6-1-2014

The Relentless Pursuit Of Construct Validity In The Design Of Employment Interviews

Maria Riaz Hamdani
University of Akron, hamdani@akron.edu

Sorin Valcea
Cleveland State University, s.valcea@csuohio.edu

M. R. Buckley
University of Oklahoma

The relentless pursuit of construct validity in the design of employment interviews

Maria Riaz Hamdani [a,*], Sorin Valcea [b], M. Ronald Buckley [c]

[a] *Department of Management, College of Business Administration, The University of Akron, Akron, OH 44325-480, USA*
[b] *School of Business, Washburn University, 1700 SW College Ave, Topeka, KS 66621-1117, USA*
[c] *Division of Management and Entrepreneurship, Michael F. Price College of Business, The University of Oklahoma, 307 W. Brooks—Room 305 D, Norman 73019-0450 USA*

1. Introduction

Employment interviews have probably 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 techniques in human resources management research (e.g., Buckley, Norris, & Wiese, 2000). In spite of significant progress, much remains to be uncovered in order to fully understand the factors that influence how well employment interviews perform. As considerable support for the criterion-related validity of the employment interview has been demonstrated (Huffcutt & Arthur, 1994), many believe that construct validity is the next big puzzle to be confronted (Buckley & Russell, 1999; Macan, 2009). Nevertheless, some researchers consider this to be a vacuous endeavor (c.f. Harris, 1999), suggesting that the interview is a versatile instrument and can be used to measure almost any construct one wishes to investigate.

We believe that the pursuit of construct validity is neither elusive nor vacuous. In the last twenty years, efforts to create valid interviews have relied mostly on interview structure and job analysis. Grounding interviews in established human resource management theory and research could improve interview validity, but it is seldom done. Our primary objective is to provide concrete suggestions for future construct validity endeavors by focusing on theoretical grounding. Numerous researchers have highlighted the need for theoretical rigor in personnel selection research in general (Bartram, 2005; Binning & Barrett, 1989), and more specifically in assessment center research (Arthur, Day, & Woehr, 2008) and employment interview research (Harris, 1999; Landy & Shankster, 1994). However, we believe our approach to be unique because: (a) we specifically focus on employment interviews; this is the first attempt to systematically investigate myriad construct validity issues that plague the design of structured interviews; (b) we not only identify the gaps in the research, but also propose a comprehensive

* Corresponding author. Tel.: +1 330 972 5444.
  E-mail addresses: hamdani@uakron.edu (M.R. Hamdani), sorin.valcea@washburn.edu (S. Valcea), mbuckley@ou.edu (M.R. Buckley).

Table 1
Key design features of construct validity studies.

Publication	Study #[a]	N	Sample type	Focal job
1. Allen et al. (2004)	1	188	Students	A lab experiment
2. Conway and Peneno (1999)	2(I[b])	179	Students	Resident assistant
	2(II[b])	137	Students	
3. Huffcutt, Weekley, Wiesner, Groot, and Jones (2001)	3 (1)	59	Navy officers	Training program selection
	3(2)	93	Managers	Experiment
4. Klehe et al. (2008)	4(1)	123	Students	Mock interviews
	4(2)	269	Students	Mock interviews
5. Krajewski et al. (2006)	5	157	Applicants	High-level management
6. Motowidlo et al. (1992)	6(1)	107	Job applicants	Entry-level management
	6(2)	164	Managers	Mock interviews
	6(3)	195	Applicants	Entry-level management
	6(4)	176	Managers	Mock interviews
7. O'Leary (2004)	7	135	Police officers	Promotion to sergeant
8. Roth, Van Iddekinge, Huffcutt, Eidson, and Schmit (2005)	8(1)	164	Sales associates	Experiment
	8(2)	435	Applicants	Customer service manager
9. Schuler (1989)	9(1)	307	Applicants	Bank clerk apprentice
	9(2)	69	Students	Experiment
10. Sue-Chan and Latham (2004)	10	75	Executive MBAs	Mock interviews
11. Van Iddekinge, Raymark, Eidson, and Attenweiler (2004)	11	427	Applicants	Customer service manager
12. Van Iddekinge et al. (2005)	12	143	Students	Experiment

Study #	Key interview dimensions (Total number of dimensions)
1	Citizenships behaviors (1)
2(I)	Motivation for seeking job, expectations from the job, apprehensions (3)
2(II)	Role modeling, programming, helping, relationships, community development (8)
3 (1)	Two general leadership, two administrative and two interpersonal dimensions (6)
3(2)	Four action, three leadership and three interpersonal dimensions (10)
4(1)	Leadership, systematic planning, information management, cooperation (4)
4(2)	Leadership, systematic planning, cooperation (3)
5	Organizing, coaching, results orientation, willingness to learn, team orientation and oral communication (6)
6(1–3)	Leadership, assertiveness, flexibility, sensitivity, organization, thoroughness, resourcefulness, drive and communication skills (9)
6(4)	Persuasiveness, tact, service orientation, strategic skills and communication (5)
7	Oral communication, problem analysis, supervisory ability (3)
8(1)	Details not provided (10)
8(2)	No information available
9(1–2)	Information behavior, conscientiousness, verbal expression, initiative, support giving, coping with, errors, coping with complaints, calmness and sociability (9)
10	Teamwork behavior in MBA program (1)
11	Interpersonal behavior, conscientiousness, stress management (3)[c]
12	Altruism, self-discipline, vulnerability (3)

Study #	Question type	Question details	No. of questions	Rating key details (type)	Rating assignment
1	PBDI and SI	None	9	No (not clear)	Discuss and assign[d]
2(I)	General	Examples	8	Example (BARS)	Average of all ratings
2(II)	PBDI and SI	Examples	$7,7 = 14$[e]	Example (BARS)	Average of all ratings
3 (1)	PBDI and SI	None	$6,6 = 12$ [e]	None (not clear)	Average of all ratings
3(2)	PBDI and SI	None	$10,10 = 20$[e]	None (not clear)	Average of all ratings
4(1)	PBDI and SI	Examples	$12,12 = 24$[e]	Example (BARS)	Discuss and assign[f]
4(2)	PBDI and SI	None	$12,12 = 25$[e]	None (BARS)	Discuss and assign[f]
5	PBDI and SI	None	$2@6 = 12$[e]	None (BARS)	Average of all ratings
6(1)	PBDI	None	7	None (BARS)	Only one interviewer
6(2)	PBDI	None	7	None (BARS)	Only one interviewer
6(3)	PBDI	None	7	None (BARS)	Only one interviewer
6(4)	PBDI	None	7	None (BARS)	Only one interviewer
7	SI	Examples	3	Example (BARS)	Discuss and assign[f]
8(1)	SI	Brief labels	6	None (BARS)	Not clear
8(2)	PBDI	Brief labels	$2@7 = 14$[g]	None (BARS)	Consensus
9(1)	3 other and SI	None	$1,6,9,9 = 25$[e]	None (BARS)	Not clear
9(2)	2 other and SI	None	$1,18,19 = 38$[e]	None (BARS)	Average of all ratings
10	SI	Examples	6	Example (BARS)	Average of all ratings
11	PBDI	None	$2@7 = 14$[g]	None (BARS)	Average of all ratings
12	PBDI or SI	None	9	None (BARS)	Average of all ratings

model providing concrete suggestions for bridging the construct validity gap; and (c) we provide a detailed example demonstrating how to implement the proposed model. 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 (Macan, 2009; Posthuma, Morgeson, & Campion, 2002). Thus, we review a significant research area and provide guidelines on construct validation issues that are of broad interest to researchers.

We acknowledge at the onset that the theoretical rigor proposed in our study may raise some questions regarding the practical application of our model. We believe such concerns are connected with the broader debate regarding the divide between practitioners and management scholars. Although, there are no short-term solutions that can overcome this divide, we believe the onus for bridging this divide mainly rests on management researchers. Without researchers' focus on construct validity issues in their own research programs, there would be little guidance for practitioners to understand these issues. This notion is supported by comments in section 14D of the Equal Employment Opportunity (EEO) Uniform Guidelines (1978), where researchers are urged to clarify construct validation strategies for practitioners. Overall, by focusing on construct validity aspects outlined in our model, researchers should eventually make construct validation feasible and inexpensive for recruitment mangers, thereby bridging the research-practice divide. We start our discussion with a brief overview of the current state of research. Next, we outline our model in detail, which is followed by a section illustrating the application of the proposed model. We conclude our discussion with theoretical and practical implications.

2. The current state of conceptual development and design

The construct validity of an employment interview has to do with understanding why an interview predicts performance. Based on Cronbach's (1990) approach to construct validation, one needs to first show which constructs (e.g., motivation or personality) account for variation in the interview ratings. Second, it should be explained how constructs measured by the employment interview are hypothesized to be related to job performance. Finally, these hypotheses should be tested empirically. In this sense, construct validity analysis is an ambitious theory building endeavor. There is no single statistical parameter that provides direct evidence for construct validity. Establishing construct validity rests on showing reliability, as well as content, convergent, discriminant and criterion validities (Schwab, 1999). Keeping this basic framework in view, in the following section we address how validity challenges arise in the *design* of structured interviews. While threats to validity may also occur during interview administration, structuring interviews has played a key role in reducing these biases. Thus, we focus mainly on interview design as a major source of threats to construct validity.

The conceptual and design issues highlighted in our study apply to structured interview studies in general. However, in addition to a general review, we also closely examined the conceptual development and design aspects of those studies that stated construct validation as their main objective. Therefore, we searched various databases such as Dissertation Abstracts, EBSCO, PsycLit, Psyc-Info, and Web of Science to identify construct validity studies. Our aim was to review and tabulate the conceptual development and design aspects of construct validity studies. A thorough database search identified twelve such studies (see Table 1.).

