance will be more positive for individuals with higher core self-evaluations.

Methods

Participants

The interviewees for this study were potential associates for a selective leadership training program of a business school in a large mid-western University. Out of 79 interviewees 57 interviewees agreed to participate in this research study. A one-way between subjects ANOVA was conducted to compare the differences between those who agreed to be research-subjects and those who would not consent to participate (non-research-subjects) in terms of numerous characteristics. There were no significant differences between the research and the non-research candidates in terms of resume quality [F (1, 77) = 1.1, p = .298], essay writing skills [F (1, 77) = .08, p = .782] and recommendation letter scores [F (1, 77) = .05, p = .819]

The typical interviewee in the study was 20 years old. Except for one interviewee, each participant had been through at least one formal selection interview (M = 4.2 interviews). Each fall the leadership program selects aspiring business associates. The leadership program conducts around 80 interviews and selects 50 to 60 qualified associates. Once selected, the associates participate in

various developmental programs throughout their tenure in the program. The leadership associates must complete an assortment of obligations each semester which are evaluated on a point system. Typically, an associate commits 30-35 hours to the leadership program per semester. These activities include attending leadership team meetings, participating in panel discussions with professionals, attending large seminar-style events, business lunches, and workshops, performing community service, attending professional development workshops, and serving in a leadership role in at least one of the projects organized by the training program. For continuing their enrollment in the program, the associates are expected to meet the minimum program requirements; such maintaining a minimum grade point average of 3.3 and observing minimum attendance requirements in the program. The highest performing associates are also offered the opportunity to serve on the student advisory board.

Interview Development

I utilized critical incident analysis in order to identify key behaviors that the successful associates demonstrate in the focal training program in this research. The input for job analysis was provided by eight former and current leadership program associates through a written survey. These associates provided approximately 50 critical incidents describing training performance. In addition, several face-to-face meetings were conducted with the program director and associated program faculty to understand the behaviors of successful associates in the training program. To further understand the desirable training behaviors, program brochures were also

examined. On the basis of job analysis, I identified three key performance dimensions. These dimensions were: organizing and planning behavior, proactive change and proactive self-development behavior.

I developed interview questions about behaviors in past situations that might draw on activities reflecting these dimensions. Initially twelve questions were developed. These questions were reviewed independently by five subject matter experts. These experts were not aware of the underlying dimensions intended to be assessed by these questions. The subject matter experts requested to provide feedback on the face validity of these questions for a selection interview. The questions where subject matter experts did not agree on the underlying dimensions were dropped. The final version of the interview included five structured, behavioral questions (See Table 1).

Insert Table 1 about here

I also developed an interview rating scale for each dimension based on job analysis information and review of extant literature. I reviewed the extant literature and looked at the existing measurement scales for designing interview rating scales that not only job relevant but are theoretically sound as well. Planning and organizing behaviors were defined as a set of activities that shows planning ahead and working in a systematic and organized way. Such behaviors provide evidence of thinking ahead, planning of time and resources, establishing rules and norms to

organize work, prioritizing of schedule and monitoring of plans (Bartram, 2005; Goldberg, et al., 2006).

The proactive self-development behavior was defined as a set of activities that shows a focus towards evaluating oneself, setting standards, noticing discrepancies, and enthusiastically working towards reducing these discrepancies while managing negative emotions such as pessimism or dislike of those who criticize (Ashford & Tsui, 1991; Frayne & Geringer, 2000; Manz & Sims, 1980; Porath & Bateman, 2006; Ryan & Deci, 2000). Proactive change was defined as a set of activities that shows a focus towards scanning the environment for opportunities to improve, showing initiative, taking action, and persevering until one reaches closure by bringing about the change (Crant, 2000; Crant & Bateman, 1993; Seibert, Kraimer, & Crant, 2001). (See Annexure 1 for the response assessment key for the behavioral interview questions).

Participants

The interviewees for this study were potential associates for a selective leadership training program of a business school in a large mid-western University. Out of 79 interviewees 57 interviewees agreed to participate in this research study. A one-way between subjects ANOVA was conducted to compare the differences between those who agreed to be research-subjects and those who would not consent to participate (non-research-subjects) in terms of numerous characteristics. There were no significant differences between the research and the non-research candidates in terms of resume quality [F (1, 77) = 1.1, p = .298], essay writing

skills [F (1, 77) = .08, p = .782] and recommendation letter scores [F (1, 77) = .05, p = .819]

The typical interviewee in the study was 20 years old. Except for one interviewee, each participant had been through at least one formal selection interview (M = 4.2 interviews). Each fall the leadership program selects aspiring business associates. The leadership program conducts around 80 interviews and selects 50 to 60 qualified associates. Once selected, the associates participate in various developmental programs throughout their tenure in the program. The leadership associates must complete an assortment of obligations each semester which are evaluated on a point system. Typically, an associate commits 30-35 hours to the leadership program per semester. These activities include attending leadership team meetings, participating in panel discussions with professionals, attending large seminar-style events, business lunches, and workshops, performing community service, attending professional development workshops, and serving in a leadership role in at least one of the projects organized by the training program. For continuing their enrollment in the program, the associates are expected to meet the minimum program requirements; such maintaining a minimum grade point average of 3.3 and observing minimum attendance requirements in the program. The highest performing associates are also offered the opportunity to serve on the student advisory board.

Procedures

The interviewees first appeared in a face-to-face selection interview. After finishing the oral interview candidates were guided to a computer laboratory where they participated in a parallel online written interview. As these were real selection interviews, to avoid any adverse effect of interview ordering, all interviews were administered in the same order i.e., the oral followed by the written interview.

The face-to-face interviews were administered by a three-member panel. The panel included the leadership program director, a faculty member, and a representative of the student advisory board. There were a total of seven faculty members and seven student members that were available to serve as interviewers. The faculty member and the student member could not remain constant on all the interviews due to intensive time commitments – the interviews were scheduled to last three full days. However, the program director was present in all interviews. The instructions on the standardized format of the interview and the behavioral rating keys were provided a week in advance to all interviewers. The researcher was available to answer any queries. The program director was trained by the researchers on how to conduct past behavioral description interviews through several one-on-one meetings.

The first half of the interview was designed to be a structured behavioral interview. In this part, the program director introduced her, the two other panel members, and the researcher to observe candidate's impression management skills. This was followed by administering five structured behavioral questions. The

standard questions were asked in the same order as shown in Table 1. The third and fourth questions were asked by the student interviewer. The remaining three were asked by the program director. No probing questions were permitted. The second half of the interview was less structured where interviewers could ask any question that they deemed suitable.

After dismissing the candidate, the interviewers rated the candidate on all the interview questions (five structured behavioral questions as well as questions asked in the second half of the interview session). The average rating of all three interviewers was assigned to the candidate. The candidates were selected in the program based on their combined scores on essay writing skills, resume, letter of references and interview performance. However, the ratings assigned by interviewers were not used for analysis. All interviews were audio taped and two trained researchers listened to all the audiotapes for rating the interviews. These researchers did not serve as interviewers.

Measures

Personality. Personality was measured using the 50-item IPIP scale for the Big-5 personality factors (Goldberg, et al., 2006). Each subscale consists of 10 items rated on a 5-point scale that ranges from strongly disagree (1) to strongly agree (5). The reliability estimates for the five factors were as follows: emotional stability (α = .80), extraversion (α = .85), conscientiousness (α = .85), agreeableness (α = .63), and openness (α = .74)[1].

[1] For the 10-item openness to experience scale the reliability estimate was .58. These items were further analyzed by plotting a graph between average score on all the items and the score on an

Core self-evaluations. Consistent with previous research (e.g., Judge, et al., 2005; Judge, et al., 1998), core self-evaluations were measured with four sub-scales. The first dimension of locus of control was measured using an 11-item scale (Ghorpade, Hattrup, & Lackritz, 1999). A sample item includes "I think that life is mostly a gamble." The responses were based on a 1 (strongly disagree) to 7 (strongly agree) scale (α = .78). The second dimension of emotional stability was measured with the 10-item subscale from the IPIP scale for measuring the Big-5 personality factors (Goldberg, et al., 2006). Responses were based on a 1 (strongly disagree) to 5 (strongly agree) scale (α = .80). An example of the items is "I often feel blue." The 10-item scale by Rosenberg (1965) was used to measure the third dimension of self-esteem. Responses were based on a 1 (strongly disagree) to 5 (strongly agree) scale (α = .83). "On the whole, I am satisfied with myself" is one of the items in the scale. The fourth dimension of generalized self-efficacy was measured with an 8-item scale proposed by Chen, Gully, and Eden (2001). Responses were based on a 1 (strongly disagree) to 5 (strongly agree) scale (α = .93). A sample item includes "Even when things are tough, I can perform quite well." Consistent with prior research (e.g., Erez & Judge, 2001; Judge, et al., 2005), core self-evaluations were treated as a higher order concept.

---

individual item. These plots indicated five items with outliers such that several subjects' scores on these items deviated from their average scores on all the items. It is possible that associates viewed some of the items socially desirable as an applicant for a leadership program. These items were related to abstract and novel ideas e.g., "Enjoy hearing new ideas." Thus, all the subsequent analyses were done by dropping these five items from the scale measuring openness to experience. A similar analysis on agreeableness was done with no clear evidence of items indicating outliers.

To investigate the validity and structure of the core self-evaluations concept, first principal component analysis was done using SAS 9.2 followed by a confirmatory factor analysis. The items within each four scales were averaged to form a single score for emotional stability, locus of control, self-efficacy, and self-esteem. Due to small sample size, the principal component analysis was done on the average scores of the four component dimensions. The principal component analysis identified a single factor solution. The one-factor solution was supported by the subsequent confirmatory factor analysis. The fit statistics for this model were as follows: $\chi 2$ (2, N = 57) = 5.93, ns; *RMR* = .07 *RMSEA* = .18; *GFI* = .95; *BFI* = .95; and *NFI* = .92. These results were used to calculate a factor score of core self-evaluations for each subject; these factors were used in the subsequent analysis.

Impression Management. A trained researcher served as an observer to measure candidates' impression management in the interviews. A 14-item scale developed by Stevens and Kristof (1995) was used to assess interviewee's impression management. The internal consistency reliability estimates were acceptable for the impression-management scale: overall 14-item scale, α = .88; five items measuring self-promotion, α = .93; four items measuring other enhancement, α = .72; three items measuring organization fit, α = .91; and two items measuring nonverbal impression management, α = .64. For each scale, the mean ratings across items were computed for use in the subsequent analyses.

Face-to-face structured past behavioral interview questions. Two researchers independently listened to the audio tapes of all the candidates. Ratings were assigned to all candidates by listening to one question at a time. These ratings were assigned according to a predefined standard behavioral rating key (See Annexure 1). In instances, where the ratings of the two raters differed by more than two points, the two raters explained their ratings to each other. After discussion, each rater independently assigned a new rating to the candidate for that question. Overall interview score was calculated by taking an average of the ratings of the five questions. The inter-rater reliability estimates for the interview questions were greater than the acceptable limit of 0.7.

Insert Table 2 about here

Written structured past behavioral interview questions. The researchers who reviewed the audio tapes also rated the written interviews using the same standard behavioral rating keys. Ratings were assigned to all candidates by reviewing one question at a time. An overall interview score was calculated by taking an average of the five interview questions. In instances, where the ratings of the two raters differed by more than two points, the two raters explained their ratings to each other. After discussion, each rater independently assigned a new

rating to the candidate for that question. The inter-rater reliability estimates for the interview questions were greater than the acceptable limit of 0.7.

Training performance. The performance of selected candidates was measured by several methods. One indicator of performance was frequency of participation in different programs. The participation scores were measured at two different points in time: three months after the enrollment and six months after the enrollment in the program. The program director also provided assessment of selected candidates on a 6-item training performance scale measuring proactive behaviors (4-items) and organizing and planning behavior (2-items). The four items for proactive behavior included the following: this person assumes leadership roles**;** this person takes initiatives that add value to the program; this person enthusiastically makes use of the self-development opportunities provided by the program activities; this person often consults you to seek performance feedback. The organizing and planning behavior was measured by the following two items "this person is punctual, regular, and always prepared" and "this person is committed and fulfills his/her tasks and duties." The internal consistency reliability were acceptable for this scale $\alpha = .90$. The program director also provided a ranking of all the associates in the following five performance quartiles: eightieth percentile and above, sixtieth percentile and above, fortieth percentile and above, twentieth percentile and above and below twentieth percentile.

To investigate the structure of the construct training performance, first principal factor analysis was done followed by a confirmatory factor analysis. The

principal factor analysis identified a single factor solution. One-factor solution was also supported by a subsequent confirmatory factor analysis. The fit statistics for this model were as follows: $\chi 2$ (5, N = 38) = 3.45, p=0.63; *RMR* = .03 *RMSEA* = .00; *GFI* = .96; and *NFI* = .96. These results were used to calculate a factor score of training performance for each subject. A one-way between subjects ANOVA was conducted to compare the effect of trainee type on training performance in research-subjects and non-research-subjects conditions. There was no significant effect of trainee type on training performance, F (1, 51) = .82, p = .371.

Others. In addition, I obtained candidates' scores on other selection criteria including essay writing skills, resume, and letters of recommendations. These scores were assigned to the candidates by the selection committee comprising program director and several faculty members. I also obtained information about candidates' grade point average and their scores on standard aptitude tests. An overall variable "application" was defined by averaging the scores across all bio-data variables that capture a candidate's quality and qualification. These variables included the following: grade point average, resume, letter of recommendations, and application essay scores.

Insert Table 3 about here

Results

All candidates first appeared in a face-to-face-structured behavioral interview followed by a structured written interview. As these were real selection interviews, it was necessary to keep the same order (i.e., face-to-face interview followed by the written) for all the candidates so as to avoid any selection biases due to interview order. For testing any order effects, a control past-behavioral question was designed. The half of the candidates was randomly selected to answer this control question in a written interview. All candidates were also asked this question in a subsequent second face-to-face structured interview. The second face-to-face interview was part of a separate research study being conducted on the same day by the researchers. The remaining half of candidates was requested to answer the written control question after the second face-to-face interview. A one-way between subjects ANOVA was conducted to compare the order effects on the interview scores for the control question. Neither the interview scores for the face-to-face question [F (1, 55) = .18, p = .675] nor the scores for the written question [F (1, 55) = .11, p = .747] provided any support for order effects.

Hypotheses 1a-b, 2a-b, and 3a-b included personality as a continuous covariate with oral and written interview as a repeated factor. The SAS "Proc Mixed" approach offers considerable flexibility in testing a wide range of nested models (Singer, 1998). Thus, these three set of hypotheses were tested with SAS Proc Mixed. The interview included five questions – four of these questions measured the dimension of proactive behavior and one question measured

organizing and planning behavior. The internal consistency estimates for all five questions and for the four proactive behavior questions were $\alpha = .53$ and $\alpha = .58$, respectively. Out of the four proactive behavior questions, dropping one question (i.e., the fourth question in Table 1) improved the reliability estimates ($\alpha = .60$)[2]. Thus, for the hypotheses testing, analyses were run on all the five interview questions, the four proactive behavior questions (i.e., by dropping question 3), and the three proactive behavior questions (i.e., by dropping question 3 and question 4).

The test of hypotheses concerning the relative performance of an interview candidate in an oral and a written interview (i.e., hypotheses 1a-b, 2a-b, and 3a-b) provided modest support. I argued in hypothesis 1a and 1b that extraversion will moderate a candidate's relative performance in an oral and a written interview such that extroverts would perform relatively better in an oral interview (hypothesis 1a), and introverts would perform relatively better in a written interview (hypothesis 1b). It is important to mention that written interview scores were assigned on the basis of relevance of behaviors and not on the basis of written communication skills. This is further evident from a weak correlation between ACT English proficiency scores and written interview scores ($r = .16$, $p = .22$). Moreover, using English proficiency scores in the subsequent models as a covariate did not show any significant effects of English writing skills. Thus, to save degrees of freedom this covariate was not used in further analysis.

[2] Alpha scores are sensitive to the number of items in a scale, such that with fewer items alpha scores tends to be lower (Cortina, 1993).

Table 4 shows that the type of an interview has a significant effect for all three dependent variables: for all questions, F (1,55) = 3.87, p=.05; for the four proactive questions F (1,55) = 7.93, p=.01, and for the three proactive questions F (1,55) = 4.63, p=.04. Extraversion did not have a main effect. However, the interaction between extraversion and the type of test was only significant for the interview questions related to proactive behavior: for all questions, F (1,55) = 3.27, p=.08; for the four proactive questions F (1,55) = 7.14, p=.01, and for the three proactive questions F (1,55) = 5.31, p=.03.

Insert Table 4, 5 and 6 about here

Table 5 shows regression coefficients for the three models. The type of an interview was defined as a class variable with the baseline model representing written interviews. The results show that the interaction between extraversion and the type of interview holds only for the interview questions related to proactive behaviors. The parameters for interaction term are as follows: for the four proactive questions, b = .42, t (55) = 2.67, p = .01, and for the three proactive questions b = .43, t (55) = 2.30, p = .03. These interactions were further probed by conducting planned comparisons of predicted interview scores in the oral and written interview conditions at the mean value of extraversion and at one standard deviation above (designated as extroverts) and below (designated as introverts) the mean values. As shown in Table 6, the interactions effects are in the predicted direction. The review

of four proactive interview questions shows that introverts score significantly higher in the written interview than the oral interview, t (55) = -2.7, p<.01. However, for extroverts there was no difference in the performances between the oral and written interview. The examination of the three proactive questions shows that for introverts there was no difference in performance between the oral and the written interview scores t (55) = -1.05, p<.30). However, in this case extroverts scored significantly higher in the oral interview than in the written interview, t (55) = 2.22, p<.03.

It is also important to note when a model was run with impression management as a moderator for the effects of type of interview on the total interview scores, the interaction term between type of interview and impression management was not significant. The parameter estimates for this model were as follows: ; for oral interview b = -.46, t (55) = -.87, p = .39, for impression management b = .34, t (55) = 2.28, p = .02, and for the interaction term between oral interview and impression management b = .11, t (55) = .70, p = .49. This is consistent with past research that shows that structuring reduces the influence of impression management on candidate's evaluation (Barrick, Shaffer, & DeGrassi, 2009).

Insert Figure 1 and Figure 2 about here

In order to further explore the interaction between extraversion and type of an interview, a factor analysis was conducted on the four proactive behavior questions. First, principal component analysis was done on the four interview questions related to proactive behaviors followed by a confirmatory factor analysis. These analyses were done separately for oral and written interview scores. The principal component analysis identified a single factor solution. One-factor solution was supported by the subsequent confirmatory factor analysis. The fit statistics for these models were as follows: oral interview, $\chi 2$ (2, N = 57) = 5.04, p=.08; *SRMR* = .07 *RMSEA* = .16; *GFI* = .96; *BFI* = .88; written interview, $\chi 2$ (2, N = 57) = .90, p=.64; *SRMR* = .03 *RMSEA* = .00; *GFI* = .99; *BFI* = 1. These results were used to calculate factor scores for the oral and the written interview scores related to the proactive behaviors. These factors were used to investigate the impact of extraversion on the type of interview. The results indicated introverts perform significantly better in the written interview, t (55) = -1.98, p<.05 whereas, extroverts perform significantly better in the oral interviews, t (55) = 1.99, p<.05. Overall, these results show that extraversion is associated with the mode of an interview such that a higher degree of extraversion enables a candidate to perform better in an oral interview whereas a lower degree of extraversion enables a candidate to perform better in a written interview. Thus, the study provides clear support for hypothesis 1a and hypothesis 1b.

