

PERSONNEL PSYCHOLOGY

2011, 64, 351–395

**EXERCISES AND DIMENSIONS ARE THE CURRENCY OF ASSESSMENT CENTERS**

BRIAN J. HOFFMAN The University of Georgia

KLAUS G. MELCHERS Universita¨t Zu¨ rich, Switzerland

CARRIE A. BLAIR

Department of Management and Entrepreneurship

Schottland Scholars Program

MARTIN KLEINMANN Universita¨t Zu¨ rich, Switzerland

ROBERT T. LADD University of Tennessee

A novel assessment center (AC) structure that models broad dimension factors, exercise factors, and a general performance factor is proposed and supported in 4 independent samples of AC ratings. Consistent with prior research, the variance attributable to dimension and exercise fac- tors varied widely across ACs. To investigate the construct validity of these empirically supported components of AC ratings, the nomologi- cal network of broad dimensions, exercises, and general performance was examined. Results supported the criterion-related validity of broad dimensions and exercises as predictors of effectiveness and success cri- teria as well as the incremental validity of broad dimensions beyond exercises and general performance. Finally, the relationships between individual differences and AC factors supported the construct validity of broad dimension factors and provide initial insight as to the meaning of exercise specific variance and general AC performance.

Assessment centers (ACs) have traditionally used multiple exercises to measure multiple dimensions pertinent to effective performance in a given job. Despite the emphasis on dimensions in the traditional de- sign and interpretation of ACs, internal structure research has typically demonstrated that postexercise dimension ratings (PEDRs) reflect the exercises in which candidate behavior is observed to a greater extent than the performance dimensions ACs were designed to measure (Bowler

The data from Sample 1 stem from a research project that was supported by a grant from the Swiss National Science Foundation (Schweizerischer Nationalfonds; Grant 100014-

117917) to Klaus Melchers and Martin Kleinmann. The authors would like to thank Chuck

Lance for providing the data used in Sample 3.

Correspondence and requests for reprints should be addressed to Brian J. Hoffman, 228

Psychology Building, The University of Georgia, Athens, GA 30602; [hoffmanb@uga.edu](mailto:hoffmanb@uga.edu)

C 2011 Wiley Periodicals, Inc.

351

& Woehr, 2006; Lance, Lambert, Gewin, Lievens, & Conway, 2004; Melchers, Henggeler, & Kleinmann, 2007; Woehr & Arthur, 2003).

Based on these findings, some have advocated the abandonment of

dimensions in favor of task-based rating strategies (Jackson, Stillman, & Atkins, 2005; Lance, 2008a; Lance, Foster, Gentry, & Thoresen, 2004; Sackett & Dreher, 1982). From this perspective, the exercise structure characterizing PEDRs is not attributable to measurement bias but instead represents valid situational specificity in performance (Lance, Foster et al.,

2004). Accordingly, AC performance should be interpreted as effective or ineffective on a variety of *tasks* (e.g., an in-basket), and dimension-based interpretations should be discarded all together.

Suggestions to abandon dimensions in the design and scoring of ACs are premature for several reasons. First, to the extent that AC internal structure research has misspecified the underlying structure of ACs, the accumulated knowledge of the psychometric properties of ACs has been misguided. Next, research investigating the psychometric properties of ex- ercises is needed before task-based approaches can be adopted (e.g., Jones

& Klimoski, 2008; Lievens, 2008; Melchers & Ko¨ nig, 2008). Finally, as noted by Woehr and Arthur (2003), although multitrait–multimethod (MTMM) research has yielded disappointing results, prior research using diverse methodologies has supported the validity of AC dimensions in the form of postcombination dimension ratings (PCDRs; also known as post consensus dimension ratings and across-exercise dimension ratings).

In light of the continued use of ACs in applied settings, recent sug- gestions that exercise effects are substantively meaningful, and further arguments to abandon dimensions altogether (Lance, 2008a), a closer in- spection of the construct validity of AC ratings is warranted. This study contributes to the literature by combining multiple methods of investi- gating validity in order to facilitate inferences of the overarching validity of AC ratings (Messick, 1995). Toward this end, previous research on the structure of performance ratings was used to inform the development of a new model of AC ratings. Next, the relationships between the fac- tor analytically supported structure of AC ratings and external measures were examined to ascertain (a) the relative criterion-related validity of the empirically supported components of AC ratings, (b) the extent to which dimension factors explained variance in outcomes beyond other facets of measurement, and (c) the nomological network of individual differences surrounding the facets of AC ratings.

*Internal Structure of Assessment Center Ratings*

*Traditional models of AC ratings.* Based on the more general MTMM methodology, AC internal structure research has traditionally assumed that dimensions reflect cross-situationally consistent aspects of performance,

and exercises reflect methods of measurement. Informed by this general MTMM model, past AC research has routinely tested four primary AC structures.