The conceptual development focuses on what the interview should measure, whereas subsequent design activities focus on how it should be measured. Typically, in employment interviews, conceptual development involves conducting some type of job analysis, such as critical incident analysis (Flanagan, 1954). Job analysis identifies the key work behaviors or job dimensions for the focal job. These, in turn, become the basis for designing structured behavioral or situational interview questions and assessment keys. This was true for all the construct validity studies listed in Table 1.

However, job analysis is not aimed at incorporating the nomological network of job performance. Theory building or construct validation requires a clear definition of the criterion domain, a proper explication of the nomological network shared by the job dimensions and their predictors, and an adequate sampling of the content domain. Nevertheless, interviews are generally designed around job analysis without any *a priori* hypothesis regarding how various job dimensions are related to the nomological network of job performance. While a few authors draw a connection from job dimensions to theoretical constructs (e.g., Allen, Facteau, & Facteau, 2004; Van Iddekinge, Raymark, & Roth, 2005), the broader theoretical links are largely ignored. In

addition, our review revealed that while interview questions are designed on the basis of job analysis (Table 1), often it is not explicated how each question is related to the nomological network. This practice results in not knowing how interview questions (and the respective job dimensions) might be related to other predictors of job performance during interview design. A lack of *a priori* knowledge of underlying constructs is especially problematic for a predictor method that can be designed to measure a wide array of constructs. Next, we highlight two major problems that are associated with having a vague nomological network.

First, a lack of *a priori* hypotheses regarding how different job dimensions are related to job performance renders the post-interview validity analysis weak. In principle, the validity efforts rely on showing that the correlations of the focal construct with the other constructs conform to the theoretical expectations. For instance, if an interview score truly captures the proactivity of a job candidate, it should be correlated with other measures of proactivity, as well as with known correlates of proactivity (e.g., conscientiousness; Thomas, Whitman, & Viswesvaran, 2010). Similarly, such an interview score should be uncorrelated with variables that are not expected to influence proactivity (e.g., agreeableness; Thomas et al., 2010). Thus, an *a priori* knowledge of the nomological network of the interview construct points to the type of evidence one needs to collect in order to validate the construct. Nevertheless, understanding the nomological network of interview constructs is largely ignored in extant research. To add further complication, overall interview scores are often used to make construct validity inferences without understanding what these scores mean.

Second, a weak explication of the predictor-criterion link confounds the *method-related* variance with the *intended-construct* variance. Campbell and Fiske (1959) suggested that any measurement attempt is a unique *trait-method* unit which elicits responses that are a combination of the *intended trait* content and the *inherent method* characteristics. For example, if organizing behavior is the *intended construct,* then the final interview score could be a combination of organizing behavior (a *trait* content) and communication skills (an inherent *method* characteristic of face-to-face interviews). Without a clear understanding of what constructs the interview is trying to measure and how each question and assessment key is related to those constructs, it becomes difficult to separate *intended-construct* variance from the *method-related* variance. For instance, some have claimed that interviews are oral tests of cognitive ability (e.g., Hunter & Hirsh, 1987)—a *method-related* variance. However, research does not provide consistent evidence for this claim (e.g., Huffcutt, Roth, & McDaniel, 1996). Before concluding one way or the other, one must examine the nomological network of the performance behaviors assessed by interview questions. A method-related effect of cognitive ability should only be claimed if the performance behaviors measured (i.e., intended constructs) are *not* expected to be related with cognitive ability. For example, the scores on an interview question measuring integrity should not be correlated with cognitive ability, because the latter is not part of the nomological network of integrity (Sackett & Wanek, 1996). Thus, a correlation between cognitive ability and the score on an integrity question would support the notion that cognitive ability is an inherent characteristic of interviews. However, such a claim cannot be made if we observed a correlation between cognitive ability and the score on a leadership question, as cognitive ability *is* part of the leadership nomological network (Hogan, Curphy, & Hogan, 1994).

Current research is moving in the direction where issues related to structure, such as antecedents of structure, various facets of structure, and the relative contribution of each facet, are being explored (e.g., Melchers, Lienhardt, Von Aarburg, & Kleinmann, 2011). These efforts, although vital for interview validity, are not sufficient to ensure construct validity. When solving the construct validity puzzle, we need to look beyond structure and job analysis. At the conceptual development and design stage, theoretical grounding is the third element that is currently missing but necessary to ensure construct validity. Next, we propose a framework that should aid in the theoretical grounding of interviews.

3. Agenda for future construct validity research

Our review indicated that the theoretical grounding of interviews is a major challenge during the conceptual development and design of interviews. The construct validation of employment interviews is essentially a theory building and hypothesis testing process (Binning & Barrett, 1989; Landy & Shankster, 1994). In order to have validity, an employment interview has to be job-relevant as well as theoretically sound. While the former is usually well addressed in current research, more work is needed to ensure theoretical rigor in employment interviews. In this section, a comprehensive four-step model is recommended that explicates how to bring theoretical rigor during the design and development stages of employment interviews.

For each step of the model shown in Fig. 2, we discuss the relevant validity inferences illustrated in Fig. 1. Inferences in Fig. 1 are based, in part, on a construct validation process first outlined by Nunnally (1978). He argued that there are four basic inferences that define the construct validity process: (a) a theoretical X–Y link between a predictor construct and a criterion construct, (b) a measurement x–X link or content-related validity inference regarding a measure x tapping the domain of construct X, (c) a similar y–Y link or content-related validity inference for the criterion measure y, and (d) an empirical x–y link or criterion-related validity inference regarding a relationship between measures x and y. It is important to note that only the fourth inference is amenable to direct empirical testing. However, for theory building or construct validation, all four inferences are of equal importance.

Binning and Barrett (1989) applied Nunnally's (1978) framework to selection research and suggested that the validation of a selection tool is a special case of theory building. The validation process and related inferences suggested in Fig. 1 are in line with both Nunnally's (1978) and Binning and Barrett's (1989) view of selection validity. We build on the previous theoretical
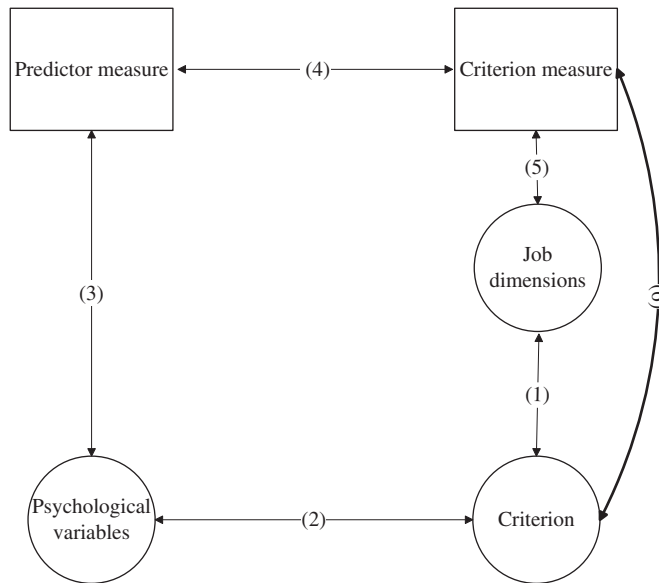
Fig. 1. Construct validity inferences for employment interviews. Note: Key validity inferences 1 through 6 are illustrated in Fig. 1.

contributions and discuss these inferences at each step of the interview model presented in Fig. 2. The following section describes the proposed model in detail. After presenting the model, we provide an example to demonstrate the application of the model.

3.1. Step 1: theoretical definition of the criterion space

Often, a job analysis is conducted to understand the behaviors or traits that are relevant for the focal job. The job performance behaviors are essentially a value judgment of organizational experts regarding the ideal behaviors, and may vary across jobs or with time. We propose that, as a first step for theoretical grounding of interviews, the job dimensions should be framed within a relevant theoretical framework. At present, little effort is made to integrate the job dimensions to any theoretical framework. Each study identifies and selects performance dimensions (e.g., handling irate customers or motivating sub-standard performers) on the basis of job analysis alone. Although these dimensions are relevant to the job, they are disconnected from the broader
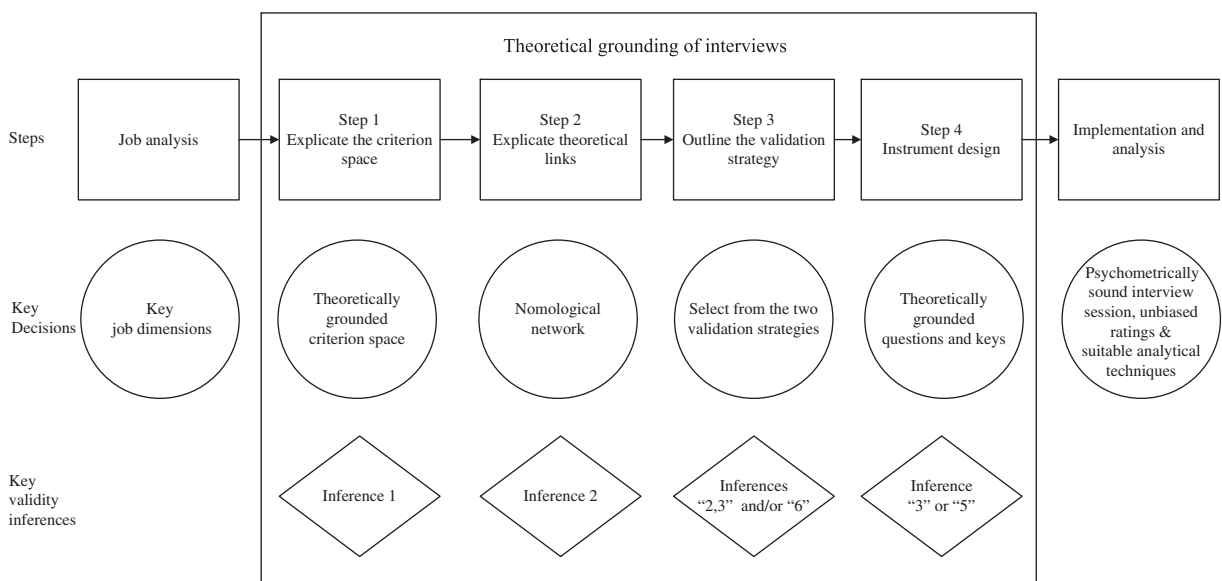

Fig. 2. Theoretical grounding of employment interviews.

theoretical framework of job performance. While corresponding theoretical constructs often exist, job dimensions are labeled casually, resulting in two sets of construct labels: one used in the literature and another used in practice (Binning & Barrett, 1989). In this sense, personnel selection procedures require the meeting of the two worlds of theory and practice. Hence, job dimensions should be translated into the theoretical language system.