I stated in hypotheses 2a that a candidate's emotional stability will moderate one's relative performance in an oral and a written interview such that candidates

with lower emotionally stability would perform relatively better in a written interview. Table 4 shows that emotional stability did not have a significant main effect. Moreover, the interaction between emotional stability and a type of interview was also not significant: for all five questions, F (1,55) = 1.08, p=.30; for the four proactive questions F (1,55) = 1.56, p=.22, and for the three proactive questions F (1,55) = 1.48, p=.33. Thus, I did not find support for hypothesis 2a. I also tested the effects of experience in interviewing on these relationships. The number of past interviews did not have any effect on the relationship between emotional stability and one's performance in the interview. Thus, hypothesis 2b was not supported.

I argued in hypothesis 3a and 3b that a candidate's core self-evaluations will moderate one's relative performance in an oral and a written interview such that subjects with high core self-evaluations would perform relatively better in an oral interview (hypothesis 3a), and those with low core self-evaluations would perform relatively better in a written interview (hypothesis 3b). Table 4 shows that neither the type of an interview nor the personality trait of core self-evaluations has a significant main effect. However, the interaction between core self-evaluation and the type of an interview was only significant for the interview questions related to three proactive behavior: for all questions, F (1,55) = 2.78, p=.10; for the four proactive questions F (1,55) = 3.63, p=.06, and for the three proactive questions F (1,55) = 5.53, p=.02. This indicates that for certain constructs (in this case proactive

behaviors) core self-evaluations might be related to a candidate's ability to present an evidence of past behaviors during an interview.

Table 5 shows that the interaction between core self-evaluations and a type of an interview holds only for the three interview questions related to proactive behavior: for all questions, b =.13, t (55) = 1.67, p=.11; for the four proactive questions, b =.16, t (55) = 1.91, p =.06, and for the three proactive questions b = .23, t (55) = 2.35, p =.02. The planned comparisons in Table 7 show that the interaction effects were in the predicted direction when examining performance on the three interview questions. For subjects with low core self-evaluations there was no significant difference in performance between the oral and the written interview, t (55) = -.16, p<.29. However, subjects with high core self-evaluations scored significantly higher in the oral interview than the written interview, t (55) = 2.25, p<.03. Overall, these results show that core self-evaluations are associated with the mode of an interview such that higher degree of core self-evaluations enables one to perform better in an oral interview. Thus, the study provides modest support for hypothesis 3b.

Insert Figure 3 about here

Although, I did not make any specific hypothesis about openness to experience, the data analysis showed some interesting results regarding this trait. Table 4 shows that the type of an interview has a significant effect for all three

dependent variables: for all questions, F (1,55) = 5.22, p=.03; for the four proactive questions F (1,55) = 4.81, p=.03, and for the three proactive questions F (1,55) = 6.32, p=.01. Openness did not have a significant main effect. However, the interaction between openness and the type of an interview was significant; for all questions, F (1,55) = 7.03, p=.01; for the four proactive questions F (1,55) = 6.43, p=.01, and for the three proactive questions F (1,55) = 5.69, p=.02.

Table 5 shows regression coefficients for the three models: for all questions, b =-.27, t (55)=-2.65, p=.01; for the four proactive questions, b =-.29, t (55) = -2.54, p=.01, and for the three proactive questions b =-.33, t (55) = -2.38, p =.02. These interactions were further probed by conducting planned comparisons of predicted interview scores in the oral and the written interview conditions at the mean values of openness and at one standard deviation above and below the mean values.

Insert Table 7 and 8 about here

As shown in Table 8 , at higher levels of openness to experience, subjects scored significantly higher in the written interview than the oral interview: all five questions, t (55) = - 2.74, p<.01); four proactive questions, t (55) = -2.60, p<.01). However, for low levels of openness there was no significant difference in performance between the oral and the written interview scores except for the three proactive interview questions. In this case, low levels of openness were associated

with better performance in an oral interview (t (55) = 2.28, p<.03), and high levels of openness did not have significant interaction with the type of interview, t (55) = -1.11, p<.27. Overall, reviewing the parameter estimates and effect sizes we can conclude that openness to experience interacts with the mode of an interview to influence an individual's performance in an interview.

Insert Figures 4-6 about here

The next set of hypotheses (hypotheses 1c-d, hypothesis 2c, and hypothesis 3c) was related to the criterion-related validity of the two modes of interview. To test hypotheses 1c and 1d, a regression model was defined with training performance as a dependent variable and oral and written interview scores as independent variables. To test these hypotheses, the overall interview score, the interview score of four proactive questions, the interview score of three proactive questions, and the interview score of the organizing and planning question (i.e., question 3) were used. These combinations resulted in four different models. In these models, training performance was defined as the factor score of training performance extracted through confirmatory factor analysis.

Table 9 shows overall results for these four models. The total interview scores didn't explain significant variance in the training performance, $R^{2\ =}$ .27 , F (1,37) = 2.35, p=.06, η^2 =.27, ω^2 = .15, 95% CI [.00, .40]. However, the interview questions related to proactive behaviors explained significant variance in the

training performance score: $R^{2} = .32$, F (1,37) = 3.08, p=.02, η^2 =.32, ω^2 = .21, 95% CI [.00, .46]. The results of these models (see Tables 10 and 11) show that extraversion (β = 1.74, t (32) = 2.38, p = .02, , η^2 =.11, ω^2 = .07, 95% CI [.00, .30]) and oral interview (β = 3.42, t(32) = 2.91, p = .01, , η^2 =.15, ω^2 = .11, 95% CI [.00, .34]) have significant main effects. The interaction term between oral interview and extraversion was also significant, β = -1.22, t (32) = -2.78, p = .01, η^2 =.14, ω^2 = .10, 95% CI [.00, .33]. I followed Cohen's (1988) guidelines regarding η^2 and ω^2 effect size estimates (i.e., small =. 0.0099, medium = 0.0588, large = .1379). The models 1 and 3 show large effect.

Insert Table 9, 10, and 11 about here

To further probe the significant interaction in model 2, I followed the method recommended by Aiken and West (1991) for analyzing interactions between continuous predictors. First regression lines were drawn for the model. The next step was to test the slopes of these regression lines. For model 2, the standard errors of the simple slopes were calculated to conduct t-tests for the slopes. The t-tests of each simple slope against zero were as follows: low extraversion, t (32) = 2.82, p =.008; medium extraversion, t (32) = 1.60, p =.12; high extraversion, t (32) = -0.87, p =.39. These results show that the relationship between training performance and oral interviews holds only for low extraversion. In other words, oral interview predicts performance for individuals low in

extroversion. Thus, hypothesis 1c is supported in opposite direction i.e., higher the score of introverts in an oral interview higher the future performance. The study did not show a support for hypothesis 1d.

Insert Figure 7 about here

To test hypotheses 2c and 3c, I proceeded as before by defining four different models for each hypothesis using as predictors the overall interview score, the interview score for the four proactive questions, the interview scores for the three proactive questions, and the scores for the organizing and planning question. Table 12 to 14 shows overall results for these hypotheses. None of these models explain significant variance in the training performance. However, due to sample size limitations, I also reviewed the confidence intervals for the η^2 and ω^2 effect sizes provided for the overall ANOVA. I followed Cohen's (1988) guidelines regarding η^2 and ω^2 effect size estimates (i.e., small =. 0.0099, medium = 0.0588, large = .1379). The upper limits for the confidence intervals show a small effect (e.g., model 1: 95% CI [.00, .04]) to a medium effect (e.g., model 4: 95% CI [.00, .04]) for emotional stability; a large effect for core self-evaluation (e.g., model 4: 95% CI [.00, .22]) and a medium effect (e.g., model 2: 95% CI [.00, .12]) to a large effect for openness to experience (e.g., model 4: 95% CI [.00, .24]). This study does not provide support for hypothesis 2c and 3c. However, effect size estimates

for core self-evaluations and openness to experience are encouraging enough to warrant future investigations.

Insert Table 12, 13 and 14 about here

Discussion

This field study investigated the role of candidate's personality in determining the validity of structured patterned behavior interviews (PBDI). The structured patterned behavior interviews are designed to tap into job-relevant patterns of past behavior by minimizing the influence of interviewer-related biases in the interview process. However, structuring does not limit the role of candidate's personality in providing job-relevant information. In the first set of hypotheses (1a, 1b, 2a, 2b, 3a and 3b), I argued that structured patterned behavior interviews do not simply tap into samples of past behavior, instead interviews tap into a candidate's ability to present past behaviors in the selection interview. Thus, an introvert should find it easier to provide job-relevant information in a structured written interview than in a face-to-face interview. The study provided some support for these assertions. More specifically, hypothesis 1a and 1b were supported confirming that an introvert performs relatively better in a written interview than in an oral interview and vice versa. The role of emotional stability (hypotheses 2a and 2b) was not supported, however, hypothesis 3b was supported showing that candidates with high core self-evaluations fair better in an oral interview.

For testing these hypotheses, interview scores were calculated for overall interview as well as for various combinations of questions. I also conducted question-wise analysis that is not shown in this study for the sake of brevity. The effect sizes in these tests varied depending on the combinations of questions used in the analysis. This is not surprising as all questions were designed to draw off of different constructs. In addition, the two questions which are intended to measure the same construct (e.g., question 2 and question 4 were measuring proactive help seeking) may differ in the ability to successfully tap into the desired construct. Some questions might simply not be very effective. Thus, it is reasonable to conduct the analysis by using different combinations of questions.

The next set of hypotheses (1c. 1d, 2c and 3c) tested whether personality influences the criterion-related validity of structured pattern behavior interviews. Hypothesis 1c was supported but in the opposite direction. Contrary to the prediction, the results for hypothesis 1c showed that oral interviews predict performance but only for introverts. The written interview did not predict performance. Thus, hypothesis 1d was not supported. It seems that in the context of this training program, an oral interview only has validity for introverted candidates. Probably, introverts who perform well in the oral interview by overcoming their dispositional inhibitions also have the personal resources to do better in the training program. The slopes of lines for moderate levels and high levels of extraversion were not significant for the oral interview. In addition, written interviews did not

predict future performance. This indicates that the performance of extraverts in the interview is less relevant to the future performance.

The remaining hypotheses for criterion-related validity were not supported. However, the effect sizes for core self-evaluations (hypothesis 3c) were more encouraging than those for emotional stability (hypothesis 2c). It is possible that these effects were not detected by the analysis due to weak power. Thus, large effect size estimates for core self-evaluations provide encouragement to explore these relationships in future studies with more power.

Although, I did not make any specific hypothesis about openness to experience it appears that openness to experience influences the validity of interview process. Individuals with high openness to experience performed relatively better in a written interview than in a face-to-face interview. One reason for this difference might be the fact that written interviews are not very common and candidates generally expect to participate in an oral interview. Thus, candidates with higher levels of openness to experiences responded better to this less common form of interview. If this is true, openness to experience should be relevant in any non-traditional, novel, and less commonly used form of selection interview.

Another reason for these effects can be a link between openness to experience, intellectual abilities, and written communication skills. If this is the case, then openness to experience should only be relevant to the written interview and not to the other verbal but novel forms of communications. Unfortunately, there is not enough work that explores the relationship between openness to

experience and communication skills. In relation to criterion validity, the effect sizes for the type of interview and openness to experience ranged from moderate to large, thus, indicating possible influence on the criterion-related validity of selection interviews. In future studies, the relationship between openness to experience and different forms of interviews should be explored in larger sample sizes.

Overall, this study has made significant contribution to the selection interview literature. This study shows that face-to-face structured behavioral interviews are influenced by personality of the candidates. Moreover, the criterion-related validity of interviews differs depending on the personality of candidates. These findings on one hand highlight that face-to-face interviews are more relevant for individuals with certain personality traits, and on the other hand, urge researchers to investigate what other forms of selection tools should be more appropriate for those with other personality traits. I would conclude that the best predictor of future behavior is past behavior as long as we take into account important personality characteristics when we design a method to assess those past behaviors.

REFERENCES

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park: Sage Publications.

Allen, M., Hunter, J. E., & Donohue, W. A. (1989). Meta-analysis of self-report data on the effectiveness of public speaking anxiety treatment techniques. *Communication Education, 38*(1), 54-76.

Ashford, S. J., & Tsui, A. S. (1991). Self-regulation for managerial effectiveness: the role of active feedback seeking. *Academy of Management Journal, 34*(2), 251-280.

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

Bartram, D. (2005). The Great Eight Competencies: A Criterion-Centric Approach to Validation. *Journal of Applied Psychology, 90*(6), 1185-1203.

Berry, D. S., & Hansen, J. S. (2000). Personality, nonverbal behavior, and interaction quality in female dyads. *Personality and Social Psychology Bulletin, 26*(3), 278.

Caldwell, D. F., & Burger, J. M. (1998). Personality characteristics of job applicants and success in screening interviews. *Personnel Psychology, 51*(1), 119-136.

Campion, M. A., Campion, J. E., & Hudson, J. P. (1994). Structured interviewing: A note on incremental validity and alternative question types. *Journal of Applied Psychology, 79*(6), 998-1102.

Chen, G., Gully, S. M., & Eden, D. (2001). Validation of a New General Self-Efficacy Scale. *Organizational Research Methods, 4*(1).

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, N.J.: L. Erlbaum Associates.

Cook, K. W., Vance, C. A., & Spector, P. E. (2000). The relation of candidate personality with selection-interview outcomes. *Journal of Applied Social Psychology, 30*(4), 867-885.

Cortina, J. M. (1993). What is Coefficient Alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98-104.

Crant, J. M. (2000). Proactive behavior in organizations. *Journal of Management, 26*(3), 435-462.

Crant, J. M., & Bateman, T. S. (1993). Assignment of credit and blame for performance outcomes. *Academy of Management Journal, 36*(1), 7-27.

Dipboye, R. L., & Gaugler, B. B. (1993). Cognitive and behavioral processes in the selection interview. In N. Schmitt & W. S. Borman (Eds.), *Personnel selection in organizations* (pp. 135-170). CA: Jossey-Bass.: San Francisco.

Erez, A., & Judge, T. A. (2001). Relationship of Core Self-Evaluations to Goal Setting, Motivation, and Performance. *Journal of Applied Psychology, 86*(6), 1270-1279.

Frayne, C. A., & Geringer, J. M. (2000). Self-management training for improving job performance: A field experiment involving salespeople. *Journal of Applied Psychology, 85*(3), 361-372.

Ghorpade, J., Hattrup, K., & Lackritz, J. R. (1999). The Use of Personality Measures in Cross-Cultural Research: A Test of Three Personality Scales Across Two Countries. *Journal of Applied Psychology, 84*(5), 670-679.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality, 40*(1), 84-96.

Harel, G. H., Arditi-Vogel, A., & Janz, T. (2003). Comparing the validity and utility of behavior description interview versus assessment center ratings. *Journal of Managerial Psychology, 18*(2), 94-104.

Huffcutt, A. I., Roth, P. L., Conway, J. M., & Stone, N. J. (2001a). Identification and Meta-Analytic Assessment of Psychological Constructs Measured in Employment Interviews. *Journal of Applied Psychology, 86*(5), 897-913.

Huffcutt, A. I., Weekley, J. A., Wiesner, W. H., Groot, T. G., & Jones, C. (2001b). Comparison of situational and behavior description interview questions for higher-level positions. *Personnel Psychology, 54*(3), 619-644.

Janz, T. (1982). Initial comparisons of patterned behavior description interviews versus unstructured interviews. *Journal of Applied Psychology, 67*(5), 577-580.

Judge, T. A., Bono, J. E., Erez, A., & Locke, E. A. (2005). Core Self-Evaluations and Job and Life Satisfaction: The Role of Self-Concordance and Goal Attainment. *Journal of Applied Psychology, 90*(2), 257-268.

Judge, T. A., Locke, E. A., Durham, C. C., & Kluger, A. N. (1998). Dispositional Effects on Job and Life Satisfaction: The Role of Core Evaluations. *Journal of Applied Psychology, 83*(1), 17-34.

Kacmar, K. M., Collins, B. J., Harris, K. J., & Judge, T. A. (2009). Core Self-Evaluations and Job Performance: The Role of the Perceived Work Environment. *Journal of Applied Psychology, 94*(6), 1572-1580.

Krajewski, H. T., Goffin, R. D., McCarthy, J. M., Rothstein, M. G., & Johnston, N. (2006). Comparing the validity of structured interviews for managerial-level employees: Should we look to the past or focus on the future? *Journal of Occupational & Organizational Psychology, 79*(3), 411-432.

Liden, R. C., Martin, C. L., & Parsons, C. K. (1993). Interviewer and applicant behaviors in employment interviews. *Academy of Management Journal, 36*(2), 372-386.

Manz, C. C., & Sims, H. P. J. (1980). Self-Management as a Substitute for Leadership: A Social Learning Theory Perspective. [Article]. *Academy of Management Review, 5*(3), 361-368.

Motowidlo, S. J., Carter, G. W., Dunnette, M. D., Tippins, N., Werner, S., Burnett, J. R., & Vaughan, M. J. (1992). Studies of the Structured Behavioral Interview. *Journal of Applied Psychology, 77*(5), 571-587.

Mount, M. K., Barrick, M. R., & Stewart, G. L. (1998). Five-factor model of personality and performance in jobs involving interpersonal interactions. *Human Performance, 11*(2), 145-165.

Porath, C. L., & Bateman, T. S. (2006). Self-Regulation: From Goal Orientation to Job Performance. *Journal of Applied Psychology, 91*(1), 185-192.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68-78.

Salgado, J. F., & Moscoso, S. (2002). Comprehensive meta-analysis of the construct validity of the employment interview. *European Journal of Work & Organizational Psychology, 11*(3), 299-324.

Santayana, G. (1905). *The Life of Reason: Introduction and Reason in common sense* (Vol. 1). New York: Charles Scribner's Sons.

Schuler, H., & Funke, U. (1989). The interview as a multimodal procedure. In R. W. Eder & G. R. Ferris (Eds.), *The employment interview: Theory, research, and practice* (pp. 183-192). Newbury Park, CA: Sage Publications.

Seibert, S. E., Kraimer, M. L., & Crant, J. M. (2001). What do proactive people do? A longitudinal model linking proactive personality and career success. *Personnel Psychology, 54*(4), 845-874.

Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics, 23*(4), 323-355.

Stevens, C. K., & Kristof, A. L. (1995). Making the right impression: A field study of applicant impression management during job interviews. *Journal of Applied Psychology, 80*(5), 587-606.

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 & Organizational Psychology, 75*(3), 277-294.