Model 1 (J-dimensions, 0-exercises) parameterizes multiple correlated dimension factors and no exercise factors, assuming that dimension level performance is fully consistent across exercises and that the exercise in which behavior is observed does not influence AC ratings. This model rarely characterizes AC ratings (Lance, Lambert et al., 2004). Model

2 (J-dimensions, K-exercises) proposes that AC ratings reflect two sys- tematic factors: dimensions and exercises, where the number of dimen- sion factors corresponds to the number of manifest dimensions rated and the number of exercise factors corresponds to the number of exercises. This model is historically the most frequently specified (e.g., Anderson, Lievens, van Dam, & Born, 2006; Bowler & Woehr, 2006; Kleinmann, Kuptsch, & Ko¨ ller, 1996), despite systemic convergence and admissibil- ity problems (Lance, Woehr, & Meade, 2007). Model 3 (0-dimensions, K-exercises) specifies no dimension factors and K-exercise factors, sug- gesting that there is no consistency in candidates’ performance across exercises.

Recently, the exercises-only model has been amended to acknowledge

that some performance in an AC is consistent across exercises. Specif- ically, Lance and colleagues (Lance, 2008a; Lance et al., 2000; Lance, Foster et al., 2004) advocated an AC structure where all of the dimen- sions assessed in a single exercise load on a single exercise factor for each exercise, and all of the dimensions rated across exercises load on a general performance factor (Model 4: 1-dimension, K-exercises). Given prior research substantiating the criterion-related validity of certain indi- vidual differences across situations (Schmidt & Hunter, 1998) and support for a general performance factor in the broader work performance litera- ture (Guilford, 1954; Hoffman, Lance, Bynum, & Gentry, 2010; Scullen, Mount, & Goff, 2000; Viswesvaran, Schmidt, & Ones, 2005), the emer- gence of a general performance factor in AC ratings should not be sur- prising. Research has typically supported the superiority of this model in terms of convergence and admissibility, not necessarily model fit (Lance et al., 2000; Lance, Lambert et al., 2004).

However, past internal structure research suffers from at least two limitations: First, Lance et al.’s (2007) simulation study found that the

1-dimension, K-exercises model is frequently supported even when it does not reflect the true (simulated) data, and the J-dimension, K-exercise model is often not supported on the basis of nonconvergence and solution inadmissibility even when it does reflect the true structure of the data. The implications of this trend are severe: prior support for the 1-dimension, K-exercise model may reflect a methodological artifact rather than the explanatory superiority of the 1-dimension, K-exercises model.

Second, in addition to empirical concerns with the 1-dimension, K- exercises model, alternative models may also provide a theoretically and empirically superior representation of AC data. Although research in the performance rating literature consistently supports a general factor of job performance (Viswesvaran et al., 2005), there is also consensus that performance is multifaceted (Hoffman, Blair, Meriac, & Woehr, 2007; King, Hunter, & Schmidt, 1980; Schmidt & Kaplan, 1971; Smith, Organ,

& Near, 1983; Viswesvaran et al., 2005). It is instructive that a key ad-

vance in management research over the past 25 years is the emergence of broadly applicable and theoretical models of the criterion domain (Austin

& Crespin, 2006). Insofar as ACs are designed as content valid samples

of performance, a correspondence between the structure of ACs and that of the relevant criterion space is both desirable and expected (Messick,

1995).

*Alternative model of the structure of AC ratings.* Conceptual tax- onomies of managerial performance (Borman & Brush, 1993; Campbell et al., 1990), theoretical (Smith, Organ, & Near, 1983) and empirical (Hoffman et al., 2007) models of organizational citizenship behavior, transformational leadership research (Avolio & Bass, 2004; Judge & Piccolo, 2004), measures of managerial skills (McCauley, Lombardo, & Usher, 1989; Yukl, 2006), and investigations of the internal structure of multisource performance ratings (Hoffman, Lance et al., 2010; Hoffman

& Woehr, 2009; Mount, Judge, Scullen, Sytsma, & Hezlett, 1998; Scullen

et al., 2000) routinely parameterize the structure of performance using “dimensions as items.” Using this approach, conceptually similar dimen- sions (e.g., analysis and judgment) are treated as indicators of broader latent factors when conceptually and empirically defining the structure of performance.