The translation of job dimensions into theoretical constructs requires reframing the dimensions using a theory of job performance (inference 1 in Figs. 1 and 2). There are several taxonomies of job performance available that can help map the job dimensions into a broader theoretical framework, including the great eight work competencies (Bartram, 2005), the eight performance components (Campbell, McCloy, Oppler, & Sager, 1993), the task and contextual performance framework (Borman & Motowidlo, 1993), and the hyper-dimensional taxonomy that identifies fifty-three job performance dimensions grouped under nine major categories (Tett, Guterman, Bleier, & Murphy, 2000). These taxonomies of the criterion domain strike a balance between a multitude of fragmented and narrowly defined job dimensions and a global measure of job performance.

From a researcher's point of view, step 1 sets the basis of a strong study design to test construct validity. A stronger theoretical focus allows for a systematic investigation of interview validity. For instance, an *a priori* focus on a theoretical framework helps researchers evaluate the suitability of the research setting for studying a particular question. For example, if researchers are interested in contextual performance, they would first select a setting where contextual performance is likely to occur and then compare job analysis information to the theoretical framework to confirm this expectation. Furthermore, the theoretical understanding of job dimensions enables researchers to draw boundary conditions around their results. This helps linking study results to a particular criterion domain without premature generalizations to the entire criterion space (i.e., all job contexts). For instance, research questions concerning the comparative validity of different interview types (e.g., Huffcutt, Weekley, Wiesner, Groot, & Jones, 2001), the accuracy of personality judgments in interviews (e.g., Barrick, Patton, & Haugland, 2000) or the criterion validity of a type of interview (e.g., Sue-Chan & Latham, 2004)—all should be systematically investigated for several criterions before making generalizable claims. Moreover, using such frameworks should also enable systematic cross-study investigations of predictor-criterion links and a synthesis of research findings. Finally, a good understanding of the construct also aids in the design of measures and data collection and analysis processes, as we discuss later.

From a practitioner's point of view, matching specific job dimensions to a more general theoretical performance framework provides a better understanding of the jobs in the organization in at least three different ways. First, the focus on theoretical constructs may aid in interpreting and integrating job analysis information in a more meaningful way. For example, different critical incidents collected during job analysis may reveal that goal setting, focus on rules, and task coordination are three important job behaviors. However, a theoretical focus would show that all three dimensions are part of the leadership behavior "initiating structure" (Judge, Piccolo, & Ilies, 2004). Thus, a diverse set of dimensions identified in job analysis might turn out to be manifestations of a single key behavior after theoretical reframing.

Second, this insight may prompt further investigation of the broader behavior to see what other dimensions are included in theory, but were not identified by job analysis. One limiting assumption of job analysis is that job incumbents and supervisors are aware of all the performance dimensions of the focal job. Thus, reference to extant literature can supplement or confirm the information gained through job analysis. For example, the literature provides other dimensions of initiating structure that may be relevant for the job, such as defining roles and establishing clear communication channels (Judge et al., 2004). Managers can then decide whether such dimensions should be included, or whether the organization can benefit from the increased simplicity, without jeopardizing its ability to recruit qualified employees.

Finally, matching narrow job dimensions to theory-driven broader constructs may reveal similarities between different jobs and reveal which competencies are most transferable from job to job within the organization. This knowledge can facilitate initiatives in training and development, job enlargement, job rotation, or lateral transfers. Overall, as a first step, researchers should match the job dimensions to one of the available theoretical frameworks of job performance (inference 1). In fact, many researchers have argued for a "criterion-centric" (as opposed to a job-centric) approach for employee selection (Bartram, 2005; Binning & Barrett, 1989), which enables defining *a priori* hypotheses of job performance and helps finding stronger empirical evidence for these hypotheses (Bartram, 2005).

3.2. Step 2: explicate the theoretical links

In the first step, the criterion space is clearly laid out (inference 1); whereas in the second step researchers should explain what antecedent constructs are related to the focal criteria (inference 2). A clear definition of the nomological network aids the construct validity process in several ways. First, it clarifies the nature of the constructs measured in the interview and outlines which constructs are central for a given job. Second, it clarifies how different constructs are related to each other. Third, it may help in identifying job-relevant constructs that were overlooked in job analysis. Finally, it provides a comprehensive map of inter-related constructs which may aid in planning the overall selection process. Specifically, an overall map may help decide which constructs to measure in a job interview and which constructs to measure through alternative selection methods. For example, the performance domain may include personality variables, such as conscientiousness, which are more easily measured through personality tests, rather than selection interviews.

Overall, the nomological network informs both researchers and practitioners on the variables they need to collect in order to show strong evidence of construct validity. In addition, researchers can also learn about constructs that have a bearing on the

performance behaviors studied and may need to be included as control variables in the study design. Similarly, human resource practitioners and line managers may learn about important antecedents of performance behaviors that need to receive increased managerial attention not just for recruitment, but also for ongoing employee development and management initiatives.

Another important advantage of developing a nomological map at the design stage concerns subsequent improvement of the performance evaluation measures used in the organization—which is of consequence to both researchers and managers. A strong theory and interview design enables the critical examination of the criterion measures. For example, suppose the interview program was designed to assess personal initiative and proactive behaviors, and included interview questions on these behaviors. The nomological network should prompt researchers or managers to collect paper–pencil measures of psychological correlates (e.g., extraversion, proactive personality, or personal energy). Given this information and assuming interview questions had construct validity (as shown by post-interview analysis), one may use these scores to examine the validity of employee performance measures later on. If the criterion measures used in the organization fail to correlate with variables in the nomological network, then one should reexamine the validity of the performance measures. Of course, this type of analysis requires a more systematic approach that involves planning and periodic data analysis. Next, we discuss various job performance frameworks that can be used to build the nomological network of job performance.

There are several job performance frameworks that can help in defining the predictor-criterion link (i.e., inference 2). Some of these theories include the person–job and person–organization fit frameworks (Kristof-Brown, Zimmerman, & Johnson, 2005), the task and contextual performance framework (Borman & Motowidlo, 1993), the adaptive performance framework (Pulakos, Arad, Donovan, & Plamondon, 2000), and the knowledge–skill–ability framework (Campbell et al., 1993). Each of these job performance frameworks underscores different aspects of work performance. The choice of framework may depend on the research agenda, on the strategy and culture of an organization, or on the nature of focal job. For example, when one is interested in studying adaptive behaviors, or when the culture or strategy of the organization places a premium on adaptability, Pulakos et al.'s (2000) framework of adaptive performance may be used.

The person–organization fit framework (Kristof-Brown et al., 2005) goes beyond simple job requirements by emphasizing the compatibility of a candidate with the organization. A candidate's fit can be assessed with respect to the broader organizational culture, the immediate work group, or the job at hand. Thus, the nomological network of performance may include knowledge, skills, and abilities, when considering person–job fit, personality and teamwork skills, when evaluating person–group fit, or personal values, when assessing person–organization fit. This framework offers flexibility in viewing the suitability of candidates from various perspectives. However, in order to define a good fit, this framework demands considerable introspection from an organization that should go far beyond conducting job analysis. An organization might need to explore its culture and values across different levels (see Bowen, Ledford, and Nathan (1991) for an application of the fit framework in selection processes). Research has identified many individual characteristics relevant for fit, including personality, needs, skills and values (Kristof-Brown et al., 2005). This literature can be used to generate hypotheses about the relationships between candidate characteristics and organizational fit.

Borman and Motowidlo's (1993) framework posits that contextual performance is an important component of job performance. Contextual performance refers to activities that increase organizational effectiveness by supporting the broader social and psychological environment within which the technical core exists (Borman & Motowidlo, 1993). Grounding interviews in this literature allows researchers to predict which constructs are likely to influence contextual performance. For example, if an employment interview is designed to distinguish between different levels of citizenship behaviors (a type of contextual performance), the dispositional factors and the values of employees which have been shown to predict this behavior (e.g., Podsakoff, MacKenzie, Paine, & Bachrach, 2000) need to be measured in the selection process, in order to estimate the construct validity of the interview. Meta-analyses and literature reviews provide a convenient way to understand the nomological network of a focal construct. Elements of the nomological network could be measured in the interview, alongside with the relevant performance behaviors they predict, or with other selection methods, such as online personality tests. Next, the correlations between these variables and the focal construct need to be checked against expected correlations based on previous research and meta-analyses.