ANNEXURES

ESSAY 1: TABLES AND FIGURES

Table 1: The Patterned Behavior Interview Questions

Past behavior Interview question	Underlying training behavior
Question 1. Think about the last five years and tell me what improvements have occurred in you? What specific steps you have taken in the past concerning to bring about these self-improvements? What results have come from these self-improvements? (Asked by the program director)	Proactive self-development behavior (Proactive change - internal)
Question 2. Tell me about your attempts to seek out a mentor in the past. Is there someone who you see as a mentor in your academic life? How did you identify this person and how do you interact with this person (Asked by the program director)	Proactive self-development behavior (Proactive help-seeking)
Question 3. Tell me about a busy week at school when you have multiple deadlines to meet or have multiple exams. How do you plan your activities for that week? (Asked by a student)	Organizing and planning behavior
Question 4. What was the most difficult subject that you ever had to learn in your academic life? Were you able to improve your knowledge of this subject? If yes, how; If not, why? (Asked by the program director)	Proactive self-development behavior (Proactive help-seeking)
Question 5. Tell me about a time when you were involved in a project that resulted in a positive change. How did you facilitate the change? What actions did you take to overcome the challenges along the way? (Asked by the program director)	Proactive behavior (Proactive change - external)

Table 2: Inter-Rater Agreement Scores for the Oral and the Written Interviews

	Face-to-face Interview					
Question	1	2	3	4	5	Total
ICC (A,1)	0.88	0.86	0.87	0.91	0.91	0.96
ICC(A,K)	0.94	0.92	0.93	0.96	0.95	0.98
Rwg	0.91	0.88	0.89	0.96	0.88	0.88

	Written Interview					
Question	1	2	3	4	5	Total
ICC (A,1)	0.91	0.85	0.88	0.87	0.84	0.90
ICC(A,K)	0.95	0.92	0.93	0.93	0.91	0.95
Rwg	0.95	0.90	0.91	0.92	0.86	0.98

Table 3: Mean Values, Standard Deviations, and Intercorrelations of Study Variables

Variables	*M*	*SD*	1	2	3	4	5	6	7
1. Age	19.66	0.94	1.00						
2. Gender	0.67	0.48	-0.06	1.00					
3. No of Interviews	4.20	2.91	0.25	-0.21	1.00				
4. Application	4.11	0.47	-0.19	-0.23	-0.06	1.00			
5. ACT	26.91	4.19	***-0.39***	**-0.28**	-0.06	***0.41***	1.00		
6. ACT- English	26.70	5.72	***-0.41***	-0.12	-0.05	0.09	***0.72***	1.00	
7. Openness	3.15	0.78	0.05	0.14	-0.08	-0.05	-0.11	-0.18	1.00
8. Conscientiousness	4.06	0.51	**-0.28**	0.10	-0.07	-0.10	-0.06	-0.16	-0.04
9. Extraversion	3.67	0.57	0.22	0.06	0.29	-0.25	-0.05	0.06	-0.09
10. Agreeableness	3.95	0.37	-0.14	**0.28**	-0.23	-0.07	-0.05	-0.05	-0.12
11. Emotional stability	3.80	0.54	-0.02	-0.23	0.15	-0.06	0.13	0.03	-0.12
12. Core self-evaluation	0.00	1.07	0.01	-0.12	0.06	-0.24	0.05	0.06	-0.17
13. Impression Management	3.28	0.52	0.15	0.05	0.11	-0.06	0.01	-0.07	-0.14
14. Oral interview	2.46	0.61	-0.05	-0.07	0.06	***0.39***	**0.28**	0.10	**-0.28**
15. Proactive personality	5.55	0.79	-0.15	-0.09	**0.31**	-0.14	0.06	-0.04	-0.09
16. Written interview	2.56	0.62	-0.16	0.05	0.01	0.00	0.19	0.16	0.06
17. Training performance	0.00	0.94	-0.12	-0.25	-0.27	0.29	0.31	-0.10	-0.08

Table 3…..continued

Variables	8	9	10	11	12	13	14	15	16	17
8. Conscientiousness	1.00									
9. Extraversion	0.05	1.00								
10. Agreeableness	***0.58***	0.11	1.00							
11. Emotional stability	***0.37***	***0.52***	0.23	1.00						
12. Core self-evaluation	0.18	***0.53***	0.03	***0.65***	1.00					
13. Impression Management	0.15	**0.33**	-0.04	0.18	**0.33**					
14. Proactive personality	***0.62***	***0.42***	0.19	***0.54***	***0.44***	***0.35***	1.00			
15. Oral interview	-0.06	0.21	-0.12	0.15	0.24	***0.39***	0.11	1.00		
16. Written interview	0.00	-0.03	0.00	0.01	0.02	**0.29**	0.03	***0.48***	1.00	
17. Training performance	0.13	-0.27	0.12	0.09	0.19	**0.35**	0.08	0.09	0.10	1.00

Notes

- Bold cells: $p < .05$; Bold and italic cells: $p<.01$;

- For all variables $N = 57$ except for the following: for the variables Age and Number of Interviews $N = 56$; for Training Performance $N = 38$

Table 4: Type 3 Tests of Fixed Effects

Effect	All five questions		Proactive behavior (4 questions)		Proactive behavior (3 questions)	
	F value	*Pr>|t|*	*F* value	*Pr>|t|*	*F* value	*Pr>|t|*
Model 1						
Type	3.87	0.05	7.93	0.01	4.63	0.04
Extraversion	0.58	0.45	0.47	0.49	0.20	0.66
Extraversion*type	3.27	0.08	7.14	0.01	5.31	0.03
Model 2						
Type	1.42	0.24	1.92	0.17	1.20	0.28
Emotional stability	0.46	0.50	0.07	0.79	0.33	0.57
Emotional stability*type	1.08	0.30	1.56	0.22	1.48	0.23
Model 3						
Type	1.38	0.24	1.24	0.27	0.69	0.41
Core Self-evaluation	1.35	0.25	0.34	0.56	0.45	0.50
Core Self-evaluation*type	2.78	0.10	3.63	0.06	5.53	0.02
Model 4						
Type	5.22	0.03	4.81	0.03	6.32	0.01
Openness	0.90	0.35	0.78	0.38	0.65	0.42
Openness*type	7.03	0.01	6.43	0.01	5.69	0.02

- For all the models, the numerator *Df* = 1 and the denominator *Df* = 55

Table 5: Results of Repeated ANOVA

	All five questions				Proactive behavior (4 questions)				Proactive behavior (3 questions)			
Predictor variable	*b*	*SE*	*t*	*p* value	*b*	*SE*	*t*	*p* value	*b*	*SE*	*t*	*p* value
Intercept	2.69	0.54	5.03	<.0001	2.97	0.55	5.31	<.0001	3.00	0.60	5.02	<.0001
Oral interview	-1.04	0.54	-1.97	0.06	-1.61	0.56	-2.82	0.01	-1.49	0.69	-2.15	0.04
Written interview	0.00	.	.	.	0.00	.	.	.	0.00	.	.	.
Extraversion	-0.04	0.15	-0.25	0.81	-0.11	0.15	-0.78	0.47	-0.15	0.16	-0.93	0.36
Ext*oral	0.26	0.14	1.81	0.08	0.42	0.15	2.67	0.01	0.43	0.19	2.30	0.03
Ext*written	0.00	.	.	.	0.00	.	.	.	0.00	.	.	.
Intercept	2.53	0.60	4.3	<.0001	2.84	0.62	4.55	<.0001	2.59	0.66	3.90	0.00
Oral interview	-0.70	0.60	-1.19	0.25	-0.93	0.66	-1.39	0.16	-0.87	0.78	-1.10	0.27
Written interview	0.00	.	.	.	0.00	.	.	.	0.00	.	.	.
Emotional stability	0.01	0.16	0.05	0.96	-0.07	0.16	-0.44	0.66	-0.03	0.17	-0.20	0.84
ES*oral	0.16	0.16	1.04	0.31	0.22	0.17	1.25	0.21	0.25	0.20	1.22	0.23
ES*written	0.00	.	.	.	0.00	.	.	.	0.00	.	.	.
Intercept	2.56	0.08	31.44	<.0001	2.57	0.08	29.77	<.0001	2.46	0.09	26.92	<.0001
Oral interview	-0.09	0.08	-1.18	0.25	-0.11	0.09	-1.11	0.22	0.08	0.10	0.83	0.44
Written interview	0.00	.	.	.	0.00	.	.	.	0.00	.	.	.
Core self-evaluation	0.01	0.08	0.15	0.88	-0.04	0.08	-0.50	0.61	-0.06	0.09	-0.74	0.46
CSE*oral	0.13	0.08	1.67	0.11	0.16	0.08	1.91	0.06	0.23	0.10	2.35	0.02
CSE*written	0.00	.	.	.	0.00	.	.	.	0.00	.	.	.
Intercept	2.40	0.35	7.03	<.0001	2.37	0.35	6.55	<.0001	2.22	0.38	5.77	<.0001
Oral interview	0.75	0.33	2.29	0.03	0.82	0.36	2.19	0.04	1.11	0.44	2.51	0.01

Written interview	0.00	.	·	.	0.00	.	·	.	0.00	.	·	.
Openness	0.05	0.11	0.47	0.65	0.06	0.11	0.54	0.67	0.08	0.12	0.64	0.52
Openness*oral	-0.27	0.10	-2.65	0.01	-0.29	0.11	-2.54	0.01	-0.33	0.14	-2.38	0.02
Openness*written	0.00	.	·	.	0.00	.	·	.	0.00	.	·	.

- For all the models, the numerator *Df* = 1 and the denominator *Df* = 55

Table 6: The Contrast Tests for Extraversion and Interview Performance

Comparisons	*b*	*SE*	*df*	*t*	*Pr>|t|*	95% Confidence Interval	
All five questions						Lower	Upper
Oral - Written interview at Extraversion = 3.1	-0.24	0.11	55.00	-2.12	0.04	-0.47	-0.01
Oral - Written interview at Extraversion = 3.67	-0.09	0.08	55.00	-1.18	0.24	-0.26	0.07
Oral - Written interview at Extraversion = 4.24	0.05	0.11	55.00	0.46	0.65	-0.18	0.28
Four Proactive behavior questions							
Oral - Written interview at Extraversion = 3.1	-0.34	0.12	55.00	-2.70	0.01	-0.59	-0.09
Oral - Written interview at Extraversion = 3.67	-0.10	0.09	55.00	-1.14	0.26	-0.28	0.08
Oral - Written interview at Extraversion = 4.24	0.14	0.13	55.00	1.10	0.28	-0.11	0.39
Three Proactive behavior questions							
Oral - Written interview at Extraversion = 3.1	-0.16	0.15	55.00	-1.05	0.30	-0.46	0.14
Oral - Written interview at Extraversion = 3.67	0.09	0.11	55.00	0.84	0.41	-0.12	0.30
Oral - Written interview at Extraversion = 4.24	0.33	0.15	55.00	2.22	0.03	0.03	0.64
Factor Score - Proactive behavior							
Oral - Written interview at Extraversion = 3.1	-0.34	0.17	55.00	-1.98	0.05	-0.69	0.00
Oral - Written interview at Extraversion = 3.67	0.00	0.12	55.00	0.01	0.99	-0.24	0.24
Oral - Written interview at Extraversion = 4.24	0.34	0.17	55.00	1.99	0.05	0.00	0.69

- For all the models, the numerator *Df* = 1 and the denominator *Df* = 55

Table 7: The Contrast Tests for Core Self-Evaluation and Interview Performance

Comparisons	*b*	*SE*	*df*	*t*	*Pr*>|*t*|	95% Confidence interval	
All five questions						Lower	Upper
Oral - Written interview at Core Evaluation = -1.07	-0.23	0.11	55	-2.01	0.05	-0.46	0.00
Oral - Written interview at Core Evaluation = 0	-0.09	0.08	55	-1.18	0.24	-0.26	0.07
Oral - Written interview at Core Evaluation = 1.07	0.04	0.11	55	0.35	0.73	-0.19	0.27
Four Proactive behavior questions							
Oral - Written interview at Core Evaluation = -1.07	-0.27	0.13	55	-2.14	0.04	-0.53	-0.02
Oral - Written interview at Core Evaluation = 0	-0.10	0.09	55	-1.11	0.27	-0.28	0.08
Oral - Written interview at Core Evaluation = 1.07	0.07	0.13	55	0.56	0.57	-0.18	0.33
Three Proactive behavior questions							
Oral - Written interview at Core Evaluation = -1.07	-0.16	0.15	55	-1.08	0.29	-0.46	0.14
Oral - Written interview at Core Evaluation = 0	0.09	0.11	55	0.83	0.41	-0.12	0.30
Oral - Written interview at Core Evaluation = 1.07	0.34	0.15	55	2.25	0.03	0.04	0.64

- For all the models, the numerator *Df* = 1 and the denominator *Df* = 55

Table 8: The Contrast Tests for Openness to Experience and Interview Performance

Comparisons	*b*	*SE*	*df*	*t*	*Pr>|t|*	95% Confidence interval	
All five questions						Lower	Upper
Oral - Written interview at Openness = 2.37	0.11	0.11	55	1.03	0.31	-0.11	0.34
Oral - Written interview at Openness = 3.15	-0.09	0.08	55	-1.22	0.23	-0.25	0.06
Oral - Written interview at Openness = 3.93	-0.30	0.11	55	-2.74	0.01	-0.52	-0.08
Four Proactive behavior questions							
Oral - Written interview at Openness = 2.37	0.13	0.13	55	1.01	0.32	-0.13	0.38
Oral - Written interview at Openness = 3.15	-0.10	0.09	55	-1.15	0.26	-0.28	0.08
Oral - Written interview at Openness = 3.93	-0.33	0.13	55	-2.60	0.01	-0.58	-0.08
Three Proactive behavior questions							
Oral - Written interview at Openness = 2.37	0.34	0.15	55	2.28	0.03	0.04	0.64
Oral - Written interview at Openness = 3.15	0.09	0.11	55	0.83	0.41	-0.12	0.30
Oral - Written interview at Openness = 3.93	-0.17	0.15	55	-1.11	0.27	-0.47	0.13

- For all the models, the numerator *Df* = 1 and the denominator *Df* = 55

Table 9: Training Performance and Extraversion - Overall ANOVA

				Overall Non-centrality Parameter				Proportion of Variation Accounted for			
Model	F	Pr>F	R-sq	Minimum Variance Unbiased Estimate	Low MSE Estimate	95% Confidence Limits		η^2	ω^2	95% Confidence Limits	
1	2.35	0.06	0.27	5.99	5.95	0.00	25.69	0.27	0.15	0.00	0.40
2	3.08	0.02	0.32	9.43	8.81	0.14	32.19	0.32	0.21	0.00	0.46
3	1.88	0.13	0.23	3.81	3.55	0.00	21.40	0.23	0.11	0.00	0.36
4	0.70	0.63	0.10	-1.71	-1.60	0.00	9.19	0.10	-0.04	0.00	0.19

Notes

1. For all the models, total *df* = 37 and model *df* = 5
2. All the models have training performance as dependent variable; these models have extraversion, oral interview, written interview, and two interaction terms (i.e., oral * extraversion + written*extraversion) as predictors.
3. For model 1, the total oral and written interview scores are used as predictors.
4. For model 2, the scores of the four proactive behaviors questions for the oral and written interview are used as predictors.
5. For model 3, the scores of the three proactive behaviors questions for the oral and written interview are used as predictors.
6. For model 4, the scores of the planning and organizing question (i.e., q3) for the oral and written interview are used as predictors

Table 10: ANOVA Results with Effect Sizes

Source	Type III SS	F	Pr>F	Non-centrality Parameter: Min Variance Unbiased Estimate	Non-centrality Parameter: Low MSE Estimate	Non-centrality Parameter: 95% Confidence Limits		Partial Variation Accounted for: Partial η^2	Partial Variation Accounted for: Partial ω^2	Partial Variation Accounted for: 95% Confidence Limits	
Model 1											
Extraversion	3.02	4.01	0.05	2.76	2.58	0.00	16.07	0.11	0.07	0.00	0.30
Oral Interview	4.35	5.78	0.02	4.42	4.12	0.05	19.65	0.15	0.11	0.00	0.34
Written Interview	0.01	0.01	0.93	-0.99	-0.93	0.00	1.97	0.00	-0.03	0.00	0.05
Extraversion*Oral	3.95	5.24	0.03	3.91	3.65	0.00	18.58	0.14	0.10	0.00	0.33
Extraversion*Written	0.02	0.03	0.87	-0.97	-0.91	0.00	3.34	0.00	-0.03	0.00	0.08
Model 2											
Extraversion	3.95	5.69	0.02	4.33	4.04	0.03	19.46	0.15	0.11	0.00	0.34
Oral Interview	5.90	8.48	0.01	6.95	6.49	0.65	24.78	0.21	0.16	0.02	0.39
Written Interview	0.12	0.18	0.68	6.26	5.84	0.46	23.39	0.19	0.15	0.01	0.38
Extraversion*Oral	5.38	7.74	0.01	-0.83	-0.78	0.00	5.44	0.01	-0.02	0.00	0.13
Extraversion*Written	0.15	0.21	0.65	-0.80	-0.75	0.00	5.66	0.01	-0.02	0.00	0.13
Model 3											
Extraversion	1.73	2.17	0.15	1.03	0.97	0.00	11.93	0.06	0.03	0.00	0.24
Oral Interview	3.78	4.75	0.04	3.45	3.22	0.00	17.59	0.13	0.09	0.00	0.32
Written Interview	0.01	0.01	0.93	-0.99	-0.93	0.00	1.95	0.00	-0.03	0.00	0.05
Extraversion*Oral	3.30	4.15	0.05	2.89	2.70	0.00	16.35	0.11	0.08	0.00	0.30
Extraversion*Written	0.03	0.03	0.86	-0.97	-0.91	0.00	3.49	0.00	-0.03	0.00	0.08

Table 10: *Continued*

Source	Type III SS	F	Pr>F	Non-centrality Parameter: Min Variance Unbiased Estimate	Non-centrality Parameter: Low MSE Estimate	Non-centrality Parameter: 95% Confidence Limits		Partial Variation Accounted for: Partial η^2	Partial Variation Accounted for: Partial ω^2	Partial Variation Accounted for: 95% Confidence Limits	
Model 4											
Extraversion	0.02	0.02	0.88	-0.98	-0.91	0.00	3.14	0.00	-0.03	0.00	0.08
Oral Interview	0.31	0.33	0.57	-0.69	-0.64	0.00	6.35	0.01	-0.02	0.00	0.14
Written Interview	0.49	0.52	0.47	-0.51	-0.48	0.00	7.18	0.02	-0.01	0.00	0.16
Extraversion*Oral	0.37	0.40	0.53	-0.63	-0.59	0.00	6.65	0.01	-0.02	0.00	0.15
Extraversion*Written	0.57	0.62	0.44	-0.42	-0.39	0.00	7.54	0.02	-0.01	0.00	0.17

Notes:

1. For all the predictors df = 1 and total *df* = 37
2. All the models have training performance as dependent variable; these models have extraversion, oral interview, written interview, and two interaction terms (i.e., oral * extraversion + written*extraversion) as predictors.
3. For model 1, the total oral and written interview scores are used as predictors.
4. For model 2, the scores of the four proactive behaviors questions for the oral and written interview are used as predictors.
5. For model 3, the scores of the three proactive behaviors questions for the oral and written interview are used as predictors.
6. For model 4, the scores of the planning and organizing question (i.e., q3) for the oral and written interview are used as predictors

Table 11: Training Performance and Extraversion – Parameter Estimates

Source	*b*	β	*SE*	*t*	*Pr* > |*t*|
Model 1					
Intercept	-10.86	0.00	5.04	-2.15	0.04
Extraversion	2.79	1.67	1.39	2.00	0.05
Oral Interview	5.10	3.10	2.12	2.40	0.02
Written Interview	0.14	0.09	1.71	0.08	0.93
Extraversion*Oral	-1.28	-3.94	0.56	-2.29	0.03
Extraversion*Written	-0.07	-0.20	0.45	-0.17	0.87
Model 2					
Intercept	-11.55	0.00	4.51	-2.56	0.02
Extraversion	2.91	1.74	1.22	2.38	0.02
Oral Interview	4.88	3.42	1.68	2.91	0.01
Written Interview	0.60	0.41	1.42	0.42	0.68
Extraversion*Oral	-1.22	-4.26	0.44	-2.78	0.01
Extraversion*Written	-0.17	-0.49	0.37	-0.46	0.65
Model 3					
Intercept	-6.96	0.00	4.42	-1.58	0.12
Extraversion	1.79	1.07	1.21	1.47	0.15
Oral Interview	3.34	2.70	1.53	2.18	0.04
Written Interview	0.12	0.08	1.40	0.08	0.93
Extraversion*Oral	-0.83	-3.20	0.41	-2.04	0.05
Extraversion*Written	-0.07	-0.19	0.38	-0.18	0.86
Model 4					
Intercept	0.62	0.00	3.10	0.20	0.84
Extraversion	-0.13	-0.08	0.89	-0.15	0.88
Oral Interview	-0.93	-1.06	1.62	-0.58	0.57
Written Interview	1.30	1.42	1.80	0.72	0.47
Extraversion*Oral	0.27	1.14	0.43	0.63	0.53
Extraversion*Written	-0.38	-1.63	0.49	-0.79	0.44

Notes:

1. For all the predictors df = 1 and total df = 37
2. All the models have training performance as dependent variable; these models have extraversion, oral interview, written interview, and two interaction terms (i.e., oral * extraversion + written*extraversion) as predictors.