In a similar vein, an important factor limiting progress in AC research is the failure to ensure the validity of the espoused constructs assessed in ACs (Arthur & Villado, 2008). One approach of determining the match between espoused and actual constructs of ACs is to empirically group narrow dimensions onto broad dimension factors (Bray, 1982; Howard,

2008; Thornton & Gibbons, 2009). As noted by Arthur, Day, and Woehr

(2008):

the focus of measurement in the typical assessment center is dubious at best. *...* [T]he assessment center literature is in great need of more rigorous procedures for both the conceptualization and operationalization of dimen- sions. We should not be simply labeling ACs constructs/dimensions in an espoused manner. Instead we should be applying the standard test develop- ment and psychometric approaches and practices to demonstrating that they are actually measuring the intended constructs *before* [emphasis in original] we proceed with their use and also the comparative evaluation of dimen- sions and exercises. In other words, researchers and practitioners should

undertake formative evaluations of ACs before proceeding with summative evaluations like determining the relative contribution of dimension versus exercise variance in AC ratings. (pp. 107–108).

Despite the recognized need to subject ACs to rigorous psychomet- ric evaluation, organizational demands often prohibit the application of recommended professional standards to ACs, and many ACs continue to be designed around more dimensions than can be empirically supported (Arthur et al., 2008). Thus, in classic test development terms, an additional advantage of conceptually grouping dimensions is to provide an evalua- tion of the extent to which manifest dimensions satisfactorily represent the broader performance construct.1

Preliminary applications of the dimensions-as-items approach to ACs have yielded favorable results. For example, Arthur, Day, McNelly, and Edens (2003) classified existing PCDRs into one of seven broad categories in their meta-analysis of the criterion-related validity of AC dimensions.

Furthermore, Howard (2008) provided a sample Exercise × Dimension

grid, which organized narrow manifest dimensions into three broad di-

mension factors. Finally, predating more recent efforts, seminal work on the AC method hypothesized the presence of broad dimensions (OSS As- sessment Staff, 1948) and suggested the use of factor analysis to group manifest dimensions onto broad dimension factors (Bray, 1982). How- ever, this approach has not been applied to investigations of the structure of the more basic unit of measurement in ACs, the PEDR.

Depicted in Figure 1, Model 5 (J-broad dimensions, K-exercises, and

1-general performance factor) represents a novel conceptualization of the structure of AC ratings that parameterizes multiple broad dimension fac- tors, multiple exercise factors, and a single general performance factor. In essence, this model adds broad dimension factors to Model 4. By treating dimension ratings as manifest indicators (i.e., items) indicative of broader constructs, this approach takes a middle ground between the traditional dimensions and exercises model (Model 2) and the more recent exer- cises plus general performance model (Model 4). Informed by almost a century of work-performance research, this model recognizes that some dimensions will be more strongly related than others (e.g., leadership and persuasiveness) and parameterizes overlapping dimensions as indices of the same underlying latent factor. From this perspective, the high degree of overlap between two similar dimensions should not be interpreted as

1 Although we acknowledge that there is recent disagreement with operationalizing AC performance using within-exercise dimension ratings (Arthur et al., 2008; Howard, 2008), these ratings are necessary to tease apart exercise from dimension variance in ACs and have been the focus of AC research for almost 3 decades. Therefore, this study necessarily operationalizes AC performance using within-exercise dimension ratings.

Exercise 1

Exercise 2

E1-D1

E1-D2

E1-D3

E1-D4

E2-D1

E2-D2

E2-D3

E2-D4

Broad

Dimension1

General

Performance

Broad

Dimension 2

*Figure 1:* **Conceptual Assessment Center Factor Structure.**

evidence for a lack of discriminant validity; instead, overlap among sim- ilar constructs should be interpreted as evidence for convergent validity. Accordingly, we make the following prediction:

*Hypothesis 1*: A model specifying broad dimensions, exercises, and a general performance factor will best represent the structure of AC ratings.

*The Nomological Network of Assessment Centers Ratings*

Although most AC construct validity research has focused on the internal structure of ACs, to a lesser extent, research has also used nomo- logical network approaches (Cronbach & Meehl, 1955). The nomological network approach assesses convergent and discriminant validity based on correlations with externally measured constructs and has yielded some- what favorable results (Woehr & Arthur, 2003). From a unitarian per- spective on validity (American Educational Research Association, 1999; Landy, 1986; Messick, 1995), research investigating the pattern of rela- tions between AC ratings and external constructs is critical to inferences with respect to the validity of AC ratings. Despite the relatively positive results of research incorporating an external approach to investigating the construct validity of ACs (cf. Woehr & Arthur, 2003), very few studies incorporating a nomological network approach have modeled an empiri- cally supported structure of AC ratings. Instead, research examining the nomological network of ACs has relied almost exclusively on PCDRs and as a result has limited inferences as to the construct validity of ACs by con- founding the impact of exercises and dimensions (Wong, Law, & Huang,

2008). Accordingly, this study investigates the nomological network of performance, success, and individual differences surrounding empirically supported components of AC ratings.