Campbell et al.'s (1993) job performance framework divides the criterion space into eight components: job specific task proficiency, non-job-specific task proficiency, communication proficiency; demonstrating effort, maintaining self-discipline; supportive teamwork behaviors; leadership behaviors, and general management and administration behaviors. Campbell et al. (1993) argued that individual differences on any of the performance components (e.g., leadership behavior) depend on various proximal and distal predictors. They posited that knowledge, skills, and motivation are the proximal predictors (or determinants) of job performance. Each determinant (e.g., interpersonal skills) in turn has its own set of antecedents or distal predictors (e.g., ability, personality, aptitudes, education, training, values, beliefs, and needs).

There is a vast body of research that informs how antecedents account for variance in a particular determinant of job performance. For example, consider the performance dimension of leadership behaviors in Campbell et al.'s (1993) framework. Job analysis may help in identifying specific behaviors (e.g., monitoring and control, work planning and role distribution). Both researchers and practitioners can benefit from tapping into vast body of research informing us about leadership behaviors and its proximal and distal antecedents (e.g., Hiller, DeChurch, Murase, & Doty, 2011; Judge, Bono, Ilies, & Gerhardt, 2002; Yukl, 2012). Researchers and practitioners might find this framework to be particularly useful because of its emphasis on the distinction between distal predictors, determinants and job performance behaviors. This distinction is important because it has bearing on the validation strategy that needs to be employed, as we discuss later in step 3. As such, this general structuring of the criterion domain is useful even when employing other theories of job performance.

Pulakos et al. (2000) built on Campbell et al.'s (1993) framework to include adaptive performance as a distinct and important type of performance behavior in today's dynamic organizations. These researchers outlined eight dimensions of adaptability: handling emergencies or crisis situations, handling work stress, solving problems creatively, dealing with unpredictable work situations, learning work tasks, technologies, and procedures, demonstrating interpersonal adaptability, demonstrating cultural adaptability, and demonstrating physically oriented adaptability. This framework is relevant for jobs where adjusting to a changing job context is critical. The framing of job dimensions with the adaptive performance literature may identify links that might otherwise remain obscure. For instance, research has shown that the nomological network of adaptive performance includes cognitive ability and openness to experience (Le Pine, Colquitt, & Erez, 2000). Such knowledge serves to build *a priori* hypotheses and helps in deciding what information to collect and plan for subsequent post-interview analysis. In sum, all of the frameworks discussed above emphasize different aspects of job performance and clarify the predictor-criterion link of an interview at the onset of the interview process.

3.3. Step 3: explicate the validation strategy

The third step in the theoretical grounding of interviews requires understanding which components of the nomological network are directly measured by the interview questions. An employment interview is a predictor method and can potentially measure a range of constructs in the nomological network. The versatility and flexibility of employment interviews has been discussed (e.g., Harris, 1999; Schuler, 1989). What is needed is to explicate how this versatility influences the construct validity process. The flexibility of interviews is a double-edged sword which poses considerable challenges in integrating the research on some theoretical basis. In interview research, it is often not clear which part of the antecedent-determinant-performance link is measured. Studies generally state that critical components of jobs were identified through a job analysis. Some studies do not even mention these dimensions (e.g., Goffin, Rothstein, & Johnston, 1996), and others merely provide a list of dimensions without mentioning whether they are performance components, determinants, or antecedents of performance.

The issue of not explicating which particular aspect of the antecedent-determinant-performance link is measured by the interview obfuscates the validation strategy that should be used. Binning and Barrett (1989) argued that the validation of a personnel selection decision should employ one of the three strategies: construct-related validation (inferences 2 and 3), content-related validation (inference 6) or criterion-related validation (inference 4). The authors termed the first two validation strategies as inherently superior to the third strategy. They discouraged over reliance on the third strategy, terming this approach susceptible to dust-bowl empiricism. As shown in Fig. 1, criterion-related validity depends on inference 4, which is based on correlational evidence without a strong theoretical justification. Thus, in our proposed model, the first two theoretically-grounded validation strategies will be emphasized.

When the nomological network is explicated, the relevant validation strategy becomes obvious. If an interview is designed to tap into past behaviors (e.g., Patterned Behavior Description Interview or PBDI) then the relevant validation strategy is content-related validation. In this case, the main task is to ensure that the interview adequately samples the criterion domain (inference 6). On the other hand, when the employment interview is designed to explore antecedents of performance (e.g., personality), the relevant validation strategy is construct-related validity. In this case, the primary focus is to identify psychological variables that are linked to the criterion (i.e., inference 2 established in step 2) and to adequately sample the psychological domain (inference 3). It is possible to have both validation strategies if an interview is designed to measure both psychological variables and job behaviors. Similarly, for construct validity studies both validation strategies are important, as these studies require collecting information on psychological variables for subsequent statistical analysis. Thus, an *a priori* explication of a validation strategy emphasizes which aspects of interview development to focus upon.

Overall, we suggest that at step 3 researchers review the job dimensions identified in the earlier steps, explicitly label them as either criterion measures or antecedents of job performance, and clarify the validation strategies being used in the interview. *A priori* explication of the nomological network and the clarification of the validation strategy may eventually help researchers in understanding 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. In other words, research may ultimately provide evidence whether one validation strategy is superior to another for making employment interview decisions.

3.4. Step 4: design of questions and response assessment keys

Whether an interview involves a construct-related validation approach or a content-related validation approach, designing a measurement instrument is an important step in the process. The validation of the measurement instrument requires developing questions that elicit relevant responses from the candidates, and designing a standardized assessment scale that is theoretically sound. Both components aim at adequately sampling the relevant construct domains. For a content-related validation strategy, the emphasis is on inference 6 and its subcomponents (inferences 1 and 5). Inference 1 is satisfied once job dimensions are grounded in the theoretical criterion space (i.e., step 1), whereas inference 5 requires establishing a link between job dimensions and interview measures. Similarly, for construct-related validity, inference 2 is satisfied in step 2 and inference 3 requires validation of interview measures. Thus, in either case, at this stage content validity is critical.

In the overall validation process, the content validity aspect is somewhat qualitative in nature, as it relies primarily on judgmental evidence. Some researchers have suggested ways to quantify these judgments. For example, one may compute a content validity ratio (Lawshe, 1975) for each item. This ratio is a function of the agreement among experts on whether an item is part of the content domain. This concept can be applied to quantify the content validation efforts in the design of interview questions and keys. Although the content validity ratio attempts to quantify and systematize the judgment process, it nonetheless is subjective.

To improve content validation efforts we need to go beyond eliciting feedback from managers and job incumbents. We need to refer to the relevant theoretical literature when designing interview questions and response keys. Theoretical rigor can be exercised whether the interview instrument is tapping into the performance behavior domain (content-related validation), such as teamwork behaviors, or the predictor domain (construct-related validation), such as proactive personality. For example, research has identified five major dimensions of teamwork behaviors (Stevens & Campion, 1999). These dimensions and their associated measurement scales can be used in the design of questions and response keys. 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). As such, basing interview questions in existing scales provides a more solid theoretical ground. In addition, cross-study comparisons can more easily be made when the constructs in each study are properly defined and the measures for those constructs can be traced back to existing psychometric scales.

The use of existing research in the design of questions and assessment keys greatly aids in testing and interpreting validity. Interview questions are often averaged, and overall interview scores are used to demonstrate construct validity, even though the different interview questions may have measured different constructs. Thus, overall scores end up mixing together different constructs, which makes the interpretation of findings problematic. For example, one study found that PBDI scores were related to "manager-relevant cognitive ability measures" (Krajewski, Goffin, McCarthy, Rothstein, & Johnston, 2006, p. 411); yet, other studies found that PBDI questions are not significantly correlated with cognitive ability (e.g., Huffcutt, Weekley, et al., 2001; Pulakos & Schmitt, 1995). Given the differences in the samples and job dimensions (i.e., different intended constructs) in these studies, it is difficult to conclude which results have greater validity. Using existing scales as a starting point in the design of interview questions makes it easier to decide what questions can be averaged together and how they may be expected to correlate with other constructs. Once again, future studies that clarify these issues at the design stage by focusing on what constructs each interview question measure may help in making more informed comparison across studies.

In addition to the above-mentioned design-related suggestions, researchers may also need to adopt some norms for publishing interview research to promote theoretical rigor in interviews. In order to build upon the previous research and to add theoretical rigor, it is important to share a complete list of interview questions and rating keys in an appendix. Unfortunately, the sensitive nature of many jobs, the limited journal space, and the risk of making the questions accessible to job candidates preclude the open sharing of this information. However, new technologies and password protected online knowledge sharing platforms may facilitate the sharing of vital information regarding interview response assessment keys for research purposes.

3.5. Implementation of the model: a practical example

To illustrate how the steps proposed above should facilitate in the construct validation of interviews, we used this model for designing an interview for a selective training program in a large public US university. The successful candidates would go through a two-year long intensive leadership training program. A job analysis identified several critical performance dimensions, including motivating others, being assertive, taking initiative, drive, organizing, integrity, and communication skills. Although it was a training program, a quick glance at Table 1 shows that the identified job dimensions are relevant for several managerial jobs as well. In step 1, we grouped these dimensions into broader theoretical constructs using Campbell et al.'s (1993) job performance framework as follows: the dimensions of motivating others and being assertive was grouped under the broader dimension of leadership behaviors; taking initiative is a type of proactive behavior and was categorized as a specific dimension of demonstrating effort; and organizing was grouped under the general management behaviors category.