Table 12: Training Performance and Emotional Stability – Overall ANOVA

				Overall Non-centrality Parameter				Proportion of Variation Accounted for			
Model	F	Pr>F	R-sq	Minimum Variance Unbiased Estimate	Low MSE Estimate	95% Confidence Limits		η^2	ω^2	95% Confidence Limits	
1	0.22	0.95	0.03	-3.96	-3.69	0.00	1.70	0.03	-0.11	0.00	0.04
2	0.22	0.95	0.03	-3.98	-3.71	0.00	1.54	0.03	-0.11	0.00	0.04
3	0.19	0.97	0.03	-4.12	-3.85	0.00	0.78	0.03	-0.12	0.00	0.02
4	0.27	0.92	0.04	-3.72	-3.47	0.00	2.79	0.04	-0.11	0.00	0.07

Notes

1. For all the models, total df = 37 and model df = 5
2. All the models have training performance as dependent variable; these models have emotional stability, oral interview, written interview, and two interaction terms (i.e., oral * emotional stability and written* emotional stability) as predictors.
3. For model 1, the total oral and written interview scores are used as predictors.
4. For model 2, the scores of the four proactive behaviors questions for the oral and written interview are used as predictors.
5. For model 3, the scores of the three proactive behaviors questions for the oral and written interview are used as predictors.
6. For model 4, the scores of the planning and organizing question (i.e., q3) for the oral and written interview are used as predictors.

Table 13: Training performance and Core Self-Evaluation – Overall ANOVA

				Overall Non-centrality Parameter				Proportion of Variation Accounted for			
Model	F	Pr>F	R-sq	Minimum Variance Unbiased Estimate	Low MSE Estimate	95% Confidence Limits		η^2	ω^2	95% Confidence Limits	
1	0.59	0.70	0.08	-2.22	-2.07	0.00	7.81	0.08	-0.06	0.00	0.17
2	0.55	0.73	0.08	-2.41	-2.25	0.00	7.26	0.08	-0.06	0.00	0.16
3	0.54	0.74	0.08	-2.46	-2.30	0.00	7.13	0.08	-0.06	0.00	0.16
4	0.84	0.53	0.12	-1.06	-0.99	0.00	10.84	0.12	-0.02	0.00	0.22

Notes

1. For all the models, total df = 37 and model df = 5
2. All the models have training performance as dependent variable; these models have core self-evaluation, oral interview, written interview, and two interaction terms (i.e., oral * core self-evaluation and written* core self-evaluation) as predictors.
3. For model 1, the total oral and written interview scores are used as predictors.
4. For model 2, the scores of the four proactive behaviors questions for the oral and written interview are used as predictors.
5. For model 3, the scores of the three proactive behaviors questions for the oral and written interview are used as predictors.
6. For model 4, the scores of the planning and organizing question (i.e., q3) for the oral and written interview are used as predictors.

Table 14: Training performance and Openness to Experience – Overall ANOVA

				Overall Non-centrality Parameter				Proportion of Variation Accounted for			
Model	F	Pr>F	R-sq	Minimum Variance Unbiased Estimate	Low MSE Estimate	95% Confidence Limits		η^2	ω^2	95% Confidence Limits	
1	0.41	0.84	0.06	-3.09	-2.88	0.00	5.17	0.06	-0.08	0.00	0.12
2	0.40	0.85	0.06	-3.13	-2.92	0.00	5.02	0.06	-0.09	0.00	0.12
3	0.56	0.73	0.08	-2.37	-2.21	0.00	7.38	0.08	-0.06	0.00	0.16
4	0.93	0.48	0.13	-0.65	-0.61	0.00	11.83	0.13	-0.01	0.00	0.24

Notes

1. For all the models, total df = 37 and model df = 5
2. All the models have training performance as dependent variable; these models have Openness, oral interview, written interview, and two interaction terms (i.e., oral * openness and written* openness) as predictors.
3. For model 1, the total oral and written interview scores are used as predictors.
4. For model 2, the scores of the four proactive behaviors questions for the oral and written interview are used as predictors.
5. For model 3, the scores of the three proactive behaviors questions for the oral and written interview are used as predictors.
6. For model 4, the scores of the planning and organizing question (i.e., q3) for the oral and written interview are used as predictors.

Figure 1: Interaction between Interview Type and Extraversion – Four Questions for Proactive Behaviors

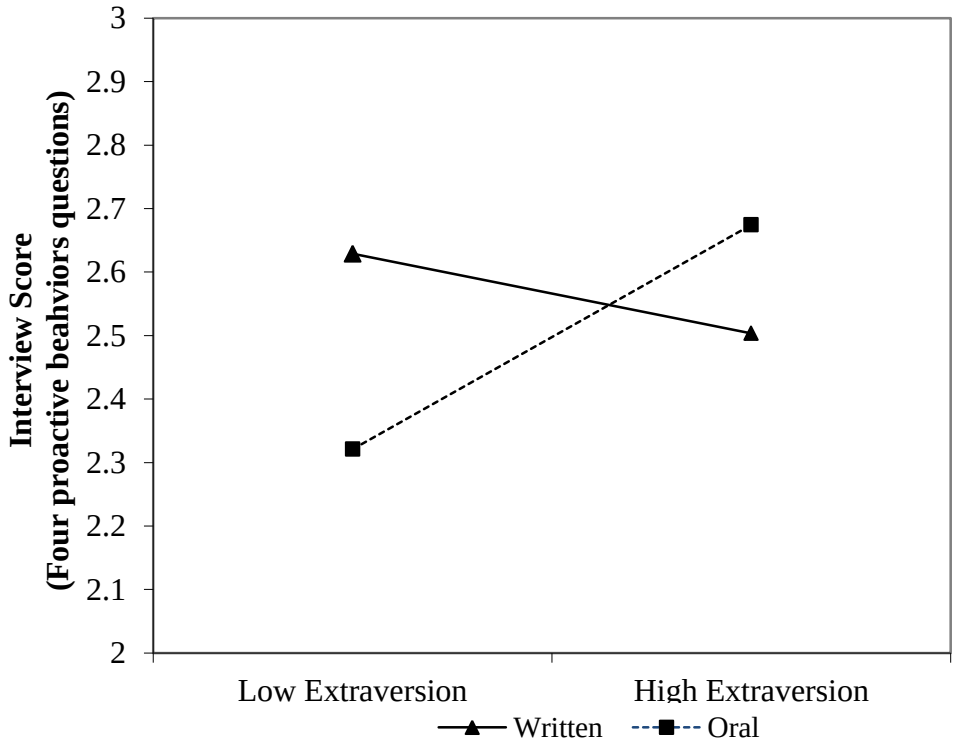


Figure 2: Interaction between Interview Type and Extraversion – Three Questions for Proactive Behaviors

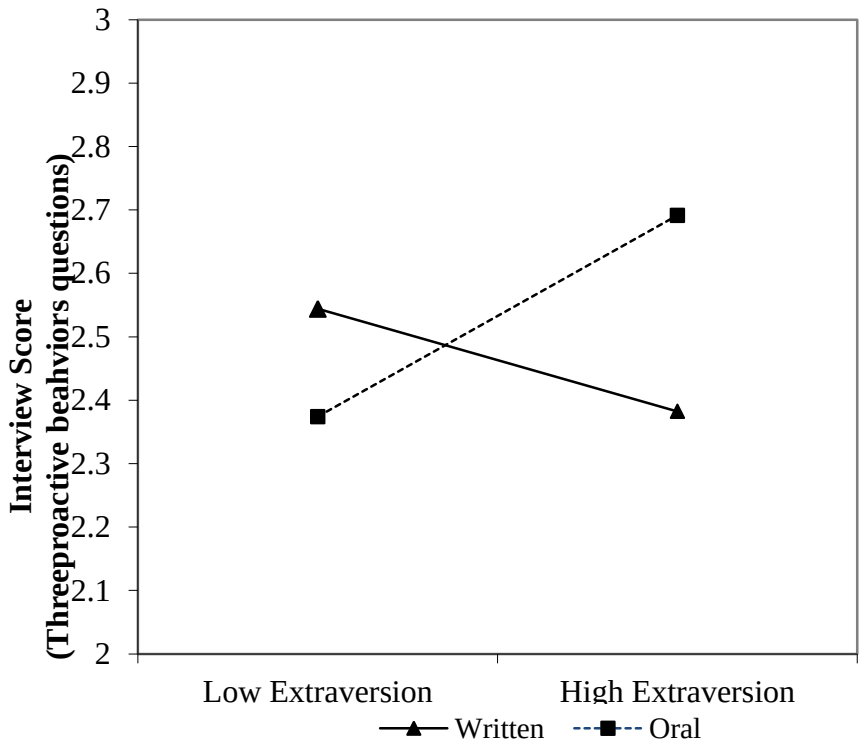


Figure 3: Interaction between Interview Type and Core Self-Evaluations – Three Questions for Proactive Behaviors

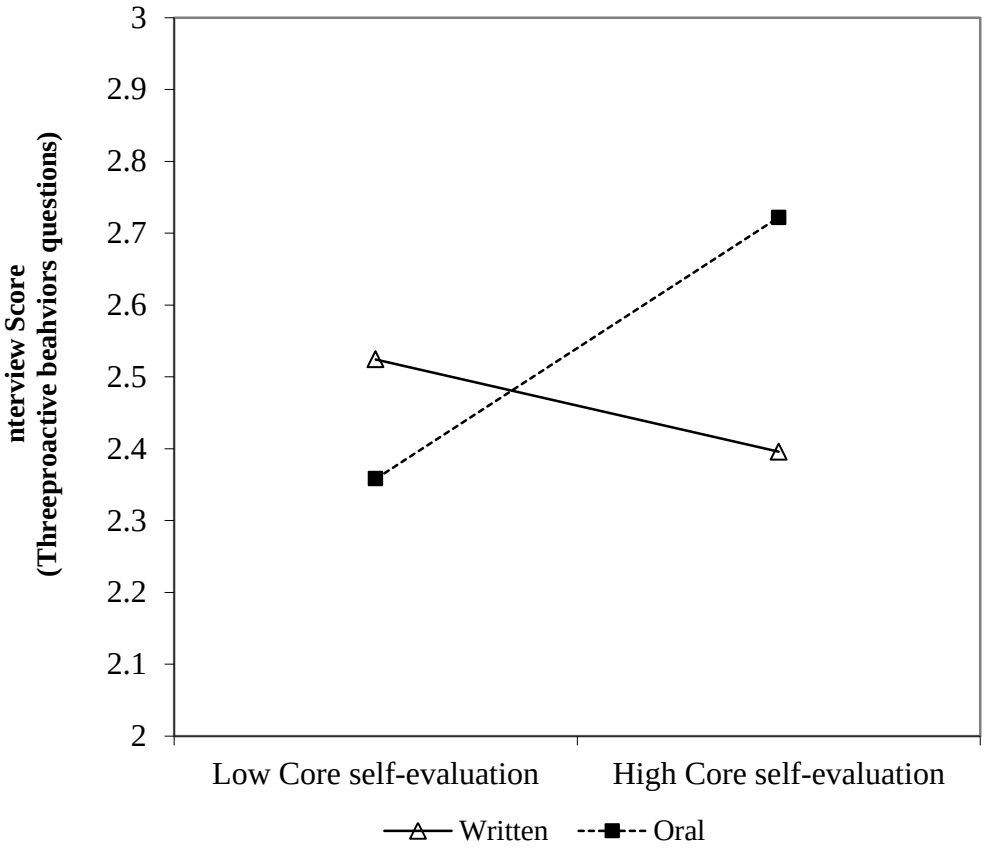


Figure 4: Interaction between Interview Type and Openness to Experience – All Five Questions

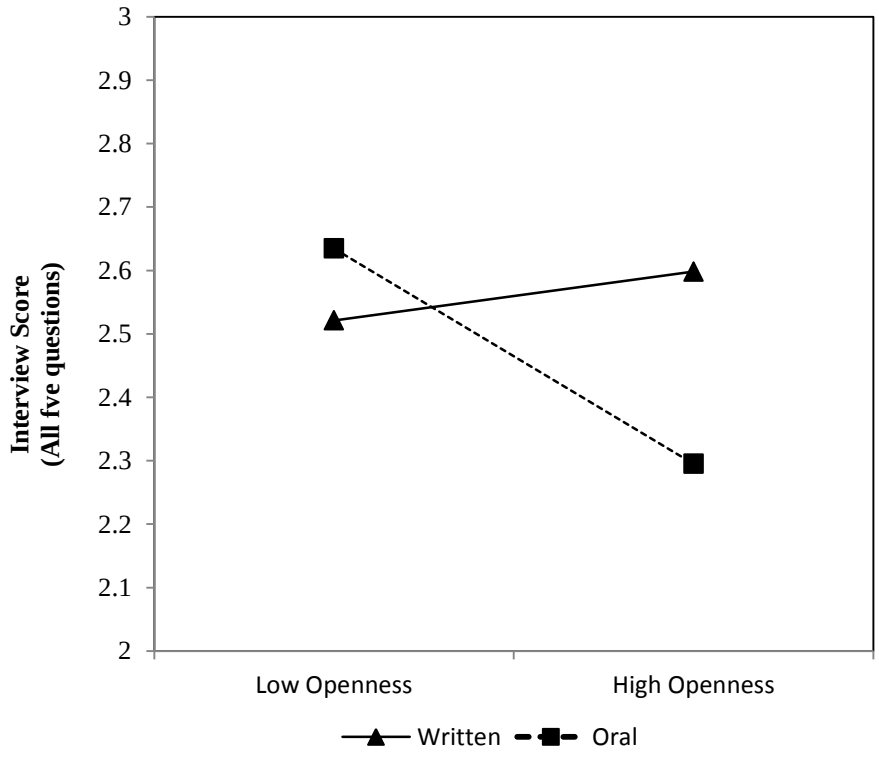


Figure 5: Interaction between Interview Type and Openness to Experience – Four Questions for Proactive Behaviors

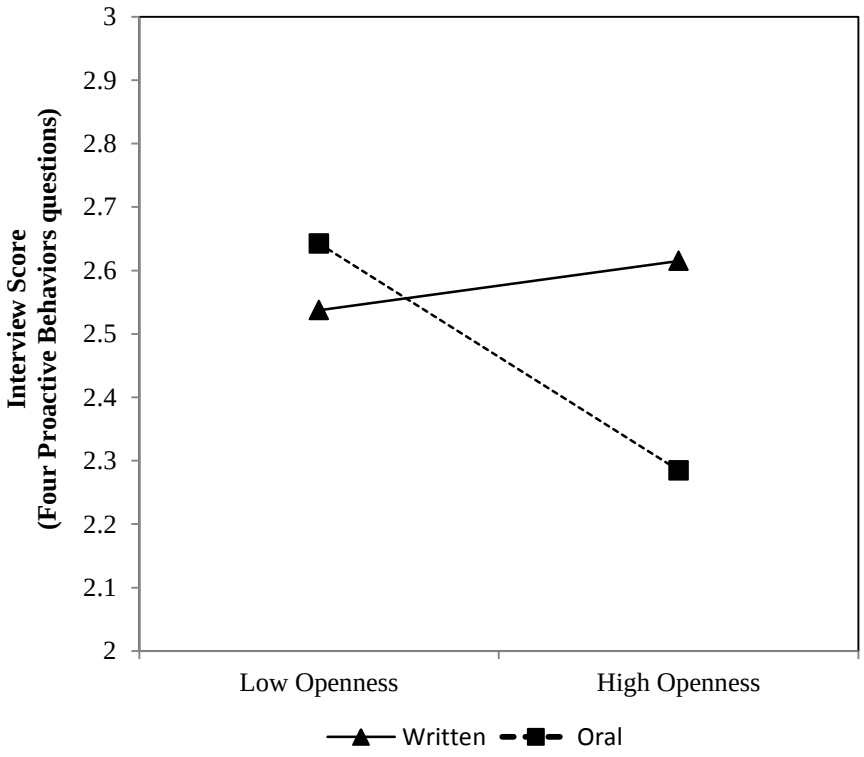


Figure 6: Interaction between Interview Type and Openness to Experience – Three Questions for Proactive Behaviors

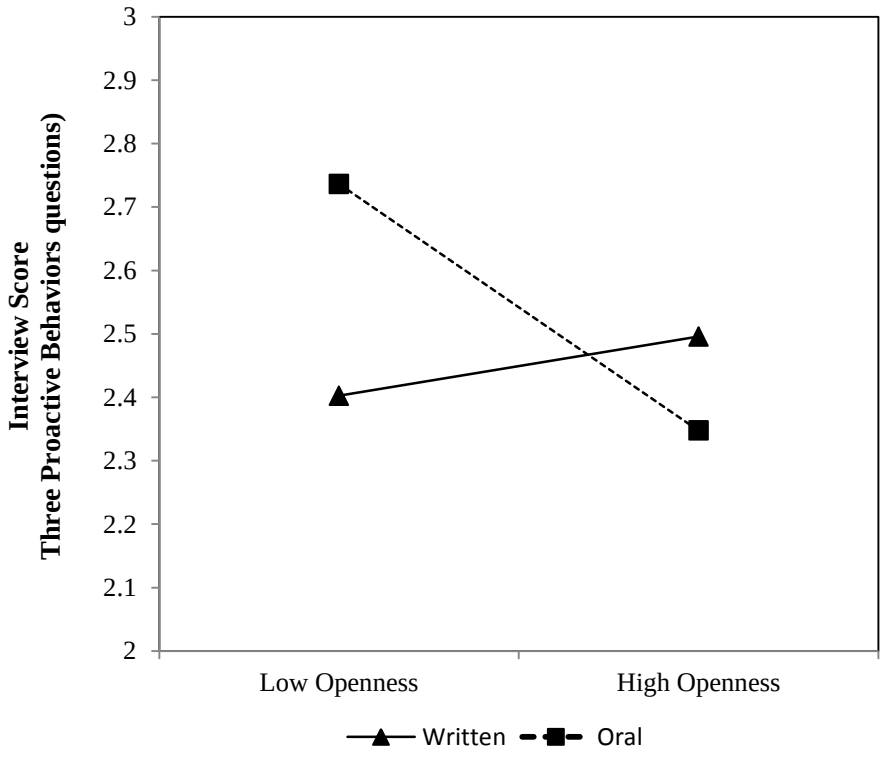
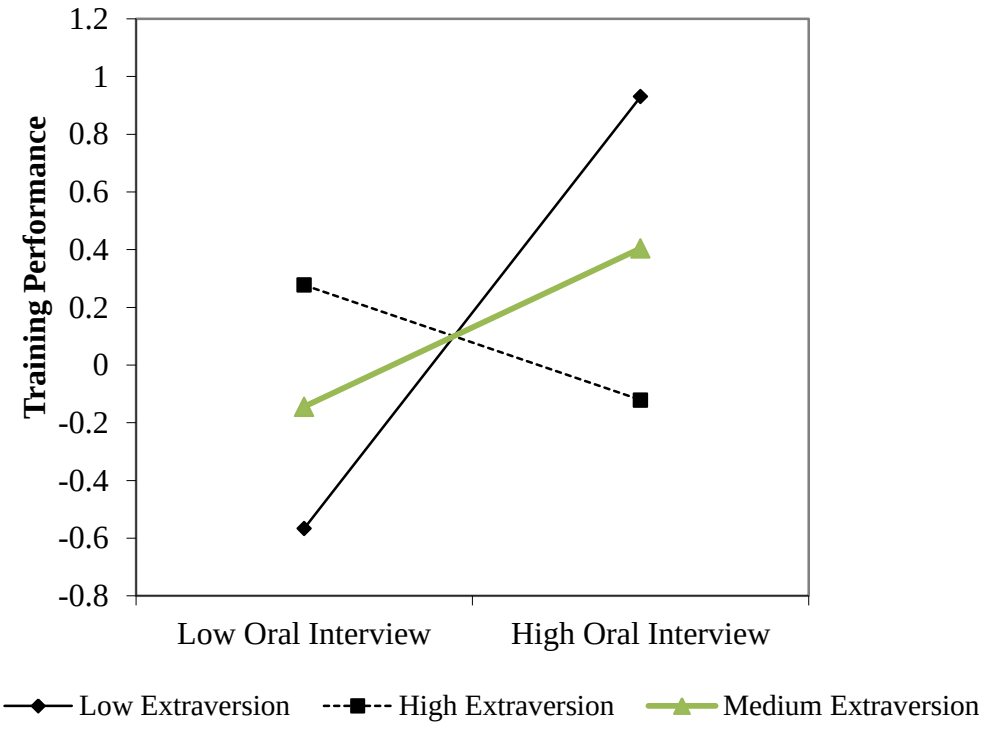


Figure 7: Interaction Between Oral Interview (Four Proactive Behaviors Questions) and Extraversion



Annexure 1: Assessment Key for Behavioral Description Interview

Questions

Ask all candidates the exact same question in a uniform manner:

1. Same person should ask these questions.
2. These questions should be asked in the same order.
3. No probing of questions from candidates.
4. However, you may repeat or clarify if a candidate does not understand the question.