The above-mentioned constructs represent job performance behaviors and are thus part of the criterion domain. However, integrity is a distal predictor (i.e., a trait), and drive and communication skills are proximal predictors of job performance. Thus, step 1 clarified the theoretical constructs relevant for the training position. In this example, the interview taps into training performance behaviors related to demonstrating effort, leadership and organizing behaviors. Moreover, step 1 clarified the nature of these constructs by showing which job dimensions are part of the criterion domain and which are predictors of the criterion. Explicating the nature of constructs helped in deciding the appropriate validation strategy in step 3. Overall, the first step focused on validating inference 1.

In step 2 the nomological network of these constructs was explicated by referring to available research. For example, the interview construct of leadership behaviors is related to the traits of extraversion (Hogan et al., 1994) and proactive personality (e.g., Bateman & Crant, 1993). The construct of proactive behaviors is also related to extraversion and proactive personality (e.g., Bateman & Crant, 1993). The focal construct of organizing behavior is related to conscientiousness (e.g., Barrick, Mount, & Strauss, 1993), and the focal construct of integrity is negatively related to deviant work behaviors (e.g., Collins & Schmidt, 1993) and positively related to leadership (e.g., Hogan et al., 1994). Similarly, drive is part of the nomological network of extraversion, conscientiousness (e.g., Barrick et al., 1993), proactive behaviors (e.g., Bateman & Crant, 1993), and leadership (e.g., Hogan et al., 1994). Finally, communication skills

are related to leadership behaviors (Hogan et al., 1994). Overall, these relationships clarify how the interview taps into the theoretical domain of job performance (Fig. 3).

The analysis of the nomological network is useful in several ways. First, we can more clearly distinguish between job elements that are part of the criterion space—like leadership behaviors and proactive behaviors—and job elements that are distal and proximal predictors. Second, the review of literature revealed a performance dimension—deviant behaviors—that was not apparent from the job analysis. Third, several correlates of the criterion domain variables were identified through this analysis. Measuring these additional variables can facilitate the validation process. Fourth, the analysis indicates that some of the job dimensions (e.g., drive) could be measured indirectly, through their correlates in the nomological network for which paper–pencil tests are available (e.g., conscientiousness). This information can help in the planning and design of the selection process by highlighting whether the constructs can be measured with other selection tests. Overall, the second step validated inference 2.

Step 3 clarified the appropriate validation strategies. The constructs related to job performance behaviors (e.g., leadership behaviors and proactive behaviors) represent the criterion domain and thus require a content-related validation strategy (inferences 1 and 5). On the other hand, integrity, drive, and communication skills are predictor constructs that require adopting a construct-related validation strategy (inferences 2 and 3). For both validation strategies, inferences 1 and 2 are already satisfied because of the earlier steps. At this stage we clarify which validity inferences (inference 3 or inference 5) are relevant for which constructs. This step can also help in deciding which additional measures to collect (e.g., paper–pencil integrity tests or verbal American College Testing (ACT) scores for communication) for construct validity analysis in the post-interview stages (e.g., multitrait–multimethod matrix). Inference 3 (i.e., the validity of the predictor measure) and inference 5 (i.e., the validity of the criterion measure) are discussed next. For brevity, we only focus on inference 5 concerning proactive behaviors.

The fourth and final step involves developing measures (i.e., questions and keys) that adequately tap into the construct domain. In our training example, research can help in assessing various interview constructs. For instance, the interview construct of proactive behaviors can be defined as a set of activities that includes scanning the environment for opportunities to improve, showing initiative, taking action, and persevering until one succeeds in bringing about change (Bateman & Crant, 1993). This knowledge helps understanding the key sub-dimensions and designing the appropriate questions, thereby augmenting job analysis.

After reviewing the literature and taking feedback from five researchers and the program director, several job-relevant as well as theoretically relevant questions were identified. See Appendices 1 and 2 for an example of a proactive behavior question. For the proactive behavior question, several question stems were designed keeping in view the sub-dimensions of proactive behaviors. For example, the stem "what was your role in the project" explores whether the person was a leading force of change or just a follower. Similarly, the last stem "what actions did you take to overcome any challenges along the way?" focuses on persevering.
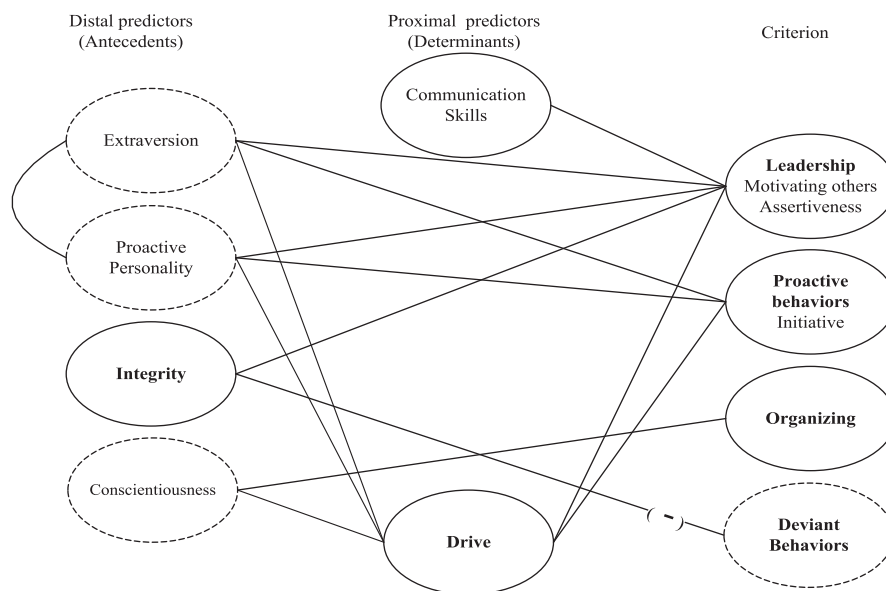

Fig. 3. Explicating the theoretical links for the position of training associates. Notes: The solid-line variables are identified through job analysis; the dotted-line variables are identified through literature review. The negative relationships are represented by (−), all other relationships are positive.

A job analysis may only identify a narrower set of proactive behaviors that job incumbents (e.g., existing training associates) are familiar with. However, a focus on theory may help in assessing proactive behaviors more comprehensively by designing questions that capture other job-relevant aspects of proactive behaviors that were overlooked in job analysis. These theoretically relevant aspects may include proactive career development or proactive idea implementation. Thus, the focus would switch from measuring specific types of proactivity to measuring proactivity in general.

The move to a more general, rather than job-specific, measure of proactive behavior may raise the question of whether the predictive validity of such an interview would be compromised. We argue that this risk is minimal, particularly when the performance dimension belongs to the contextual or adaptive performance categories. Unlike task performance behaviors, which can be more job-specific, contextual and adaptive performance behaviors are relevant across a wide spectrum of jobs. Thus, the interview should aim to better distinguish between proactive and non-proactive candidates, rather than between candidates who have or have not displayed very specific instances of proactivity. The latter approach may work well when the applicant pool includes only candidates with work experience in work contexts highly similar to the focal job, but may prove unnecessarily limiting in the case of a more diverse applicant pool, or when previous job experience is limited (i.e., entry-level job applicants). Even in the case of task behaviors, an effort should be made to strike a balance between the job-specific performance behaviors derived through job analysis, and theory driven criterion-centric constructs that have been shown to be relevant to the performance domain.

Theory can also help with designing assessment keys. The available established scales for the focal constructs can aid in designing theoretically relevant response assessment keys. Appendices 1 and 2 provide examples of sample assessment keys that we defined keeping in view the theoretical definition of proactive behaviors and the established scales (e.g., Bateman & Crant, 1993). In appendices, we provide a sample list of four items from Bateman and Crant's (1993) scale of proactive behaviors that aided in designing the response keys. For example, the scale item "I enjoy facing and overcoming obstacles to my ideas" helped in explaining behavioral anchors to measure persistence. The key shown in Appendix 1 is a behaviorally anchored rating scale, which is commonly used in interview research (see Table 1).

One major shortcoming of behaviorally anchored rating scales is that each rating level is illustrated by providing a benchmark answer for that level. In other words, this type of a scale predicts how an ideal or a poor candidate should respond. A rater has to exercise considerable judgment to rate actual candidate's responses that may not follow the patterns shown in benchmark answers. Instead of defining what an ideal or a poor response should be, an assessment key could focus on defining the components of an ideal response. One solution to this problem is to focus on critical behaviors individually rather than defining benchmark answers for each level. Therefore, we propose a new type of interview assessment key termed behavioral item key (BIK). The proposed key is more precise in defining the items that are central in assessing the focal construct. In this sense, BIK is comparable to behavioral observation scales or the measurement scales used in survey research.

Appendix 2 provides an example of a BIK developed for proactive behaviors. In this scale four items are defined to represent the theoretical sub-dimensions of proactive behaviors. Each candidate's response can be assessed based on these four items. Once sub-dimensions or items are defined, the behavioral anchors can be specified for the finer dimensions with the help of existing survey scales. The average score on the four items can be calculated to assign a final score on the question. Overall, BIK provides clarity in evaluating the candidate's response by focusing on key sub-dimensions rather than a benchmark answer. BIK also provides additional information to understand which sub-dimensions differentiate the candidates from each other. Finally, the assigned scores on these sub-dimensions can be used in further analysis. Thus, content validation of questions and response keys for proactive behaviors establishes inference 5.