QUESTION NO 1

Think about the last five years and tell me what measures you have taken in the past concerning your self-development. What specific improvements have resulted from this?

Dimension: Self-development behavior *(a set of activities that shows a focus towards evaluating one's self, setting standards, noticing discrepancies, and enthusiastically working towards reducing the discrepancies)*

1. Sets personal standards of excellence.
2. Engages in self-assessment to create a self-awareness of weaknesses.
3. Proactively seeks multiple ways to improve self.
4. Seeks feedback specifically negative.
5. Monitors self-improvement.
6. Manages negative emotions (e.g., pessimism, helplessness, and dislike of those who criticize)

Response Key

Excellent (5) if the following elements are present:	**Examples**
1. Mention a weakness that he/she was aware of. Hint on how they knew of it. 2. Mention what they did to address the issue, including what resources they used. To be a 5 it needs to involve other people as well. Also, it has to show	I have always wanted to be good at public speaking because I want to start a business consulting firm. During various class presentations, I noticed that others around me were much better at public speaking. I decided I needed to do something about it. I talked to various senior students and teachers in the school, and found out about Toastmasters. I have been active member of this group since last two years, and have

some proactivity on the part of the respondent in seeking resources. 3. Mention the ways they improved in. Mention how they knew they had improved. Just saying that they improved is not enough.	availed many opportunities to improve my skills through its membership e.g., talking to others, seeking advice. In my interactions with others, I have paid specific attention to the aspects I was weak in such as narrating interesting stories. I practiced the advice given from others. I never became defensive when someone pointed out my flaws. For example, a fellow member in Toastmasters pointed out that when I am presenting I somehow start speaking in a very affected manner. I guess I would do so to hide my nervousness. I kept working on these issues and I believe I have made significant improvements in my presentation skills; I was a very shy person and had stage freight. However, I soon realized that it is a skill that can be developed. First, I started paying attention towards individuals who I thought were good at public speaking. I would always approach them and ask them what enables them to deliver good presentation. I will make detailed notes of tips given by these role models. Then I searched a lot of public speaking articles on the internet and in the library. These helped me understand specific challenges that I was facing regarding voice control, nervousness, and organizing the message. I also started requesting friends and family to listen to my presentation and give me feedback. It was often embarrassing to expose my weaknesses to my friends and family but I realized that this is the only way I can improve. A teacher suggested that I should video tape my presentations and should play back these clips to observe my weaknesses and monitor improvement. I have significantly improved my skills over the last few years.
Good (4) 1. Mention a weakness that he/she was aware of.	**----------**

2. Mention what they did to address the issue, including what resources they used. Show an evidence of taking initiative. **3.** Mention the ways they improved in.	
Acceptable (3) **if the following elements are present:**	**Examples**
1. Weakness or area of improvement is identified either beforehand or in the ways they improved in. 2. Mention what they did but no resources are identified. There appears to be little or no proactivity. (Mentioning hardwork or gaining experience do not qualify as a resource; penalize clichés such as time management)	As I was weak in public speaking, I tried to volunteer more for making presentations. Before, any presentation, I practiced a lot. Through these methods, I have improved my skills considerably.
Fair (2) Either 1. Know of a weakness, but provide no evidence of actions taken to address it, or just mention future intentions. OR 1. Vague about area of improvement, but talks in general about self-development actions	I have enrolled in various skills development programs. I participated in extra-curricular activities.
Poor (1) **if the following elements are present:**	**Examples**
1. Very little evidence of self-awareness 2. No evidence of clear self-development goals 3. A lack of focus towards seeking opportunities for self-	I have worked really hard to get good grades. Or I have been successful to come this far through my desire to improve. Or I believe in constant learning, we are never perfect. (clichéd)

development 4. Might give a clichéd answer.	

QUESTION 2

Tell me about your attempts to seek out a mentor in the past. Is there someone who you see as a mentor in your academic life? How did you identify this person and how do you solicit advice from this person?

Dimension: Proactive Self-development through mentorship *(a set of activities providing evidence of self evaluation, setting standards, noticing discrepancies, and enthusiastically working with a mentor to reduce these discrepancies).*

Excellent (5) if the following elements are present:	**Examples**
I. Provides an evidence of at least one mentoring relationship (class instructors don't count if they only provide advice for the respective class). II. Demonstrate they actively looked for a mentor. III. Demonstrate active efforts to cultivate a relationship (e.g., regular visits etc). IV. Makes an effort to know the mentor (should at the very least mention something about the mentor's qualities or show appreciation for the person) V. Show an interest in being coached. VI. Identify area of weakness on which they want to be coached or supported (social, instrumental, technical support). Item # I+II + at least 2 others	In my senior years at High School, I was not very clear whether I would be attending the college or not. I heard about this mentoring program in my school where you were matched up with a graduate student in a college. I enrolled in this program and was matched up with a graduate student. I soon realized that this student had so much to offer me. He had a wealth of knowledge but I needed to ask the right questions. I started actively seeking feedback from him about college life, various career options, and opportunities. I would eagerly wait for our weekly meetings and email him whenever I would need any guidance. I would not wait for him to offer guidance, I would approach him myself. Through this feedback seeking and mentoring relationship, I refined my career goals, really understood what it takes to be a successful college student, and how to realize my goals. We have become great friends and still share thoughts on career goals.
Good (4)	**----------**

Item # I + at least 3 others (excluding #II) OR Item # I+II + any other	
Acceptable (3) if the following elements are present:	**Examples**
Item # I + any 2 others OR Any 3 (excluding # I)	I always regularly attend meetings with my advisors. I listen carefully to what they have to say. For example one of High school teachers pointed out that I need improving my writing skills and she showed me various resources that I can use to improve my skills.
Fair (2) Items # I+II + any other OR Any 2 (excluding # I)	**----------**
Poor (1) if the following elements are present:	**Examples**
1. Very little evidence of self-awareness 2. A lack of focus towards seeking mentoring opportunities. 3. Might give a clichéd answer about having a role model.	My mother/uncle/father is my biggest mentor (clichéd). OR I always seek advice from my teachers and parents. OR Every person has to teach you something if you want to learn (clichéd).

QUESTION 3

Tell me about a busy week at school when you have multiple deadlines to meet or have multiple exams. How do you plan your activities for that week?

Dimension: Planning and organizing *(A behavior that shows planning ahead and working in a systematic and organized way. An evidence of following through set plans and procedures.)*

Excellent (5) if the following elements are present:	**Examples**
1. Makes a conscious effort to think in advance.	I am a very organized person. I enter all my commitments in an online Microsoft

2. **Regularly** plans for resources required and timelines available in a sufficient detail. Detail planning. 3. Establishes some rules and norms. For instance, uses tools to organize information, checks schedule on set times, sets reminders etc 4. Mention about priorities as a rule/norm. 5. Monitors the plan and takes action if there is a discrepancy. (Any 4 items of above).	calendar on a regular basis. I always set up reminders for at least a week in advance. I have a habit of checking my weekly schedule every morning to see where I am headed. My online schedule helps me know in advance about my project deadlines. Based on these reminders, I take special note of weeks with heavy workload. In addition, I categorize and prioritize different projects according to the difficulty and urgency levels. I think in detail about how much time each activity will take. This way, I plan ahead, and know what activities need to be done when. In addition, I monitor my progress according to the plans and if some adjustments need to be made I make them right away. You invest time in planning but it saves you lot of troubles and slips.
Good (4)	**----------**
Acceptable (3) if the following elements are present:	**Examples**
1. Understands the importance of planning activities. 2. Can plan activities when required to do so. 3. However, there seems to be insufficient evidence of detailed regular planning, resource allocation, and monitoring activities.	When I realize that many deadlines are approaching in a given week, I sit down and think about what to do. I think about how much time each activity will take. This way I know well in advance what needs to be done. For example, last week I had two exams due and a team meeting with class fellows in a different college. I knew of this clash in advance due to my planning ahead and thus convinced my team members to meet the week after.
Fair (2)	**----------**
Poor (1) if the following elements are present:	**Examples**
1. Shows a lack of focus towards thinking in advance about multiple commitments and projects. 2. Approaches activities in an ad hoc manner 3. Makes a clichéd statement.	I work extra hour in those weeks; I plan effectively and work efficiently; I try to do my best; Such weeks are difficult by I try to do my best; I make a detailed plan and then stick to it (clichéd).

QUESTION 4
What was the most difficult subject that you ever had to learn in your academic life? Were you able to improve your knowledge of this subject? If yes, how; If not, why?

Dimension: Self-development behavior *(a set of activities that shows a focus towards self evaluation, setting standards, noticing discrepancies, and enthusiastically working towards reducing the discrepancies.)*

Response Key

Excellent (5) if the following elements are present:	**Examples**
i. Sets personal standards of excellence. ii. Engages in self-assessment to create self-awareness of one's weaknesses. iii. Proactively seeks multiple ways to improve self. iv. Seeks feedback specifically negative. v. Monitors self-improvement. vi. Manages negative emotions (e.g., pessimism, helplessness, and dislike of those who criticize). vii. Mention why it was important. (At least 4 of the above items should be present and some form of interpersonal interaction)	Mathematics and Trigonometry have been very difficult subjects for me. I took it as a personal challenge to improve my skills in these subjects. This required lot of patience and persistence on my part. However, I decided to be positive about my efforts and not get discouraged. I approached one of my fellow students. She was particularly good in Mathematics. We worked out a routine of studying together that really helped me grasp the key concepts. I also started consulting our instructor more often. In addition, I allocated weekends for practicing the questions that were very difficult. I specifically sought help in the areas that I was weak in and practiced those questions more so. Sometimes, I felt embarrassed to ask the instructor or my friend something over and over or felt that I was annoying others. But I needed to learn the subject no matter what, so I had to overcome these hang ups.
Good (4) Item # III (needs to mention several resources, including people) + any other item. OR Any 3 of the above Items.	----------
Acceptable (3)	**Examples**

if the following elements are present:	
Item #III (may or may not include people resources) OR Any two of the above.	Mathematics has been a very difficult subject for me. However, I allocated more time to it. I also consulted others to help me with the concepts that were difficult to understand; I was weak in Physics. Therefore, I made an extra effort to learn this subject. For example, I will talk to the instructor after the class to help clarify the issues. In order to do so, I would always take notes in the class and right away clarify any confusion that I might have.
Fair (2) Typically just working hard or using only one help source (e.g., TA, office hours), combined with some understanding of the importance of learning.	**----------**
Poor (1) **if the following elements are present:**	**Examples**
1. Very little evidence of self-awareness 2. No evidence of clear self-development goals 3. A lack of focus towards seeking opportunities for self-development 4. Might give a clichéd answer. 5. No mention why it was important	I have worked really hard to get good grades; Mathematics. I tried my best and eventually decided that it is not my cup of tea; I did whatever there was to do to excel in Physics (clichéd).

QUESTION 5:
Tell me about a time when you were involved in a project that resulted in a constructive change. How did you facilitate the change? What actions did you take to overcome the challenges in implementing the change?

Dimension: Proactive behavior *(a set of activities that shows a focus towards scanning the environment for opportunities to improve, showing initiative, taking action, and persevering until one reaches closure by bringing about the change.)*

Response Key

Excellent (5) if the following elements are present:	**Examples**
I. Shows a keen focus towards making things better, doing something new, bringing some change. II. Takes an initiative to identify an opportunity for change. III. Actively acts on the change opportunity. IV. Takes up a stand and resists opposition and persists till the change is brought about. V. The respondent has to have a major role and the points above need to be strongly emphasized.	I was once participating in arranging a charity event in my school for kids. I suggested that we should have a music event. My committee members said this would need extra funding and work. I argued that the extra effort would bring more results. They did not seem convinced. I went out and talked to some of my friends who had arranged similar events. After talking to them I prepared an expense and gains report and showed it to my committee. My numbers convinced them to consider this event as a serious option. They asked me if I will take the responsibility to arrange the music segment. I said sure why not. My biggest challenge was to gain the necessary support; I talked to my friends, family members, and contacts that I knew would help in a charity event. It was tough doing all this with the other school work and exams. But I focused on the end objective. I believe when you are sincerely making an effort many people will be willing to help you because they will trust your intentions and commitment to deliver. The music event generated the largest amount of funds and we were able to help the deserving kids.
Good (4) Item 5 + one of the points 2-4 above is not strongly emphasized	**----------**
Acceptable (3)	**Examples**

if the following elements are present:	
1. Shows a desire to make things better. 2. Item V might be absent 3. One element of proactive behavior (Item II-IV: identify opportunities, take initiative, and persist) are not equally strong. For instance; - Comes up with ideas, makes some effort to get them implemented but shows no evidence of facing opposition.	My school was organizing a charity event. One of my friends asked me to help him in arranging a music event. I decided to help him. I stood by my commitment. It was difficult with my other school activities but all the effort was worth helping the kids in distress; My school was organizing a charity event. I suggested we include a music show in the event. Everyone loved the idea and we worked towards making it a success.
Fair (2) Item V is absent Only two of the Item # 1-IV are shown (weakly).	----------
Poor (1) **if the following elements are present:**	**Examples**
1. Does not have a desire to make things better. 2. All three elements of proactive behavior (identify opportunities, take initiative, and persist) are weak. For instance; - Waits for others to highlight what needs to be done and barely does what is expected of him/her. - Comes up with lots of ideas but never takes the initiative to implement those. - Suggests some actions that don't convey essential elements of proactive behavior. - Makes some clichéd statement	I have always helped in the community projects and charity events; I am an active member of various charity and community service organizations; I don't remember a specific event but I am dedicated to make a change; Once we were having a team meeting for a charity event. I suggested that we should meet daily instead of weekly. This change helped us improve a lot.

STUDY 3: A VALIDITY OF GENERAL INTERVIEW QUESTIONS

Abstract

The third study explores the validity of general type of interview questions. Despite research evidence that supports the use of PBDI- and SI-type questions, managers continue using general questions in the actual interviews. In this study, I propose how to make the general interview questions more effective. I propose that researchers can increase the validity of general questions when they have an *a priori* understanding of underlying constructs, they ground the interview design in relevant theory, and they make the intended dimensions transparent to the candidates. The results provide modest support for the validity of general interview questions.

Study 3: A Validity of General Interview Question

How can we divine the experiences of an individual without access to significant amounts of information concerning the performance of said focal individual? Unless endowed with superior intuition, most must rely upon information collected directly from the individual. The best way to collect information directly from an individual is through an interview. Many options exist with respect to what can be asked in an interview. Importantly, what type of questions should be asked during an employment interview? This question is closely related to the validity of employment interviews. In past, researchers have proposed various taxonomies of interview questions on the basis of the content of the questions (Campion, Palmer, & Campion, 1997; Janz, 1982; Salgado & Moscoso, 2002).

Overall, interview questions can be classified into the following seven major types. The first category includes relatively structured situational questions (Latham, Saari, Pursell, & Campion, 1980). Situational questions involve hypothetical situations, often posed as dilemmas. The candidates are expected to state how they would behave in these situations (e.g., "suppose a co-worker asked for your help but you are in the middle of writing a report that is due in two hours. How would you respond to this co-worker?"). The second major category of questions is known as patterned behavior description interviews (Janz, 1982). The behavioral questions probe how candidates actually behaved in the previous jobs when faced with situations similar to the current job (e.g., "tell me about a situation

where you helped resolve a conflict?"). These two types of questions have received considerable attention in studies related to structured employment interviews.

The third category includes credential questions which are designed to gather details about a candidate's qualifications, achievements, and biographical facts (e.g., "what was your major in the graduate studies?"). The fourth category gathers information regarding past job experiences and activities (e.g., "what were your major responsibilities as a shift manager?"). The fifth category involves job knowledge type questions which probe candidate's procedural knowledge (e.g., "what factor of safety is recommended for designing a steel footbridge?"). The sixth category includes willingness questions that explore a candidate's reaction to job conditions such as travel, working hour, physical environment, and field conditions (e.g., "would you be able to travel during the week days?").

The seventh category includes general or psychological type of questions. These questions prompt candidates to share information about them (e.g., "what are your major strengths as a team player?"). The general interview questions are of particular interest, as these can tap into a candidate's personality, beliefs, attitudes, opinions and motivation. Conway and Peneno (1999) conducted a study to investigate the validity of different types of questions and found that psychological questions tap into constructs that are different from those tapped by situational and behavioral questions. McDaniel and colleagues (1994) showed through a meta-analysis that psychological questions yield a mean validity of 0.29 for job performance (compared to 0.50 for situational interviews and 0.39 for behavioral

interviews), whereas when the criterion was training success the mean validity of psychological questions was estimated to be 0.40 (compared to 0.36 for behavioral questions, for situational questions the validity estimates were not available). It is important to note that the general or psychological questions included in the meta-analysis were unstructured. The encouraging validity estimates of unstructured general questions indicate the possibility of increasing the validity of these questions by introducing structure in these interviews.

In the present study, my aim is to explore how psychometric properties of general type of questions can be increased for adding validity to the interview process. Despite ample evidence of superior empirical validity of behavioral and situational interview, the employers continue to use general types of interview questions. Another reason to continue researching on the validity of general questions is favorable applicant reactions and superior face validity of general interview questions compared to that of situational and behavioral questions (Conway & Peneno, 1999).

An additional benefit of asking general questions is the possibility of assessing a wider domain of job-related constructs in an employment interview. Different interview types tap different interview constructs. For instance, some have argued that PBDI are more effective for higher-level jobs as these are less specific in content than the situational interviews (Huffcutt, Weekley, Wiesner, Groot, & Jones, 2001). When there is less specificity in a question, candidates are forced to rely on their own cognitive resources to understand the question and craft

a logical and appropriate answer. The general questions provide another interviewing technique where candidates are posed with questions that tap into maturity, experience, work values, and psychological attributes of the candidate. Thus, it is important to explore how we might incrementally improve interview validity.