The objective of this example was to illustrate how theory can guide and reframe the information gained through job analysis. The example does not suggest replacing job analysis with theoretical grounding, but rather advocates complementing job analysis with theory. Overall, theoretical grounding enables a more comprehensive view of what an interview is measuring. In addition, theoretical grounding of interviews enables researchers to enhance the measurement properties of interview questions by making response assessment keys methodologically sound as well as theory driven.

4. Discussion

Previous research has highlighted the need for theoretical rigor in personnel selection research. We set out to address this issue and provided guidelines on how to increase the construct validity of employment interviews. Our review of construct validity studies revealed that theoretical grounding is often times lacking in interview research. Thus, we laid out a four-step plan to introduce theoretical rigor in the design and development of interviews. We believe these guidelines are timely and of broad interest to researchers. In this section, we discuss the theoretical and practical implications of our model.

4.1. Theoretical implication

Previous validation efforts have generally relied on meta-analyses (e.g., Huffcutt, Roth, Conway, & Stone, 2001) to clarify what constructs are measured by employment interviews. This approach runs the risk of confounding trait and method variances. Each method employed for testing a specific construct is a trait-method unit (Campbell & Fiske, 1959). Thus, different types of

interview questions or interview methods in general will elicit responses that capture the intended constructs (e.g., proactive behaviors) as well as method-related constructs (e.g., impression management, in the case of certain types of interview questions). An interview score contaminated with method variance is less valid.

In employment interview research, when researchers focus on construct validity, it is often not explicitly stated whether the intention is to focus on method-related variance (e.g., what is measured by situational interview (SI) questions and PBDI questions in general, regardless of the intended construct) or intended construct variance (e.g., whether leadership behaviors can be assessed through interviews). Meta-analyses cannot differentiate between intended construct-based variance and inherent method-based variance in the validity estimates. Such differentiation requires conceptual clarity at the design stage and an *a priori* knowledge of what is being measured. We argue that theoretical rigor at the design stage is a prerequisite for separating the trait and method variances. Thus, the model proposed in this study primarily focuses on clarifying what the interview aims to measure and how to measure it in a theoretically rigorous manner.

Theoretical grounding can be used to clarify method variance associated with different types of interview questions. Many studies appear to assume that the responses elicited by similarly worded situational or behavioral questions will tap into similar constructs and should be correlated (e.g., Huffcutt, Weekley, et al., 2001). Research has shown that seemingly trivial changes in context or content can produce significant changes in behavior (Bycio, Alvares, & Hahn, 1987). Trait activation theory (Tett & Guterman, 2000) can be used to examine whether similarly worded SI and PBDI questions do actually tap into similar constructs. Although comparable, it is possible that different types of questions may activate different traits. A situational question asking "what will you do if an angry customer calls?" might activate knowledge, creativity and optimism-related trait systems, whereas a patterned behavior question "what did you do when an angry customer called?" might activate trait systems related to memory or fear of negative attribution about past actions. Thus, although the two types of questions differ only slightly and are comparable on face value, they may indeed generate different reactions and activate different personality systems. The utility of trait activation theory in addressing construct validity issues has already been highlighted in assessment center research (e.g., Haaland & Christiansen, 2002).

In addition to theoretical grounding, researchers should also pay heed to interview administration when comparing the validities of SI and PBDI questions. An interview that contains different types of questions in a single session is called a multimodal interview (Schuler, 1989). The response assessments resulting from a multimodal interview are influenced by same sources of biases resulting from having a common interviewer, interview time and ambiance. This may artificially inflate correlations between different question types. Thus, for comparative validity studies, SI and PBDI questions should ideally be measured in separate sessions to avoid method-related correlations in the constructs. We introduce the term multiphase interviews for cases which include more than one interview session per candidate, each focusing on a single type of question. Overall, we suggest that multiphase interviews are particularly suited for validity comparisons of different types of interview questions.

While we have illustrated how theoretical rigor during the conceptual development and design of interviews can enhance construct validity, we have not addressed all construct validity issues. The subsequent stages of employment interviews pose considerable challenges for construct validity. The actual administration and assessment of an employment interview is an evaluation process embedded in a social context involving six distinct steps: observation, impression formation, information storage, information integration, and formal rating (DeNisi, Cafferty, & Meglino, 1984). Thus, during the actual administration of interviews, contextual, psychological, and cognitive processes influence construct validity.

One such factor that may influence the validity of the interview questions is the transparency of interview dimensions. In assessment center research it has been shown that making the dimensions transparent to the candidate improves construct validity (Lievens, 1998). In employment interview research, a recent study reported that the transparency of dimensions improves the performance of candidates, increases the construct validity, and does not influence the criterion-validity of interviews (Klehe, Konig, Richter, Kleinmann, & Melchers, 2008). Based on these results, Klehe et al. (2008) stated that transparency of interview dimensions should be considered a facet of interview structure. However, further research is needed to investigate the effects of transparency on different constructs, types of questions and job levels.

4.2. Practical implications

We have outlined a method for grounding interviews in theory. However, from a practitioner's perspective the feasibility of our suggestions is a valid concern. In this section, we address some of these concerns. We argued that interview design should be grounded in a theory of job performance, and highlighted issues associated with sole reliance on job analysis for the design of interviews. However, job analysis is a critical part of designing interviews. Our aim is not to inhibit the use of job analysis. Instead, we argue that theory should be used to inform, correct and supplement the information gained through job analysis. The application example provided earlier was intended to clarify some of these issues.

One major issue in the application of the proposed model is the job-centric nature of real interviews. Interview dimensions are not theory driven and instead focus on the unique requirements of a given job. Thus, our theory driven approach may appear irrelevant for real interviews. However, as argued earlier, inference 1 (Fig. 1) emphasizes the possible connection between theoretical constructs and job dimensions. The apparent disconnect between job dimensions and theoretical

constructs is partly due to two different language systems being used by researchers and practitioners. This disconnection is further exacerbated by casual labeling of job dimensions, which results in construct proliferation. In assessment center research, a meta-analysis (Arthur, Day, McNelly, & Edens, 2003) showed how 168 job dimensions were encompassed within an overarching set of six constructs. Thus, theories of job performance summarize seemingly unique job dimensions within a broader unifying framework. For measurement purposes, intermediate levels of classifications can be defined by focusing on specific sub-dimensions, such as proactive behavior, within the broader category of demonstrating effort (e.g., Campbell et al., 1993). Overall, we argue that researchers and practitioners should also focus on common performance elements across comparable jobs in different organizations. While there is a degree of uniqueness to all jobs, there is also commonness. It would serve both researchers and practitioners to account for that by grounding interview research in theories of job performance.

One may question how practical our suggestions are, given that they require theoretical knowledge. Managers might not see the value of investing resources in understanding theories of job performance. Those who are willing to make such efforts might not find this knowledge easily accessible. This issue is connected with the broader debate regarding the divide between practitioners and management scholars. This divide is a major practical limitation of our model with no short-term solution. First, we as researchers need to implement theoretical rigor in our program of interview research. After all, interview researchers have not stopped structuring interviews just because practitioners prefer conducting unstructured interviews. Second, efforts to bridge the practitioner–researcher divide should focus on how to make interview-relevant theoretical knowledge more accessible. For example, publications in practitioner outlets, workshops, and training programs can highlight how job dimensions are connected with theoretical frameworks of job performance. In the long run, this type of practitioner–researcher collaboration may eventually create a knowledge base that readily links common job dimensions with theoretical constructs and associated proximal and distal predictors.

We acknowledge that it might not be practical for managers to implement an exhaustive, ongoing construct validity program for all the hiring positions in an organization. Therefore, we recommend that managers should conduct such an analysis for a few select positions. For instance, in an investment bank, hiring management trainees is an ongoing, high-visibility activity that often involves campus visits, on-site visits, and several rounds of job interviews of fresh business graduates. These organizations expend considerable resources in hiring for such positions. Thus, the strategic importance and frequency of hiring for a position might justify introducing theoretical rigor for such jobs. In such instances, it should be beneficial for the organization to plan in advance for improving construct validity aspects of the interviews. Such focused approach may help in improving or redesigning the selection criteria for the key jobs. Job analysis alone cannot and does not guarantee the validity of interviews. The construct validity analysis provides vital empirical evidence that can direct their future endeavors for improving selection processes.

In addition, such focus by practitioners might encourage collaboration between researchers and practitioners, thus narrowing the research-practice divide. We concede that introducing theoretical rigor can be time consuming and might require additional training. However, practitioners might find many interview researchers who are willing to collaborate on such projects as these objectives align with their research agenda. Furthermore, it is not possible to advance the research on construct validity of employment interviews without an extensive collaboration between researchers and practitioners. We hope that the ideas forwarded in this study would aid in stimulating such collaboration.

The construct validity analysis appears to be a scientific issue that should primarily concern interview researchers. In general, practitioners are mainly concerned with the criterion validity—the ability to predict future performance with interview scores. Nevertheless, we argue that construct validity is an equally important element of the selection process. Construct validity informs us on why the test predicts performance and thus has bearing on its criterion validity as well. Managers who know why a candidate is likely to perform well are better equipped to provide the support that an employee needs to perform well. Thus, by improving the construct validity of interviews, we can concomitantly improve their criterion validity. Despite that, practitioners may still be concerned with the utility of investing efforts in increasing the construct validity of interviews.

The utility of a selection method depends on several factors, such as the validity of the selection method, the number of workers hired per year through the method, the standard deviation of job performance, the selection ratio, and the cost of administering the selection method (Schmidt & Hunter, 1998). Thus, the validity of interview scores is a critical goal for practitioners too, as it impacts the utility of interview decisions. An employment interview is of value to the extent that it is valid. If bringing theoretical rigor in the interview process can aid in increasing the validity of a selection interview, then it should also translate into enhancing the practical value or utility of the selection method.