I propose a four-prong strategy to improve the validity of general interview questions. First, the general interview questions should follow the structuring standards applicable for behavioral and situational interviews. Second, general questions should be designed to tap into theoretically relevant constructs. Third, the psychometric properties of these questions should be improved by making the underlying constructs transparent to the applicant. Fourth, these questions should be asked through a multiphase interview approach. I will explain these steps in detail in the following section.

My first step is related to the theoretical grounding of interviews. Campbell et al. (1993) proposed an antecedent-determinant-performance model of job performance. According to this framework, declarative knowledge, procedural knowledge, and motivation are three key proximal determinants of any job performance dimension. These determinants have various distal antecedents such as personality, ability, attitudes, beliefs, values, and effort. The past behavioral interview questions and situational questions are designed to primarily elicit responses about job behaviors. Although, we can use these responses to make correlational inferences about antecedents and determinants of job performance in

the antecedent-determinant-performance model, the first two elements of the model can't be directly measured by behavioral or situational questions.

Conway and Peneno (1999) showed that general questions tap constructs such as personality that are different than those tapped by situational and behavioral questions. This indicates that general questions can be designed to assess predictor constructs such as attitudes, belief, values, and personality in order to develop a broader understanding of a candidate's potential to perform well on the job. The situational and behavioral description interviews are generally designed to tap key performance dimensions or constructs in the criterion space. However, general interview questions provide the flexibility to directly elicit candidate's response regarding job relevant antecedents and determinants.

Conway and Peneno (1999) used general questions that were standardized and had a scoring guide. I propose that theoretical grounding of general questions is another important step in improving interview validity. This essentially requires designing general questions to tap specific job-related constructs derived from a job analysis. In other words, by theoretical grounding of general questions, an interviewer will know a priori what constructs are being measured by the general type of questions, and how these constructs fit into the nomological network of job performance.

The third step concerns the transparency of interview dimensions. Following research in the field of assessment centers, some researchers in employment interviews have shown that transparency of interview questions

improves the construct validity of interviews (Klehe, Konig, Richter, Kleinmann, & Melchers, 2008). Psychological or general questions are considered to be open-ended in nature. In other words, a job candidate can provide a wide variety of responses. This creates difficulty in assessing the responses to these questions. By making the constructs of these questions transparent to the candidate, an applicant is provided a narrower domain of appropriate responses to focus upon. This should result in some consistency in interview responses provided by candidates and thus improve the structure of the interview.

For example, "where do you see yourself ten years from now?" is one the most commonly asked interview question (Christie, 2009; CNN.com, 2005; Doyle, 2009; Lee, 2008). An interviewer needs to understand why this question is being asked. The second step regarding theoretical grounding ensures that interviewers understand the purpose of the questions. However, a candidate might not understand the purpose of this question. If a candidate is informed that this question intends to assess his or her career focus or long-term planning skills, or both, an interviewer might obtain a more focused and relevant answer. Overall, theoretical grounding and transparency help eliminate those questions which lack focus and clarity (e.g., "Tell me something about you." – another commonly asked question). Thus, I propose that theoretical grounding and transparency should improve the psychometric properties of general interview questions thereby increasing its validity.

As a fourth step, to increase the validity of interview questions, I propose using a multiphase approach (i.e., different question types asked in different interview sessions) instead of a multimodal approach (different types of interview questions asked in the same interview session). Schuler & Funke (1989) proposed that interviews are a multimodal process. In others words, in order to tap the full domain of criterion constructs, the interviewers needs to ask different types of questions. This approach was suggested based on the evidence that different types of interviews tap different constructs. For instance, situational interviews have shown to be related to maximum performance, whereas, patterned behavioral description interviews (PBDI) have shown to capture typical performance (Klehe & Latham, 2006). Similarly, PBDI are correlated with motivation and experience and situational interviews tap job knowledge (Conway & Peneno, 1999). Thus, in a multimodal interview an interviewer asks different types of questions in a single interview so as to tap different predictors of job performance.

However, one of the issues that plague the multimodal approach is common method bias. Past research has shown that similar to exercise-effects found in assessment centers research (Arthur & Day, 2011) interviews have methods effect (Van Iddekinge, Raymark, Eidson, & Attenweiler, 2004). For instance, in Conway and Peneno's (1999) study, although they expected situational and behavioral interviews to tap different constructs, they found high correlations (0.85) between the two sets of questions. The high correlation might be due to methods effect as both types of questions were asked in a single interview by the same interviewers.

However, in this study, the general questions were asked in a separate session, and the correlation of general questions with the situational questions (.38) and the behavioral questions (0.31) was much smaller. Although different interview questions can potentially assess different constructs, in a multimodal interview, methods effect confound these differences.

This problem can potentially be resolved by designing multiphase interviews. A multiphase interview can be designed using two approaches – part multiphase and true multiphase. In a part-multiphase approach, an interview can be divided in several parts e.g., behavioral questions followed by general questions. In such a design, an interview panel can be divided such that different interviewers are assigned to different parts. Each interviewer will ask and rate the set of questions assigned to her respective parts. Although, this approach should eliminate the direct effects of different types of questions over each other, it will not guarantee that an interviewer's cognitive processes and judgments are free of any unconscious evaluation of the non-assigned interview part. Nonetheless, this approach might be more practical or have higher utility as only one interview session is required decreasing time and resources commitment.

In the true multiphase approach, different types of interview questions are asked in independent stand-alone interviews sessions. In the past, some studies have used this technique for a comparison of different interview questions (e.g., Conway & Peneno, 1999) – though not terming the approach as true multiphase. However, no study examines the effects of multimodal interviews with multiphase

interviews. This is unfortunate, given the potential of multiphase interviews to resolve several methodological issues in the interview research. For instance, use of multiphase interviews enables one to design less constructs per interview (in a true-multiphase approach) or per interviewer (in a part-multiphase approach). In 1965, after an extensive review of the employment interview literature, Ulrich and Trumbo (1965) recommended using less constructs per interview to decrease response evaluation burden on interviewers. Many researchers have done further analysis over the years, and have reached the similar conclusion (e.g., Van Iddekinge, et al., 2004). Hence, general questions asked in a multiphase manner can tap into distal antecedents (such as personality, values, beliefs, attitudes) of job performance. The multiphase design minimizes methods bias and thus, creates a stronger test of the hypothesis that structured general interview questions add incremental validity.

Overall, I propose that when general interview questions are designed on a theoretical basis, asked in a transparent and a multiphase manner, these questions should add incremental validity to the interview process.

> ***Hypothesis 1:*** *Theoretically grounded structured general interview questions have criterion-related validity.*
>
> ***Hypothesis 2:*** *Transparency of general questions is positively related to the validity.*
>
> ***Hypothesis 3:*** *The structured general questions will add incremental validity to the behavioral questions.*

Methods

Participants

The subjects of this study were applicants for a selective undergraduate leadership training program of a business school at a large mid-western University. Every fall semester, the leadership program selects aspiring associates. The leadership program conducts around 75 to 80 interviews each year and selects 50 to 60 qualified associates. Once selected, the program associates participate in various developmental activities throughout their two to three years stay in the program. The leadership associates must complete a mix of obligations each semester which are scored on a point system. On average, an associate is expected to invest around 30-35 hours per semester in the leadership program. The program activities include leadership team meetings, panel discussions with the leading professionals, large seminar-style events, business lunches, professional developmental workshops, and community service. Each program associate is expected to assume, at least once, a leadership role in organizing the training program activities. For continuing their enrollment in the program, an associate must maintain the minimum program requirements (i.e., maintaining a minimum GPA of 3.3 and completing minimum participation hours in the training program.) The highest performing associates are also offered the opportunity to serve on the student advisory board.

Out of 79 candidates this year, a total of 57 students agreed to participate in my study. A one-way between subjects ANOVA was conducted to compare the

differences between the research-subjects and non-research-subjects in terms of different bio-data characteristics. At an alpha level of .05 level, the two groups were not significantly different in terms of resume quality [F (1, 77) = 1.1, p = .30], essay writing skills [F (1, 77) = .08, p = .78] and recommendation letter scores [F(1, 77) = .05, p = .82]. The typical interviewee in the study was 20 years old. Except for one interviewee, each participant had been through at least one formal selection interview (M = 4.2, SD = .48).

Interview development

I utilized the critical incident analysis technique to identify the key performance behaviors shown by successful associates in the training program. The input for job analysis was provided by eight former and current leadership program students through a survey. These students provided approximately 50 critical incidents describing training performance. In addition, several face-to-face meetings were also conducted with the program director to understand the behaviors of the successful students in the training program. To further understand the desirable training behaviors, I also examined the program brochures and communication materials. On the basis of job analysis, I identified the following three key performance dimensions: organizing and planning behavior, proactive behaviors and proactive self-development behavior.

Klimoski (1993) suggested developing a theory-driven conceptualization of performance domain while designing selection tests. Following this suggestion, I used the theory of job performance (Campbell, McCloy, Oppler, & Sager, 1993) to

explore the theoretical model of individual differences influencing the identified critical performance dimensions (i.e., criterion domain). In the design of general interview, I focused on assessing the predictors related to personality and motivation. The rating scales for these questions were also defined by utilizing the information gained through the job analysis and reviewing the extant literature.

The extant literature was reviewed to explore the distal predictors of the key job behaviors. The planning and organizing behavior includes setting objectives, planning, managing time, managing resources and monitoring progress (Bartram, 2005). Researchers have shown that planning and organizing behaviors are correlated positively with conscientiousness (Fagenson-Eland & Baugh, 2001; Robertson, Baron, Gibbons, MacIver, & Nyfield, 2000). A proactive personality (Parker, Williams, & Turner, 2006; Seibert, Kraimer, & Crant, 2001), need for achievement, need for domination (Bateman & Crant, 1993; Fagenson-Eland & Baugh, 2001), core self-evaluations (Bono & Colbert, 2005), and goal orientations (Porath & Bateman, 2006) have been shown to be potent predictors of proactive self-development and proactive behaviors. Moreover, the literature on learning and training success have shown that cognitive ability (Ree & Earles, 1991), conscientiousness (Busato, Prins, Elshout, & Hamaker, 2000; Duff, Boyle, Dunleavy, & Ferguson, 2004), goal orientations (Colquitt, LePine, & Noe, 2000; Colquitt & Simmering, 1998; Wolters, Yu, & Pintrich, 1996), long-term career focus (Noe & Schmitt, 1986) and need for achievement (Bartels, Magun-Jackson,

& Ryan, 2010; Loon & Casimir, 2008) influence outcomes in learning and training situations.

As a second step, I reviewed the general interview questions commonly asked by practitioners. Various popular career building and job search advisement columns published by CNN.com, Workforce, and Monster.com regularly provide lists of common selection interview questions (e.g., Christie, 2009; CNN.com, 2005; Doyle, 2009; Lee, 2008). I used these sources to explore suitable general interview questions for the training program. Based on these sources, I developed several general interview questions that might tap into the identified predictors (Career focus, goal orientations, need for achievement, need for dominance, extraversion, conscientiousness, and confidence).

The initial pool of interview questions was reviewed independently by two subject matter experts. For each question, transparent and non-transparent versions of the scripts were designed. First, the non-transparent scripts were shown to the experts. At this stage, these experts were not aware of the underlying predictors intended to be tapped by these questions. The subject matter experts were also requested to provide feedback on the face validity of these questions for a selection interview. The questions for which subject matter experts did not agree on the underlying dimensions were dropped. Second, the transparent scripts were reviewed and revised by the experts. The final version of the interview included four general interview questions (see Table 1). Finally, I developed an interview rating scale for each question. I reviewed the extant literature as well as used the

job analysis information for defining the behavioral rating scales. These rating scales were reviewed by one subject matter expert.

Insert Table 1 about here

Procedures

The general interviews were administered by a two-member panel comprising two trained researchers. There were a total of five trained researchers. One member of the panel served as a main interviewer who asked all the questions. The main interviewer was present in all the interview sessions and conducted all the interviews. The other member of the panel served as a note taker. The note taker could not remain constant due to intensive time commitments of the selection interviews – the interviews were scheduled to last three days. Each interview started with the main interviewer greeting the candidate. After a brief introduction, the interviewer started with a control patterned behavior question which was part of another study. This was followed by asking the four general interview questions. A total of four interview conditions were defined. In each condition the order and the transparency of questions were varied. The candidates were randomly assigned to these four conditions. To maintain the highest degree of structure probing questions were not allowed. The interviewer was provided the instructions on the standardized format of the interview and the behavioral rating keys a week in advance. After dismissing the candidate, the interviewer rated the candidates on all

the interview questions. All candidates completed a post-interview questionnaire after being dismissed from the interview for a manipulation check.

Measures

Personality. Personality was measured using the 50-item IPIP scale for the Big-5 personality factors (Goldberg et al., 2006). Each subscale consists of 10 items rated on a 5-point likert type scale that ranges from strongly disagree (1) to strongly agree (5). The reliability estimates for the five factors were as follows: emotional stability (α = .80), extraversion (α = .85), conscientiousness (α = .85), agreeableness (α = .63), and openness (α = .74)[3].

Core self-evaluations. Consistent with previous research (e.g., Judge, Bono, Erez, & Locke, 2005; Judge, Locke, Durham, & Kluger, 1998), core self-evaluations was measured with four sub scales. The first trait – locus of control – was measured by an 11-item scale designed by Ghorpade, Hattrup and Lackritz (1999). A sample item includes "I think that life is mostly a gamble." The responses were based on a scale ranging from 1 (strongly disagree) to 7 (strongly agree) (α = .78). The second sub trait – neuroticism – was measured with a 10-item IPIP scale included in the Big-5 personality factors (Goldberg, et al., 2006). Responses were based on a 1 (strongly disagree) to 5 (strongly agree) scale (α = .80). An example item includes "I often feel blue." The 10-item scale by Rosenberg

[3] For the 10-item openness to experience scale the reliability estimate was .58. These items were further analyzed by plotting a graph between average score and individual item scores. These plots indicated five items with outliers such that several subjects' individual scores on these items deviated from their average scores on all the items. It is possible that students viewed some of the items socially desirable as an applicant for a leadership program. These items were related to abstract and novel ideas e.g., "Enjoy hearing new ideas." Thus, all subsequent analysis was done by dropping these five items from the scale measuring openness to experience. A similar analysis on agreeableness was done with no clear evidence of the contamination of items.

(1965) was used to measure the third trait – self-esteem. Responses were based on a 1 (strongly disagree) to 5 (strongly agree) scale (a = .83). "On the whole, I am satisfied with myself" is one of the items in the scale. The fourth sub trait of generalized self-efficacy was measured with an 8-item scale proposed by Chen, Gully, and Eden (2001). Responses were based on a 1 (strongly disagree) to 5 (strongly agree) scale (a = .93). A sample item includes "Even when things are tough, I can perform quite well." Consistent with prior research (e.g., Erez & Judge, 2001; Judge, et al., 2005), core self-evaluations was treated as a higher order concept.

To investigate the validity and structure of the core self-evaluations concept, as a first step, a principal factor analysis was done using SAS 9.2 followed by a confirmatory factor analysis. Due to small sample size, the principal component analysis was done on the average scores of the four traits. The principal factor analysis identified a single factor solution. One-factor solution was also supported by the subsequent confirmatory factor analysis. The fit statistics for this model were as follows: $\chi 2$ (2, N = 57) = 5.93, ns; *RMR* = .07 *RMSEA* = .18; *GFI* = .95; *BFI* = .95; and *NFI* = .92. These results were used to calculate a factor score for the trait of core self-evaluations for each subject; these factor scores were used in all the subsequent analyses.

Goal orientations. The three goal orientations were measured using a 13-item scale developed (VandeWalle, Cron, & Slocum, 2001). The responses were based on a scale ranging from 1 (strongly disagree) to 7 (strongly agree). For

leaning goal orientation, α = .86. A sample item includes "I like classes that really force me to think hard." For performance prove goal orientation, α = .75, and a sample item includes "It's important that others know that I am a good student." Finally, for performance avoid goal orientation, α = .69, and a sample item includes, "I would rather drop a difficult class than earn a low grade."

Need for achievement. The need for achievement was measured using a 10-item IPIP scale (Goldberg, et al., 2006). The scale showed acceptable alpha level, α = .73. A sample item includes, "I do more than what's expected of me."

Need for dominance. The need for dominance was measured using a 10-item IPIP scale (Goldberg, et al., 2006). The scale showed acceptable alpha level, α = .85. A sample item includes, "I seek to influence others."

Career focus. Long-term career focus was measured using a 7-item scale. I had developed this scale for one of my research projects by modifying an existing purposefulness scale (Organ & Greene, 1974). In my earlier project the scale showed an acceptable alpha level = .76. The alpha levels were also acceptable in this study, α = .84. The seven items are as follows: I set specific career goals for myself; I see a definite pattern when I look back on things I have done; I look for opportunities that will help me progress towards my career goals; I've spent a lot of time thinking what line of work best suits me; I try to do things with a clear purpose and direction in mind; I know what I will be doing one year from now; I set specific goals for myself."

Others. I also obtained candidates' scores on other selection criteria including: two essays which were part of the program application, resume, and letter of recommendations. These scores were assigned to the candidates by the selection committee comprising program director and several faculty members. I also obtained information about candidates' grade point average and their scores on standard aptitude tests (i.e., ACT and SAT). I defined an overall variable "application" by averaging the scores across all bio-data variables that capture a candidate's quality and qualification. These variables included the following: grade point average, resume, letter of recommendations, and application essays.

Face-to-face structured general interview questions. After dismissing a candidate, the main researchers assigned the ratings to the candidate on each interview questions. These ratings were assigned according to the predetermined behavioral rating keys. The ratings were assigned on a scale that ranged from 1 to 5. (See Annexure 1 for the response assessment keys designed for all four questions.)

Training performance. The performance of selected candidates was measured by several methods. One indicator of performance was frequency of participation in different program activities. The participation score were measured at two different points in time: three months after the enrollment and six months after the enrollment in the program. The program director also provided assessment of selected candidates on a 6-item training performance scale measuring proactive behaviors (4 items) and organizing and planning behavior (2 items). The four

items for proactive behavior included the following: this person assumes leadership roles**;** this person takes initiatives that add value to the program; this person enthusiastically makes use of the self-development opportunities provided by the program activities; this person often consults you to seek performance feedback. The organizing and planning behavior was measured by the following two items "this person is punctual, regular, and always prepared" and "this person is committed and fulfills his/her tasks and duties." The internal consistency reliability for this scale were acceptable (α = .90). The program director also provided a ranking of all the associates in the following five performance quartiles: eightieth percentile and above, sixtieth percentile and above, fortieth percentile and above, twentieth percentile and above and below twentieth percentile.

To investigate the structure of the training performance construct, first principal factor analysis was done followed by a confirmatory factor analysis. The principal factor analysis identified a single factor solution. One-factor solution was also supported by a subsequent confirmatory factor analysis. The fit statistics for this model were as follows: $\chi 2$ (5, N = 38) = 3.45, p=0.63; *RMR* = .03 *RMSEA* = .00; *GFI* = .96; and *NFI* = .96. These results were used to calculate a factor score of training performance for each subject.

Results

Insert Tables 2, 3 and 4 about here

A one-way between subjects ANOVA was conducted to compare the effect of trainee type on training performance in research-subjects and non-research-subjects conditions. There was no significant effect of trainee type on training performance [F (1, 51) = .82, p = .37]. The first hypothesis stated that theoretically grounded general interviews questions have a criterion-related validity. A simple regression analysis was run to test this hypothesis. In this model, training performance was used as a dependent variable. To test the criterion-related validity, first, overall general interview score was calculated and used as an independent variable in the model. The model was also tested by using each question as a predictor. These models were also compared with a model that uses structured patterned behavior interview as a predictor (described in the second essay).