Such benefits, however, need to be balanced with the additional costs of achieving increased construct validity. For example, including additional elements from the intended construct's nomological network in the selection process may raise the development and administration costs. Additionally, complex assessment methods may place increased burden on both the candidate and the interviewer, with negative implications for the validity of the interview. Fortunately, the wide availability of research on job performance can help with the design of parsimonious selection methods. For example, meta-analyses can point out the strongest and most reliable correlates of the intended construct. Similarly, existing validated measures often come in several formats, ranging from lengthy and comprehensive measures to parsimonious but simpler measures. Thus, a theoretically rigorous selection process need not be cumbersome and difficult to implement.

Finally, we also want to highlight the significance of construct validity from legal stand point. In section 5B of the EEO Uniform Guidelines (1978), construct validity is noted as one of the acceptable validation strategies. However, in section 14D attention is drawn to the fact that construct validation is the most complex validation strategy that needs further development. Practitioners are cautioned about the lack of substantial literature that can guide them in the construct validity process. Overall, at present the agency's approach toward construct validity is more cautious and limited as outlined in sections 14 and 15. Thus, as researchers we are tasked to provide clearer guidelines for construct validity procedures. We believe the model presented here answers the call of EEOC to clarify the construct validity process.

In summary, based on our model we suggest that practitioners pay attention to the following three areas. First, managers should make an effort to understand how a particular job interview fits in the context of overall job performance. Bringing theoretical rigor in interview design encourages practitioners to assume an integrated perspective of different jobs across an organization. Such a theoretical focus is necessary to identify key job behaviors and then design interviews than can accurately measure such behaviors. Second, managers should focus on how interview dimensions are connected with other important constructs (e.g., personality, skills and job performance). This knowledge should help managers to shift their focus from predicting who will perform well (i.e., criterion-validity) to understanding why employees perform well (construct validity). Understanding why employees perform better facilitates superior performance management. Third, managers should identify key positions that would benefit from theoretical rigor.

5. Conclusion

In this paper we suggest that the theoretical grounding of interviews is an essential, but currently missing, element in employment interview research. So far, structure has been the basic method of ensuring construct validity in the interview process. We would suggest that future research needs to move in a different direction. We have attempted to detail this path by proposing a model suggesting how to introduce substantial theoretical rigor in interviews. Both researchers and practitioners need to spend more time on gaining a detailed and intricate understanding of the focal job for which an interview has been developed. Constructs measured in interviews should be theoretically grounded in the nomological network of job performance. Unless we take heed of these issues, we fear that our future progress in developing valid employment interviews in both our research and practice endeavors may be muted at best, and misdirected, at worst.

Appendix 1. Behaviorally anchored rating scale

Question: Tell me about a time when you were involved in a project that resulted in a positive change? What role did you play in this project? How did you facilitate the change? What actions did you take to overcome the challenges along the way?

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

Excellent (5)—if the following elements are present:
  The candidate expresses a keen focus toward making things better; S/he takes personal initiative to identify opportunities for change and plays a major role in identifying and implementing the change; S/he acts on the identified opportunities and resists opposition and persists until the change is brought about; S/he deals with obstacles positively rather than complaining or giving up.
Example: Last year, I proposed to organize a music show for the charity organization I volunteer for. I went out and talked to some of my friends who had arranged similar events in the past. This helped me prepare a budget to present to my committee who were initially lukewarm toward the idea. My numbers convinced them to consider this event as a serious option. I took the responsibility to manage the entire show. My biggest challenge was to gain the necessary support. I talked to my friends, family members, and professional contacts that I knew would be of help. It was tough doing all this with my other commitments. But I believe when you are sincerely making an effort, many people will be willing to help you because they will see your commitment to deliver. The music event generated the largest amount of funds for the charity organizations from an event.
Acceptable (3)—if the following elements are present:
  The candidate shows a desire to make things better; S/he might not have played a significant role in the change, or the candidate comes up with ideas, makes some effort to get them implemented, but gives up when challenges arise.
Example: The charity organization I volunteer with asked to help in arranging a music event. Although it was difficult with my other commitments, it was worth the effort to do something for others.
Poor (1)—if the following elements are present:
  The candidate does not show a desire to make things better; S/he waits for others to highlight what needs to be done and barely does what is expected; Comes up with ideas but never takes the initiative to implement these ideas; There is no clear evidence of a concrete change.
Examples: "I am an active member of various charity and community service organizations."
A clichéd response such as "only change is constant."

Sample items from Bateman and Crant (1993) *scale that aided the design of assessment key.*
I enjoy facing and overcoming obstacles to my ideas; I tend to let others take the initiative to start a new project (reverse coded); I excel at identifying opportunities; Wherever I have been, I have been a powerful force for constructive change.

Appendix 2. Behavioral item key

Question: Tell me about a time when you were involved in a project that resulted in a positive change. What role did you play in this project? How did you facilitate the change? What actions did you take to overcome the challenges along the way?

Underlying 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.)

The total score on the question should be calculated by averaging the score on the four items.

| The response shows a desire to bring improvement. | 1 | 2 | 3 | 4 | 5 |

Explanation:

The candidate's response shows a focus toward making things better, doing something new, and bringing about a positive change. For example, a candidate should express interest in helping others or improving the business. The candidate should express excitement about turning ideas into reality. The response should clearly indicate a desire to change things; just doing an assigned role is not enough. A clichéd response such as "only change is constant" might be rated low.

Examples:

- I was involved in a philanthropic project for kids where I was devoted to bring changes that would make the event bigger and better for kids. *(This response can be rated at 5. The respondent explicitly communicates desire to bring change.)*
- I was involved in a philanthropic project for kids where I helped the chairperson implement some changes. *(This response can be rated at 3. The respondent is somewhat vague about his/her desire to bring improvements.)*
- I was involved in a philanthropy event organized by my sorority. I helped facilitate a positive change by being very exact with our numbers in our budget. *(This response can be rated at 1. The respondent simply describes one's assigned role; participation in sorority/fraternity events is common place on university campuses.)*
- *(A sample item from proactive behaviors scale* (Bateman & Crant, 1993) *that helped defining this dimension: "I feel driven to make a difference in my community, and maybe the world.")*

| Evidence of initiative | 1 | 2 | 3 | 4 | 5 |

Explanation:

The candidate identifies opportunities in a proactive manner. The approach should be proactive and not reactive, passive or adaptive. The candidate takes initiative to solve an issue or bring about a positive change. The candidate takes a lead role. Firefighting when a problem emerges does not provide evidence of initiative.

Examples:

- For example, instead of offering sandwiches, I thought of an idea to involve local chefs. Thus, I started seeking out many local chefs who would be willing to donate their time and food to the charity event. *(This response can be rated at 5. The respondent identifies a novel opportunity and takes steps to bring it to fruition.)*
- We brainstormed different ideas of improvement. *(This response can be rated at 3. The personal role is more ambiguous and less central. Nonetheless, there is an indication of attempts to change the status quo.)*
- I was assigned to investigate how to improve the menu for a charity food event. *(This response can be rated at 1. The response indicates a more passive role.)*
- *(The sample items from the proactive behaviors scale* (Bateman & Crant, 1993) *that helped defining this dimension: "I tend to let others take the initiative to start new project (reverse coded)"; "I excel at identifying opportunities.")*

| The candidate's response shows evidence of persistence. | 1 | 2 | 3 | 4 | 5 |

Explanation:

The candidate takes up a stand and resists opposition and persists till the change is brought about; The candidate does not get discouraged by opposition and persists by taking on challenges with a positive attitude. Championing ideas, motivating others, influencing others, creating a buy-in through rational arguments and overcoming interpersonal issues are some of the relevant behaviors.

Examples:

- There were many challenges that I faced during the planning process. One was convincing chefs that how their involvement will improve the experience for deserving kids. With dedication and the will to not give up, I was able to pull off the project and it was a great success. *(This response can be rated at 5. The challenge is clearly described and evidence of persistence in convincing others is provided.)*
- We worked as a team and encouraged each other to remain persistent. *(This response can be rated at 3. The respondent plays a less central role. Nonetheless, evidence of persistence is provided.)*
- In charity projects there can be challenges, we just need to be open to others' ideas. *(This response can be rated at 1. The response is too general and vague.)*
- *(A sample item from the proactive behaviors scale* (Bateman & Crant, 1993) *that helped defining this dimension: "I enjoy facing and overcoming obstacles to my ideas.")*

| Positive outcomes | 1 | 2 | 3 | 4 | 5 |

Explanation:

The candidate shows evidence of bringing a successful change or of being innovative. The scope (number of people—one person, group, or community) and intensity of change (how deep the change was) may help in deciding the nature of outcomes. The routine outcomes such as "I always meet project deadlines" or general time management skills can be rated low.

Examples:

- Being the chair for this project deeply impacted how I view life. I knew I was responsible for providing the great lunches and activities for these children that would not have had this opportunity otherwise. In the end, I felt I was able to bring some joy to these children's lives and all the hard work paid off right away. *(This response can be rated at 5. The response shows that the change had a significant impact on both the respondent and the beneficiaries.)*
- Our work positively impacted the community and ourselves. *(This response can be rated at 3. The response is somewhat vague and the role of the respondent is not clear.)*
- Charity work has made me a better person. *(This response can be rated at 1. The answer is clichéd and does not provide sufficient evidence of impacting others.)*
- *(A relevant sample item from the proactive behaviors scale* (Bateman & Crant, 1993): *"Wherever I have been, I have been a powerful force for constructive change."*

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. http://dx.doi.org/10.1207/S15327043HUP1701_1.