As shown in Table 3, none of the models were significant at α =. However, due to small sample size, I also reviewed η^2 and ω^2 effect sizes provided for the overall ANOVA. I followed Cohen's (1988) guidelines regarding η^2 and ω^2 effect size estimates (i.e., small =. 0.0099, medium = 0.0588, large = .1379). The upper limits for the confidence intervals show a large effect for the overall interview score (model 1: $R^{2\ =}$ .10 , F (1,37) = 3.80, p=.06, η^2 =.10, ω^2 = .07, 95% CI [.00, .29], and a medium effect for overall structured behavior interview scores (model 6: $R^{2\ =}$ .01 , F (1,37) = 0.58, p=.58, η^2 =.01, ω^2 = -.02, 95% CI [.00, .14]). The effect sizes of structured general interviews are comparable with those of structured behavioral interview. These results are encouraging for exploring these

relationships in larger size samples. Overall, this study does not provide enough evidence to support hypothesis 1a but the study is highly suggestive of the potential of structured and theoretically grounded general interview questions to add validity to the interview process.

Insert Tables 5 and 6 about here

The next hypothesis was related to the validity of general interview questions based on the transparency. To test this hypothesis simple regression was run on the transparent and non-transparent sub samples. These models used training performance as a dependent variable and individual questions as predictors. As shown in table 5, the second hypothesis was supported only for the third question (Question C :"What are three things that will help you to be successful in XYZ leadership training program?") For model 3 the statistics for the two condition are as follows: transparent condition, $R^{2\,=}$ .27 , F (1,19) = 6.65, p=.02, η^2 =.27, ω^2 = .22, 95% CI [.01, .52]; non-transparent condition, $R^{2\,=}$ .02 , F (1,17) = .28, p=..60, η^2 =.02, ω^2 =- .04, 95% CI [0, .25].

Although, the transparent condition had superior-criterion validity, the t-test results for this question showed that the interview scores in the two conditions were not statistically different from each other ($R^{2\,=}$ .01 , F (1,56) = 0.74, p=.39, η^2 =.01, ω^2 = - .00, 95% CI [0, .12]). In fact, for all four questions, the interview scores were not different in the two conditions. For all other questions transparency

did not show any significant influence on the criterion-related validity. In the full sample only question C had an F value greater than 1 with a *p* value = .0591 (See Table 3), and it is the only question (i.e., model 3b) that shows evidence for the influence of transparency condition on the test validity. This might indicate that in terms of the relevance of the questions, question C was superior to other questions. However, as shown in table 5, for model 3b the confidence intervals show a very broad range of effect sizes.

Insert Tables 7 and 8 about here

I also used MTMM approach suggested in previous research (e.g., Klehe, et al., 2008; Schuler, 1989) to test the influence of transparency on the construct validity of interview questions. Consistent with previous research, different interview dimensions were defined as different traits, different interview types (written behavioral, oral behavioral and general interview) as different methods in the analyses. The non-transparent condition showed a mean monotrait–heteromethod (MTHM; convergent) correlation of .05 and a mean heterotrait–monomethod (HTMM; discriminant) correlation of .13, as compared to a mean heterotrait–heteromethod (HTHM) correlation of -.01. These estimates show a poor convergent validity. In the transparent condition, these correlations were .17 (MTHM; convergent), .00 (HTHM; discriminant), and .02 (HTHM). This indicates an increase in the convergent and discriminant validity. This provides modest

support for construct validity of transparent condition. Overall, the criterion and construct validity tests indicate support for hypothesis 2.

To check manipulation for transparency, at the end of each interview, candidates were asked to rate all four questions on the degree of clarity. More specifically, the candidates were asked to provide their degree of clarity regarding what the interviewer was trying to assess through each question. As shown in table 6, the t-test for these ratings showed a significant effect of manipulation only for question D ("What drives you to take on new projects, initiatives, or challenges in life?"). On average, for question D, the participants had a greater degree of clarity for the transparent condition (M= 6.43, SE =.13) than for the non-transparent condition (M= 5.93, SE =.20). This difference was significant $t(47)=2.05$, p=.05. Moreover, it represented a medium effect, r=.29. However, it is important to note that for all other questions, the means for transparent conditions are larger than those for the non-transparent condition. Of particular interest is question A. For this question, the participants had a greater degree of clarity for the transparent condition (M= 6.07, SE =.12) than for the non-transparent condition (M= 5.43, SE =.31). This difference was not significant $t(35)=1.95$, p=.06. However, it did represent a medium effect, r=.31.

Overall, I can argue that transparency of questions helped candidates gain a certain degree of confidence regarding the purpose of the interview questions. Only question C showed superior criterion-related validity for a transparent condition. However, the manipulation check for this condition does not show a significant

effect of transparency. Probably the candidates in the non-transparent condition over-estimated their ability to understand the purpose of the question. However, in the transparent condition, additional explanations by the interviewer helped in eliciting the relevant responses from the candidates. These speculations could only be confirmed by eliciting open ended responses from the candidates regarding the purpose of each question. However, due to concerns for response burden, I did not ask for open-ended construct explanations. These issues can be explored in the future studies.

The third hypothesis stated that the theoretically grounded general interview questions should add incremental validity over the behavioral interview questions. As shown in table 3, the overall F test for the behavioral interview was not significant which does not warrant the test of this hypothesis. Thus, hypothesis 3 was not supported.

Discussion

This study was designed to explore the validity of structured general questions. These questions were designed to predict the performance of the selected associates in the training program. The behavioral assessment keys were designed based on job analysis and review of extant literature. The influence of transparency of questions on the validity on general interview questions was also explored in this study. Overall, I argued that theoretical grounding of the interview and making interview questions transparent and multiphase should increase the validity of

general interview questions. The results of the study provide modest support for my assertions.

In the first hypothesis, I stated that structured general interview questions are valid predictors of future performance of candidates. In a meta-analysis of employment interviews, McDaniel and colleagues (1994) reported a mean validity of .15 for general or psychological interview questions and that of .24 for structured interview questions. I expected that validity estimate of structured, job-relevant and theoretically grounded general interview question should improve over the mean validity estimate of $r = .15$. As shown in Table 2 the correlation between structured general interview and training performance is $r = .31$, $p = .0591$. This estimate is comparable with the mean validity estimate of .35 reported for the structured behavioral questions in a recent meta-analysis (Taylor & Small, 2002). Thus, it provides encouraging support to explore structured general interviews as valid predictors of performance in future studies.

However, neither the overall regression model that includes general interview questions as a predictor nor the model using structured behavioral questions as a predictor of training performance was significant. One primary reason for a failure to detect a significant relationship might be an insufficient power or sensitivity of the design. According to Cohen's (1988) power analysis tables, a design with a power ($1-\beta$) of .80 and an alpha level (α) of .05 can detect an effect size (r) of .35 when minimum sample size (N) is 60, whereas for $r = .30$ the

minimum sample size is 80. In my study, the sample size is 38 which can only detect an effect size of .50 or above.

The second hypothesis explored the influence of transparency on the validity of general interview questions. Each candidate was asked two of the four general interview questions in a transparent condition and the other two in a non-transparent condition. The effect of transparency was analyzed question-wise by dividing the data in a transparent and a non-transparent condition. Table 5 shows results of these regression models. As shown in table 5, transparency showed positive influence on the interview validity only for question C. I also conducted MTMM analyses that showed improvement in the convergent and the discriminant validities of transparent questions over non-transparent questions (See Tables 7 and 8). This provides modest support for hypothesis 2. I also conducted t-tests to check the difference between the mean interview scores in the two conditions for all four questions. The candidates did not score higher in either of the two interview conditions. Thus, transparency does not influence candidates to provide more or less desirable answers. The failure to detect the influence of transparency in the other three questions can either be due to a small sample size or due to irrelevancy of the interview questions in predicting performance.

The check of manipulations for transparent and non-transparent conditions showed moderate effect sizes for question A and question D (see Table 6). In the other two questions the sample mean for the transparent condition was always greater than the non-transparent condition. However, the results of manipulation

test do not align with the results of regression for hypothesis 2 (Compare table 5 and table 6). More specifically, for questions A and D, the t-test shows that transparency increases candidates understanding of what an interviewer is asking but for these questions the results of regression analysis do not show an increase in the criterion-related validity with transparency. On the other hand, for question C, the t-test does not show significant mean differences in understanding the purpose of the question in the two conditions but the results of regression show an improvement in the criterion-validity with transparency. This lack of congruence between the two tests might indicate that the candidates do not always perceive a discrepancy between what is being asked by the interviewer and what is being perceived by them. They can either over-estimate or under estimate their ability to understand the purpose of the question. It is also possible that some questions are generally better understood and gain marginal benefits by transparency. One way of further exploring these issues is by asking candidates to write the purpose of each question during the manipulation check. I was not able to design these questions due to response load on the candidates.

To sum up, this study provides modest support for the validity of structured general interview questions. As there is some support for validity and transparency of structured general interview questions, it should encourage researchers to further explore these issues in a larger sample size. I believe it would be rather premature to dismiss the relevance and utility of general interview questions in the selection research at this stage. The flexibility and practical appeal of general interview

questions furnishes impetus for researchers to explore how to make the design and assessment of these questions more scientific. I believe my study is an important step in this direction and will facilitate our efforts to divine information about job candidates.

REFERENCES

Arthur, W., Jr., & Day, E. A. (2011). Assessment centers. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology, Vol 2: Selecting and developing members for the organization.* (pp. 205-235). Washington, DC: American Psychological Association.

Bartels, J. M., Magun-Jackson, S., & Ryan, J. J. (2010). Dispositional approach-avoidance achievement motivation and cognitive self-regulated learning: The mediation of achievement goals. *Individual Differences Research, 8*(2), 97-110.

Bartram, D. (2005). The Great Eight Competencies: A Criterion-Centric Approach to Validation. *Journal of Applied Psychology, 90*(6), 1185-1203.

Bateman, T. S., & Crant, J. M. (1993). The proactive component of organizational behavior: A measure and correlates. *Journal of Organizational Behavior, 14*(2), 103-118.

Bono, J. E., & Colbert, A. E. (2005). UNDERSTANDING RESPONSES TO MULTI-SOURCE FEEDBACK: THE ROLE OF CORE SELF-EVALUATIONS. *Personnel Psychology, 58*(1), 171-203. doi: 10.1111/j.1744-6570.2005.00633.x

Busato, V. V., Prins, F. J., Elshout, J. J., & Hamaker, C. (2000). Intellectual ability, learning style, personality, achievement motivation and academic success of psychology students in higher education. *Personality and Individual Differences, 29*(6), 1057-1068. doi: 10.1016/s0191-8869(99)00253-6

Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1993). A theory of performance. In N. Schmitt, W. C. Borman & Associates (Eds.), *Personnel selection in organizations* (pp. 35-70). San Francisco, CA: Jossey-Bass Publishers.

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

Chen, G., Gully, S. M., & Eden, D. (2001). Validation of a New General Self-Efficacy Scale. *Organizational Research Methods, 4*(1).

Christie, I. (2009). Common Interview Questions. *Monster* Retrieved Dec 22, 2009, from http://career-advice.monster.com/job-interview/Interview-Questions/Common-Interview-Questions-1/article.aspx

CNN.com. (2005). Answering 6 common interview questions. *Career.Builder.com* Retrieved Dec 22, 2009, from http://www.cnn.com/2005/US/Careers/12/09/six.questions/index.html

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, N.J.: L. Erlbaum Associates.

Colquitt, J. A., LePine, J. A., & Noe, R. A. (2000). Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology, 85*(5), 678-707. doi: 10.1037/0021-9010.85.5.678

Colquitt, J. A., & Simmering, M. J. (1998). Conscientiousness, goal orientation, and motivation to learn during the learning process: A longitudinal study. *Journal of Applied Psychology, 83*(4), 654-665.

Conway, J. M., & Peneno, G. M. (1999). Comparing structured interview question types: Construct validity and applicant reactions. *Journal of Business and Psychology*, *13*(4), 485-506.

Doyle, A. (2009). Interview Questions and Best Answers. *About.com* Retrieved Dec 22, 2009, from http://jobsearch.about.com/od/interviewquestionsanswers/a/interviewquest.htm

Duff, A., Boyle, E., Dunleavy, K., & Ferguson, J. (2004). The relationship between personality, approach to learning and academic performance. *Personality and Individual Differences, 36*(8), 1907-1920. doi: 10.1016/j.paid.2003.08.020

Erez, A., & Judge, T. A. (2001). Relationship of Core Self-Evaluations to Goal Setting, Motivation, and Performance. *Journal of Applied Psychology, 86*(6), 1270-1279.

Fagenson-Eland, E. A., & Baugh, S. G. (2001). Personality Predictors of Protégé Mentoring History. [Article]. *Journal of Applied Social Psychology*, *31*(12), 2502-2517.

Ghorpade, J., Hattrup, K., & Lackritz, J. R. (1999). The Use of Personality Measures in Cross-Cultural Research: A Test of Three Personality Scales Across Two Countries. *Journal of Applied Psychology, 84*(5), 670-679.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and

the future of public-domain personality measures. *Journal of Research in Personality, 40*(1), 84-96.

Huffcutt, A. I., Weekley, J. A., Wiesner, W. H., Groot, T. G., & Jones, C. (2001). Comparison of situational and behavior description interview questions for higher-level positions. *Personnel Psychology, 54*(3), 619-644.

Janz, T. (1982). Initial comparisons of patterned behavior description interviews versus unstructured interviews. *Journal of Applied Psychology, 67*(5), 577-580.

Judge, T. A., Bono, J. E., Erez, A., & Locke, E. A. (2005). Core Self-Evaluations and Job and Life Satisfaction: The Role of Self-Concordance and Goal Attainment. *Journal of Applied Psychology*, *90*(2), 257-268.

Judge, T. A., Locke, E. A., Durham, C. C., & Kluger, A. N. (1998). Dispositional Effects on Job and Life Satisfaction: The Role of Core Evaluations. *Journal of Applied Psychology, 83*(1), 17-34.

Klehe, U.-C., Konig, C. J., Richter, G. M., Kleinmann, M., & Melchers, K. G. (2008). Transparency in Structured Interviews: Consequences for Construct and Criterion-Related Validity. *Human Performance, 21*(2), 107-137.

Klehe, U.-C., & Latham, G. (2006). What Would You Do - Really or Ideally? Constructs Underlying the Behavior Description Interview and the Situational Interview in Predicting Typical Versus Maximum Performance. *Human Performance, 19*(4), 357-382.

Latham, G. P., Saari, L. M., Pursell, E. D., & Campion, M. A. (1980). The situational interview. *Journal of Applied Psychology, 65*(4), 422-427.

Lee, K. S. (2008). Tackling Interview Questions That Put You to the Test. *Certification Magazine, 10*(10), 10.

Loon, M., & Casimir, G. (2008). Job-demand for learning and job-related learning: The moderating effect of need for achievement. *Journal of Managerial Psychology, 23*(1), 89-102. doi: 10.1108/02683940810849684

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

Noe, R. A., & Schmitt, N. (1986). The influence of trainee attitudes on training effectiveness: Test of a model. *Personnel Psychology, 39*(3), 497-523.

Organ, D. W., & Greene, C. N. (1974). The Perceived Purposefulness of Job Behavior: Antecedents and Consequences. *Academy of Management Journal, 17*(1), 69-78.

Parker, S. K., Williams, H. M., & Turner, N. (2006). Modeling the antecedents of proactive behavior at work. *Journal of Applied Psychology, 91*(3), 636-652.

Porath, C. L., & Bateman, T. S. (2006). Self-Regulation: From Goal Orientation to Job Performance. *Journal of Applied Psychology, 91*(1), 185-192.

Ree, M. J., & Earles, J. A. (1991). Predicting training success: Not much more than g. *Personnel Psychology, 44*(2), 321-332.

Robertson, I. T., Baron, H., Gibbons, P. J., MacIver, R., & Nyfield, G. (2000). Conscientiousness and managerial performance. *Journal of Occupational and Organizational Psychology, 73*, 171–180.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Salgado, J. F., & Moscoso, S. (2002). Comprehensive meta-analysis of the construct validity of the employment interview. *European Journal of Work & Organizational Psychology, 11*(3), 299-324.

Schuler, H. (1989). Construct validity of a multimodal employment interview. In B. J. Fallon, H. P. Pfister & J. Brebner (Eds.), *Advances in industrial organizational psychology* (pp. 343-354). Amsterdam, North-Holland: Elsevier Science Publishers.

Schuler, H., & Funke, U. (1989). The interview as a multimodal procedure. In R. W. Eder & G. R. Ferris (Eds.), *The employment interview: Theory, research, and practice* (pp. 183-192). Newbury Park, CA: Sage Publications.

Seibert, S. E., Kraimer, M. L., & Crant, J. M. (2001). What do proactive people do? A longitudinal model linking proactive personality and career success. *Personnel Psychology, 54*(4), 845-874.

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 & Organizational Psychology, 75*(3), 277-294.

Ulrich, L., & Trumbo, D. (1965). The selection interview since 1949. *Psychological Bulletin, 63*(2), 100-116.

Van Iddekinge, C. H., Raymark, P. H., Eidson, J. C. E., & Attenweiler, W. J. (2004). What Do Structured Selection Interviews Really Measure? The Construct Validity of Behavior Description Interviews. *Human Performance, 17*(1), 71-93.

VandeWalle, D., Cron, W. L., & Slocum, J. W. (2001). The role of goal orientation following performance feedback. *Journal of Applied Psychology, 86*(4), 629-640.

Wolters, C. A., Yu, S. L., & Pintrich, P. R. (1996). The relation between goal orientation and students' motivational beliefs and self-regulated learning. *Learning and Individual Differences, 8*(3), 211-238.

Table 1: General Interview Questions

General interview question	Underlying traits	Condition 1	Condition 2	Condition 3	Condition 4
Question A. Where do you see yourself ten years from now? Additional script for transparent question: This question is getting at your overall approach towards managing your long-term career goals.	Long-term career orientation or purposefulness	Transparent Order = 1st question	Non Transparent Order = 1st question	Non Transparent Order = 3rd question	Transparent Order = 3rd question
Question B. Why do you want to join XYZ leadership training program? Additional script for transparent question: This question is gauging your motivation to enroll in the leadership training program. Different students have different objectives for joining this program.	Goal orientation	Transparent Order = 2nd question	Non Transparent Order = 2nd question	Non Transparent Order = 4th question	Transparent Order = 4th question
Question C. What are three things that will help you to be successful in XYZ leadership training program? Additional script for transparent question: This question is getting at your skills, abilities, and traits that would enable you to be successful in this program	Key skills and abilities Confidence; Need for achievement; Need for dominance; Motivation to learn; Extraversion; Conscientious; Drive and determination	Non Transparent Order = 3rd question	Transparent Order = 3rd question	Transparent Order = 1st question	Non Transparent Order = 1st question
Question D. What drives you to take on new projects, initiatives, or challenges in life? Additional script for transparent question: This question is gauging your motivation to take on new activities. People have different objectives in view when they take up new activities	Goal orientation	Non Transparent Order = 4th question	Transparent Order = 4th question	Transparent Order = 2nd question	Non Transparent Order = 2nd question

Table 2: Mean Values, Standard Deviations, and Inter-correlations of Study Variables

Variables	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12
1. Age	19.66	0.94	1.00											
2. Gender	0.67	0.48	-0.06	1.00										
3. No of Interviews	4.20	2.91	0.25	-0.21	1.00									
4. Application	4.11	0.47	-0.19	-0.23	-0.06	1.00								
5. Openness	3.15	0.78	0.05	0.14	-0.08	-0.05	1.00							
6. Conscientiousness	4.06	0.51	**-0.28**	0.10	-0.07	-0.10	-0.04	1.00						
7. Extraversion	3.67	0.57	0.22	0.06	0.29	-0.25	-0.09	0.05	1.00					
8. Agreeableness	3.95	0.37	-0.14	**0.28**	-0.23	-0.07	-0.12	***0.58***	0.11	1.00				
9. Emotional stability	3.80	0.54	-0.02	-0.23	0.15	-0.06	-0.12	***0.37***	***0.52***	0.23	1.00			
10. Core self-evaluations	0.00	1.07	0.01	-0.12	0.06	-0.24	-0.17	0.18	***0.53***	0.03	***0.65***	1.00		
11. Overall general interview	2.90	0.57	-0.15	-0.06	-0.08	**0.31**	-0.06	0.08	-0.01	0.05	0.15	0.07	1.00	
12. Training performance	0.00	0.94	-0.12	-0.25	-0.27	0.29	-0.08	0.13	-0.27	0.12	0.09	0.19	0.31	1.00

- For all variables N= 57 except for the following: for Age and Number of interviews, N=56, for Training performance N=38
- For bold cells, $p < .05$; For bold and italicized cell, $p < .01$

Table 3: Training Performance and Interview Scores - Overall ANOVA

				Overall Non-centrality Parameter				Proportion of Variation Accounted for			
Model	*F*	*Pr>F*	*R-sq*	Minimum Variance Unbiased Estimate	Low MSE Estimate	95% Confidence Limits		η^2	ω^2	95% Confidence Limits	
1	3.80	0.06	0.10	2.59	2.44	0.00	15.58	0.10	0.07	0.00	0.29
2	0.64	0.42	0.02	-0.39	-0.37	0.00	7.63	0.02	-0.01	0.00	0.17
3	2.60	0.12	0.07	1.45	1.37	0.00	12.92	0.07	0.04	0.00	0.25
4	3.63	0.06	0.09	2.42	2.28	0.00	15.21	0.09	0.06	0.00	0.29
5	0.48	0.49	0.01	-0.55	-0.52	0.00	7.00	0.01	-0.01	0.00	0.16
6	0.60	0.49	0.01	-0.46	-0.41	0.00	7.50	0.03	-0.02	0.00	0.27
7	1.85	0.17	0.10	1.48	1.39	0.00	13.95	0.10	0.04	0.00	0.27

Notes

1. For all the models, total *df* = 37 and model *df* = 1; All the models have training performance as a dependent variable.
2. For model 1, the total general interview score is used as a predictor; for model 2, the score on question A is used as a predictor; for model 3, the score on question B is used as a predictor; for model 4, the score on question C is used as a predictor; For model 5, the score on question D is used as a predictor.
3. For model 6, the overall score of behavioral interview of the study 2 is used as a predictor.
4. For model 7, the overall score of behavioral interview and the score of general interview are used as predictors.

Table 4: Training Performance and Interview Scores – Parameter Estimates

	b	β	*SE*	*t*	*Pr* > |*t*|
Model 1					
Intercept	-1.44	0.00	0.75	-1.91	0.06
Total general interview score	0.49	0.31	0.25	1.95	0.06
Model 2					
Intercept	-0.42	0.00	0.54	-0.77	0.45
Question A	0.14	0.13	0.17	0.80	0.43
Model 3					
Intercept	-0.68	0.00	0.45	-1.52	0.14
Question B	0.23	0.26	0.14	1.61	0.12
Model 4					
Intercept	-1.45	0.00	0.77	-1.87	0.07
Question C	0.44	0.30	0.23	1.90	0.06
Model 5					
Intercept	-0.21	0.00	0.35	-0.62	0.54
Question D	0.08	0.11	0.12	0.69	0.49
Model 6					
Intercept	-0.40	0.00	0.72	-0.55	0.58
Total behavioral interview score	0.15	0.09	0.27	0.57	0.58

Notes

1. For all the models, total *df* = 37 and model *df* = 1; All the models have training performance as a dependent variable.
2. For model 1, the total general interview score is used as a predictor; for model 2, the score on question A is used as a predictor; for model 3, the score on question B is used as a predictor; for model 4, the score on question C is used as a predictor; For model 5, the score on question D is used as a predictor; For model 6, the overall score of behavioral interview of the study 2 is used as a predictor.

Table 5: Training Performance and Transparent-Non-Transparent Interview Scores - Overall ANOVA

				Overall Non-centrality Parameter				Proportion of Variation Accounted for			
Model	*F*	*Pr>F*	*R-sq*	Minimum Variance Unbiased Estimate	Low MSE Estimate	95% Confidence Limits		η^2	ω^2	95% Confidence Limits	
1a	3.37	0.08	0.16	1.99	1.74	0.00	14.90	0.16	0.11	0.00	0.43
1b	0.02	0.88	0.00	-0.98	-0.84	0.00	3.12	0.00	-0.06	0.00	0.15
2a	2.70	0.12	0.13	1.40	1.22	0.00	13.36	0.13	0.08	0.00	0.40
2b	1.47	0.24	0.08	0.28	0.24	0.00	10.21	0.08	0.03	0.00	0.36
3a	0.28	0.60	0.02	-0.76	-0.65	0.00	6.05	0.02	-0.04	0.00	0.25
3b	6.65	0.02	0.27	4.91	4.30	0.12	21.90	0.27	0.22	0.01	0.52
4a	0.28	0.76	0.04	-1.52	-1.28	0.00	5.22	0.04	-0.09	0.00	0.22
4b	0.33	0.72	0.04	-1.41	-1.23	0.00	5.73	0.04	-0.07	0.00	0.22

Notes:

1. For all the models, total *df* = 19 and model *df* = 1; All the models have training performance as a dependent variable.
2. For model 1, the score on question A is used as a predictor; for model 2, the score on question B is used as a predictor; for model 3, the score on question C is used as a predictor; For model 4, the score on question D is used as a predictor.
3. For the models with a suffix “a” the interview scores represent a non-transparent condition, and for the models with suffix “b” the interview scores represent a transparent condition.

Table 6: Manipulation Check for Transparency: Results for the Independent T-Test

Question	Transparent		Non-Transparent			95%					
	M	*SE*	*M*	*SE*	Difference	Confidence level		*t*	*Df*	*Pr* > |*t*|	*r*
A	6.07	0.12	5.43	0.31	0.64	-0.03	1.31	1.95	35.25	0.06	0.31
B	6.66	0.10	6.50	0.12	0.16	-0.16	0.47	0.98	55.00	0.33	0.13
C	6.18	0.23	6.07	0.15	0.11	-0.46	0.68	0.39	47.15	0.70	0.06
D	6.43	0.13	5.93	0.20	0.50	0.01	0.98	2.05	47.33	0.05	0.29

Table 7: MTMM Matrix – Transparent questions

	Method 1: General Interview				Method 2 - Written Behavioral Interview			
	Question A	Question B	Question C	Question D	Self-Development	Mentoring	Planning	Challenging Project
Question A	1.00	0.36	-0.02	-0.12	0.07	-0.02	0.08	-0.01
	-	0.05	0.93	0.55	0.70	0.90	0.67	0.94
Question B	0.36	1.00	0.34	0.21	0.13	-0.34	0.06	-0.23
	0.05	-	0.07	0.27	0.50	0.07	0.74	0.23
Question C	0.05	0.49	1.00	-0.37	0.18	-0.25	0.22	-0.21
	0.80	0.01	-	0.05	0.35	0.21	0.26	0.29
Question D	-0.17	0.10	-0.37	1.00	-0.22	0.22	0.02	0.23
	0.39	0.61	0.05	-	0.26	0.28	0.92	0.23

	Method 3: Self-Report Measures							
	Career Focus	Avoidance Orientation	Learning Orientation	Performance Orientation	Extraversion	Need for Achievement	Need for Dominance	Core Self-Evaluations
Question A	0.28	0.42	-0.09	0.29	-0.02	-0.13	0.28	-0.07
	0.14	0.02	0.64	0.12	0.94	0.51	0.14	0.70
Question B	0.14	0.49	-0.05	-0.06	0.07	0.09	0.09	0.25
	0.46	0.01	0.80	0.77	0.70	0.65	0.65	0.20
Question C	-0.02	0.05	-0.12	0.10	0.20	0.14	0.05	0.34
	0.92	0.79	0.54	0.61	0.31	0.48	0.79	0.08
Question D	0.18	0.10	-0.07	0.05	-0.09	0.08	-0.11	-0.09
	0.37	0.62	0.73	0.79	0.65	0.68	0.57	0.65

Table 7: *Continued*

Notes:

- For questions A and B, $N = 29$; For questions C and D, $N = 28$
- In the correlation tables, for any variable the values in the first row represent the correlations and those in the second row show *p* values.
- Question A measures long-term career focus
- Question B measures motivations to join the training program or goal orientation.
- Question C measures the traits of extraversion, achievement and dominance focus, conscientiousness, and confidence
- Question D measures motivations to undertake challenging projects.

Table 8: MTMM Matrix – Non-Transparent questions

	Method 1: General Interviews				Method 2 - Written interview			
	Question A	Question B	Question C	Question D	Self-Development	Mentoring	Planning	Challenging Project
Question A	1.00	0.05	0.05	-0.17	0.27	-0.19	-0.31	0.21
	-	0.81	0.80	0.39	0.17	0.33	0.10	0.29
Question B	0.05	1.00	0.49	0.10	0.05	-0.11	0.28	-0.18
	0.81	-	0.01	0.61	0.80	0.57	0.15	0.37
Question C	-0.02	0.34	1.00	0.03	0.02	0.17	0.24	-0.01
	0.93	0.07	-	0.89	0.91	0.38	0.20	0.95
Question D	-0.12	0.21	0.03	1.00	-0.04	-0.18	-0.19	-0.02
	0.55	0.27	0.89	-	0.82	0.36	0.32	0.91

	Method 3: self-report measures							
	Career Focus	Avoidance Orientation	Learning Orientation	Performance Orientation	Extraversion	Need for Achievement	Need for Dominance	Core Self-Evaluations
Question A	-0.14	-0.13	-0.05	-0.21	-0.22	0.01	-0.34	0.18
	0.46	0.50	0.81	0.28	0.26	0.94	0.08	0.36
Question B	-0.02	-0.05	-0.19	-0.06	0.08	0.22	-0.07	0.40
	0.92	0.79	0.33	0.78	0.68	0.27	0.71	0.03
Question C	0.14	0.25	0.01	0.11	0.34	0.18	0.25	0.14
	0.46	0.19	0.95	0.58	0.07	0.36	0.18	0.45
Question D	-0.11	0.10	-0.29	0.04	0.04	0.02	0.19	0.09
	0.57	0.59	0.13	0.82	0.83	0.90	0.34	0.65

Table 8: *Continued*

Notes:

- For questions A and B, $N = 28$; For questions C and D, $N = 29$
- In the correlation tables, for any variable the values in the first row represent the correlations and those in the second row show *p* values.
- Question A measures long-term career focus
- Question B measures motivations to join the training program or goal orientation.
- Question C measures the traits of extraversion, achievement and dominance focus, conscientiousness, and confidence
- Question D measures motivations to undertake challenging projects.

Annexure 2: Assessment Key General Interview Questions

Ask all candidates the exact same question in a uniform manner:

5. Same person should ask these questions.
6. These questions should be asked in the same order.
7. No probing of questions from candidates.
8. However, you may repeat or clarify if a candidate does not understand the question.

QUESTION A: Where do you see yourself ten years from now?

Dimension: Long-term planning and purposefulness *(A purposeful and thoughtful management of concrete long-term goals.)*

Excellent (5) if the following elements are present:	**Examples**
a) Strong goal: The candidate must have a clear Job (level + specialization) and Industry/Specialization mentioned (applies to owning a business too- the business should be specific). (specific means - some details about domain/area of interest/e.g., HR manager in service industry) and should also mention at least one of the following: b) Mention how they plan to achieve that. c) Say why they have such goals d) some idea of career progression.	Ten years from now, I would like to be a manager in a public accounting firm and would like to be a few years from making partner; I am planning to attend MBA school after I graduate, ideally in the West Coast, i.e. California. Hopefully after a decade I will be able to use my MBA in the music industry in the West Coast, such as at a record company or artist management firm
Good (4) Strong goal: The candidate must have a clear Job and Industry/Specialization mentioned (applies to owning a business too). (specific means - some details about domain/area of interest/e.g., HR manager is service industry)	Practicing International Law in San Francisco, New York, or Europe; Dean of students for a law school, working at a law firm, or working toward becoming a judge; Director of human resources in large or federal company; As an executive in the marketing/sales side of a leading renewable energy firm or other

	Fortune 500 company. In addition, I expect to have started at least one business by then.
Acceptable (3) **if the following elements are present:**	**Examples**
The candidate mentions the type of work or company, but only one element is specific. OR moderate goal: The candidate must have a clear Job or "vague job and clear Industry"(applies to owning a business too).	Executive level management at a large corporation in the oil and gas industry; work as a programmer; Marketing director at a firm
Fair (2)	I want to work in a multinational, work in sales, work in IT; I see myself working in Oklahoma in the energy industry; I see myself managing people more than things. I see myself as an accountable decision maker in a company and vital.
Poor (1) **if the following elements are present:**	**Examples**
Vague or clichéd answer Or I want to work for a big company or want to work for a company that motivates them, recognizes their skills, or help them build skills, without saying which skills.	I want to be successful; I don't know; Hopefully fully employed; At this point, I'm split between family and career. Family will more than likely win.

QUESTION B: Why do you want to join JCPLP program?

Dimension: Motivation *(What type of motivation the person has. Learning, performance-prove or performance avoid.).*

Learning

Excellent (5) if the following elements are present:	**Examples**
a) Shows learning goal orientation by expressing an interest/enjoyment to assume a challenging or difficult task so as to improve. b) Also shows interest in developing at least 2 of the skills JCPLP develops - leadership skills, networking skills, teamwork, communication skills.	As a college student I want to learn and improve myself as much as I can by taking challenges. I have a roommate who is member of JCPLP, and upon discussion with him, I have come to know what a great program JCPLP is for a management student. I really want to improve my leadership and teamwork skills through this program,
Good (4)	As a college student I want to learn and improve myself as much as I can. . I really want to improve my teamwork skills through this program,
Acceptable (3) if the following elements are present:	**Examples**
a) Shows a general level learning orientation b) Also has some idea of what skills JCPLP develops.	I want to improve my skills as a manager such as how to lead; I want to learn things that are necessary to be a successful HR manager/leader.
Fair (2)	I want to learn as much as I can.
Poor (1) if the following elements are present:	**Examples**
Cliched No idea	My friends / instructor asked me to apply I don't know

Performance Prove

Excellent (5) if the following elements are present:	**Examples**
a) Shows performance prove	As a college student I want to

orientation by expressing an interest/enjoyment to showcase skills and competence. b) Also shows interest in showcasing at least 2 of the skills JCPLP develops - leadership skills, networking skills, teamwork, communication skills.	surround myself with people who are competent and driven. I have a roommate who is member of JCPLP, and upon discussion with him, I have come to know what a great program JCPLP is for a management student. I have good leadership and teamwork skills that make me fit for this program.
Good (4)	I want to prove to others that I have what it takes to be a leader and JCPLP exactly does this for price college students. It shows us how to be competent.
Acceptable (3) **if the following elements are present:**	**Examples**
c) Shows a general level performance-prove orientation. d) Also has some idea of what skills JCPLP develops.	I am a good leader and JCPLP will help me test my skills.
Fair (2)	I want to show my skills.
Poor (1) **if the following elements are present:**	**Examples**
Cliche No idea	My friends / instructor asked me to apply I don't know

QUESTION C: What are three things that will help you to be successful in the JCPenney Leadership program?

Dimension: Key JCPLP success factors *(Proactive behavior, self-development focus, learning orientation, and drive and determination).*

Excellent (5) **if the following elements are present:**	**Examples**
KEY JCPLP elements a. Proactive behavior (take initiative, brings change, do new things, ideas, loves challenges, spots opportunities), b. self-development focus, (learning orientation, growth need, seeks mentors, likes to learn new things, openness to experience) c. drive and determination, struggle, hardworking, persevere, conscientiousness, planning. MINOR may include one of the following - extraversion - leaderships experience - communication skills - networking skills - teamwork • All 3 from major • 2 major + 1 minor	I am very organized, I love challenges, and I am passionate about learning how to be a good leader. I understand that JCPLP is looking for students who want to be a leader.
Good (4) • 2 items from major + any other • 1 from major (a or b) + 2 minor	----------
Acceptable (3) **if the following elements are present:**	**Examples**
1 from major (a or b)+ any other OR 1 Major-c + 1/2 minor	
Fair (2) • Major c + NR OR 1-Major C (rep) • 2 minor • 3 minor	----------
Poor (1)	**Examples**

if the following elements are present:	
Only one minor Clichéd answer Vague answer	I have what it takes to be successful

QUESTION D: What drives you to take on new projects, initiatives, or challenges in life?
Dimension: Motivation *(What type of motivation the person has. Learning, performance-prove or performance avoid.).*

Learning

Excellent (5) **if the following elements are present:**	**Examples**
a) Shows learning goal orientation by expressing an interest/enjoyment to assume **a challenging or difficult** task so as to improve. (MUST mention improvement or learning) b) Mention an area of improvement OR past projects that he/she took on.	As a college student I want to learn and improve myself as much as I can by taking challenges. I really want to improve my leadership and teamwork skills through this program. **Keywords:** projects that push me, enjoy challenges, enjoy learning
Good (4) c) Shows learning goal orientation by expressing an interest/enjoyment to assume a challenging or difficult task so as to improve. (MUST mention improvement or learning) Or d) Shows learning goal orientation (without enjoyment/liking challenges) and also mentions an example	As a college student I want to learn and improve myself as much as I can. I take on projects that will help me develop skills or abilities. Keywords: same as above without any examples.
Acceptable (3) **if the following elements are present:**	**Examples**
e) MUST mention learning or improvement. Shows a general learning orientation, but does not mention either enjoying challenges or an example	

Fair (2) Shows a general level learning orientation in the form of an intrinsic interest in the project, but does not specifically mention development or learning as a goal. May suggest interest in new things or other indirect indication of learning goals.	I take on projects in areas that I'm passionate about; I enjoy challenges; I am driven.
Poor (1) **if the following elements are present:**	**Examples**
Cliched No idea No indication of learning goal.	My friends / instructor asked me to apply; I don't know

Performance Prove

Excellent (5) **if the following elements are present:**	**Examples**
a) Shows performance prove orientation by expressing an interest/enjoyment to showcase skills and competence. b) Also gives an example. c) Show extrinsic reasons for taking projects.	As a college student I want to prove myself to be competent and driven. I have good leadership and teamwork skills that make me fit for this program; Keywords: showing they are better than others (to someone), showing competence, proving oneself, getting ahead, attracting employer attention.
Good (4) d) Same as above, but NO example.	I want to prove to others that I have what it takes to be a leader: I take projects that showcase my skills; Keywords: Same as above. No examples given.
Acceptable (3) **if the following elements are present:**	**Examples**
e) Shows a general level performance-prove orientation, such as mentioning that they want to succeed. f) Mention approval of others.	It's good to hear others praise when I succeed at something; help me in my career; Keywords: getting praise, doing well, success (defined)

Fair (2) g) Any indirect indication of extrinsic motivation.	Keywords: want to succeed
Poor (1) **if the following elements are present:**	**Examples**
Cliche No idea	My friends / instructor asked me to apply I don't know