Arthur, W., Jr., Day, E. A., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology*, *56*(1), 125–153. http://dx.doi.org/10.1111/j.1744-6570.2003.tb00146.x.

Arthur, W., Jr., Day, E. A., & Woehr, D. J. (2008). Mend it, don't end it: An alternate view of assessment center construct-related validity evidence. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *1*(1), 105–111. http://dx.doi.org/10.1111/j.1754-9434.2007.00019.x.

Barrick, M. R., Mount, M. K., & Strauss, J. P. (1993). Conscientiousness and performance of sales representatives: Test of the mediating effects of goal setting. *Journal of Applied Psychology*, *78*(5), 715–722. http://dx.doi.org/10.1037/0021-9010.78.5.715.

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. http://dx.doi.org/10.1111/j.1744-6570.2000.tb02424.x.

Bartram, D. (2005). The great eight competencies: A criterion-centric approach to validation. *Journal of Applied Psychology*, *90*(6), 1185–1203. http://dx.doi.org/10.1037/0021-9010.90.6.1185.

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. http://dx.doi.org/10.1002/job.4030140202.

Binning, J. F., & Barrett, G. V. (1989). Validity of personnel decisions: A conceptual analysis of the inferential and evidential bases. *Journal of Applied Psychology*, *74*(3), 478–494. http://dx.doi.org/10.1037/0021-9010.74.3.478.

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. http://dx.doi.org/10.1177/109442819922005.

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, CA: Jossey-Bass.

Bowen, D. E., Ledford, G. E., Jr., & Nathan, B. R. (1991). Hiring for the organization, not the job. *The Academy of Management Executive*, *5*(4), 35–51. http://dx.doi.org/10.5465/AME.1991.4274747.

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. http://dx.doi.org/10.1108/EUM0000000005329.

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, CA: Sage Publications. http://dx.doi.org/10.4135/9781452205519.

Bycio, P., Alvares, K. M., & Hahn, J. (1987). Situational specificity in assessment center ratings: A confirmatory factor analysis. *Journal of Applied Psychology*, *72*(3), 463–474. http://dx.doi.org/10.1037/0021-9010.72.3.463.

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

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.

Collins, J. M., & Schmidt, F. L. (1993). Personality, integrity, and white collar crime: A construct validity study. *Personnel Psychology*, *46*(2), 295–311. http://dx.doi.org/10.1111/j.1744-6570.1993.tb00875.x.

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. http://dx.doi.org/10.1023/A:1022914803347.

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

DeNisi, 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. http://dx.doi.org/10.1016/0030-5073(84)90029-1.

Equal Opportunity Employment Commission, C. S. C., U.S. Department of Labor, & U.S. Department of Justice (1978). Uniform guidelines on employee selection procedures. *Federal Register*, *43*, 38290–38309.

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, *51*(4), 327–358. http://dx.doi.org/10.1037/h0061470.

Goffin, R. D., Rothstein, M. G., & Johnston, N. G. (1996). Personality testing and the assessment center: Incremental validity for managerial selection. *Journal of Applied Psychology*, *81*(6), 746–756. http://dx.doi.org/10.1037/0021-9010.81.6.746.

Haaland, S., & Christiansen, N. D. (2002). Implications of trait-activation theory for evaluating the construct validity of assessment center ratings. *Personnel Psychology*, *55*(1), 137–163. http://dx.doi.org/10.1111/j.1744-6570.2002.tb00106.x.

Harris, M. M. (1999). What is being measured? In R. W. Eder, & M. M. Harris (Eds.), *The employment interview handbook* (pp. 143–157). Thousand Oaks, CA: Sage Publications. http://dx.doi.org/10.4135/9781452205519.

Hiller, N. J., DeChurch, L. A., Murase, T., & Doty, D. (2011). Searching for outcomes of leadership: A 25-year review. *Journal of Management*, *37*(4), 1137–1177. http://dx.doi.org/10.1177/0149206310393520.

Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods*, *1*(1), 104–121. http://dx.doi.org/10.1177/109442819800100106.

Hogan, R., Curphy, G. J., & Hogan, J. (1994). What we know about leadership: Effectiveness and personality. *American Psychologist*, *49*(6), 493–504. http://dx.doi.org/10.1037/0003-066X.49.6.493.

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. http://dx.doi.org/10.1037/0021-9010.79.2.184.

Huffcutt, A. I., Roth, P. L., Conway, J. M., & Stone, N. J. (2001). Identification and meta-analytic assessment of psychological constructs measured in employment interviews. *Journal of Applied Psychology*, *86*(5), 897–913. http://dx.doi.org/10.1037//0021-9010.86.5.897.

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

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. http://dx.doi.org/10.1111/j.1744-6570.2001.tb00225.x.

Hunter, J. E., & Hirsh, H. R. (1987). Applications of meta-analysis. In C. L. Cooper, & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (pp. 321–357). Chichester, U.K.: Wiley.

Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology*, *87*(4), 765–780. http://dx.doi.org/10.1037//0021-9010.87.4.765.

Judge, T. A., Piccolo, R. F., & Ilies, R. (2004). The forgotten ones? The validity of consideration and initiating structure in leadership research. *Journal of Applied Psychology*, *89*(1), 36–51. http://dx.doi.org/10.1037/0021-9010.89.1.36.

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. http://dx.doi.org/10.1080/08959280801917636.

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 and Organizational Psychology*, *79*(3), 411–432. http://dx.doi.org/10.1348/096317905X68790.

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*(2), 281–342. http://dx.doi.org/10.1111/j.1744-6570.2005.00672.x.

Landy, F. J., & Shankster, L. J. (1994). Personnel selection and placement. *Annual Review of Psychology*, *45*(1), 261–296. http://dx.doi.org/10.1146/annurev.ps.45.020194.001401.

Lawshe, C. H. (1975). A quantitative approach to content validity1. *Personnel Psychology*, *28*(4), 563–575. http://dx.doi.org/10.1111/j.1744-6570.1975.tb01393.x.

Le Pine, J. A., Colquitt, J. A., & Erez, A. (2000). Adaptability to changing task contexts: Effects of general cognitive ability, conscientiousness, and openness to experience. *Personnel Psychology*, *53*(3), 563–593 (Article).

Lievens, F. (1998). Factors which improve the construct validity of assessment centers: A review. *International Journal of Selection and Assessment*, *6*(3), 141–152. http://dx.doi.org/10.1111/1468-2389.00085.

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

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

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. http://dx.doi.org/10.1037/0021-9010.77.5.571.

Nunnally, J. C. (1978). *Psychometric theory.* New York: McGraw-Hill.

O'Leary, R. S. (2004). *A construct validity comparison of alternative structured interview scoring methodologies.* Unpublished doctoral dissertation. Auburn: Auburn University.

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*(3), 513–563. http://dx.doi.org/10.1177/014920630002600307.

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. http://dx.doi.org/10.1111/j.1744-6570.2002.tb00103.x.

Pulakos, E. D., Arad, S., Donovan, M. A., & Plamondon, K. E. (2000). Adaptability in the workplace: Development of a taxonomy of adaptive performance. *Journal of Applied Psychology*, *85*(4), 612–624. http://dx.doi.org/10.1037/0021-9010.85.4.612.

Pulakos, E. D., & Schmitt, N. (1995). Experience-based and situational interview questions: Studies of validity. *Personnel Psychology*, *48*(2), 289–308. http://dx.doi.org/10.1111/j.1744-6570.1995.tb01758.x.

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. http://dx.doi.org/10.1111/j.1468-2389.2005.00323.x.

Sackett, P. R., & Wanek, J. E. (1996). New developments in the use of measures of honesty integrity, conscientiousness, dependability trustworthiness, and reliability for personnel selection. *Personnel Psychology*, *49*(4), 787–829. http://dx.doi.org/10.1111/j.1744-6570.1996.tb02450.x.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*(2), 262–274. http://dx.doi.org/10.1037/0033-2909.124.2.262.

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.

Schwab, D. P. (1999). *Research methods for organizational studies* (2nd ed.)Mahwah, NJ: Erlbaum.

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. http://dx.doi.org/10.1016/S0149-2063(99)80010-5.

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 and Assessment*, *12*(4), 312–320. http://dx.doi.org/10.1111/j.0965-075X.2004.00286.x.

Tett, R. P., & Guterman, H. A. (2000). Situation trait relevance, trait expression, and cross-situational consistency: Testing a principle of trait activation. *Journal of Research in Personality*, *34*(4), 397–423. http://dx.doi.org/10.1006/jrpe.2000.2292.

Tett, R. P., Guterman, H. A., Bleier, A., & Murphy, P. J. (2000). Development and content validation of a "hyperdimensional" taxonomy of managerial competence. *Human Performance*, *13*(3), 205–251. http://dx.doi.org/10.1207/S15327043HUP1303_1.

Thomas, J. P., Whitman, D. S., & Viswesvaran, C. (2010). Employee proactivity in organizations: A comparative meta-analysis of emergent proactive constructs. *Journal of Occupational and Organizational Psychology*, *83*(2), 275–300. http://dx.doi.org/10.1348/096317910X502359.

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. http://dx.doi.org/10.1207/S15327043HUP1701_4.

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. http://dx.doi.org/10.1037/0021-9010.90.3.536.

Yukl, G. (2012). Effective leadership behaviors: What we know and what questions need more attention? *The Academy of Management Perspectives*, 66–85. http://dx.doi.org/10.5465/amp.2012.0088.