*Performance and success.* Prior research consistently supports the validity of ACs in the prediction of managerial performance and of in- dicators of success such as promotion and salary (Arthur et al., 2003; Gaugler, Rosenthal, Thornton, & Bentson, 1987; Hermelin, Lievens, & Robertson, 2007; Meriac, Hoffman, Woehr, & Fleisher, 2008). Although the preponderance of AC criterion-related validity research has focused on the validity of a single summary index of AC performance, the over- all assessment rating (Gaugler et al., 1987; Hermelin et al., 2007), this approach does not allow for inferences with respect to the driving force behind the criterion-related validity of ACs (e.g., exercises, dimensions, or general performance). Arthur et al.’s (2003) meta-analysis of the criterion- related validity of AC PCDRs was an important step forward and revealed stronger multivariate validity coefficients than those found when only the summary overall assessment rating is used.

We are aware of only two studies that have investigated the criterion- related validity of empirically supported variance components (Lance et al., 2000; Lance, Foster et al., 2004). This research provided strong sup- port for the criterion-related validity of AC exercises and, at best, modest support for the criterion-related validity of the general performance factor. However, these studies were unable to support an AC structure including dimension factors, prohibiting inferences regarding the criterion-related validity of AC dimensions. Thus, an additional purpose of this study is to estimate the relative criterion-related validity of the proposed system- atic components of AC ratings (broad dimensions, exercise, and general performance factors). Based on past research that supports the criterion- related validity of dimensions (Arthur et al., 2003; Meriac et al., 2008) and of exercise factors and a general performance factor (Lance et al.,

2000; Lance, Foster et al., 2004), we hypothesize:

*Hypothesis 2*: Broad AC dimensions, exercises, and the general per- formance factor will be positively related to manage- rial performance and success.

A crucial, yet previously unexamined piece of information in understanding the value of conceptualizing AC performance using cross- situationally consistent versus situationally specific aspects of perfor- mance is the degree to which these systematic variance components contribute uniquely to the criterion-related validity of AC ratings. Re- cent research has questioned the value of dimensions to the design and interpretation of ACs (Lance et al., 2000; Lance, Foster et al., 2004; Lance,

2008a); this research has not directly compared the variance explained in

outcomes by exercises to that explained by dimensions. Although previ- ous evidence supports the validity of dimensions in the form of PCDRs (Arthur et al., 2003; Meriac et al., 2008), this research has not separated the role of dimensions from the role of exercises and general performance. However, based on Arthur et al.’s and Meriac et al.’s results, we predict that broad dimensions will account for variance in outcomes beyond that explained by exercises and general performance.

*Hypothesis 3*: Broad dimensions will explain variance in managerial performance and success beyond exercise and general performance factors.

*Individual differences.* Although the nomological network of per- formance and success criteria surrounding AC ratings is informative to understanding the predictive efficacy of AC ratings, such evidence does not provide inferences with respect to the constructs being mea- sured in ACs (Arthur & Villado, 2008; Hoffman & Woehr, 2009; Shore, Thornton, & Shore, 1990). A useful alternative is the inclusion of individ- ual difference correlates of AC ratings to draw inferences with respect to convergent and discriminant validity (Shore et al., 1990; Woehr & Arthur,

2003). In this study, we investigated the relationship between AC factors

and constructs from the five-factor model, dominance, and intelligence.

Given that AC analytical and performance style dimensions typically include cognitively loaded dimensions, we expected general mental abil- ity to be a significant correlate of AC problem-solving. Consistent with these suggestions, prior AC research has supported the overlap between cognitively loaded AC dimensions and intelligence (Dilchert & Ones,

2009; Meriac et al., 2008; Shore et al., 1990).

Next, prior research suggests that Extraversion and the associated facet dominance is among the most robust personality predictors of lead- ership (Hoffman, Woehr, Maldegan, & Lyons, 2010; Judge, Bono, Ilies,

& Gerhardt, 2002; Lord, DeVader, & Alliger, 1986), and past AC research supports the overlap between AC leadership and dominance (Kudisch &

Hoffman, 2002) and Extraversion (Dilchert & Ones, 2009; Meriac et al.,

2008).

Finally, effective interpersonal performance in ACs typically entails two related sets of behaviors. First, effective interpersonal performance requires assessees to responsibly resolve personnel issues and behave in a consistent, predictable way toward coworkers. Due to their characteris- tic predictability and high level of responsibility, conscientious assessees should perform better on interpersonally oriented dimensions. Consis- tent with this suggestion, LePine and Van Dyne (2001) found that Con- scientiousness was more strongly related to cooperative behaviors than task-oriented behaviors. Second, effective interpersonal performance also includes effective communication of ideas, demonstrating concern for the

well-being of others, and working effectively with others in one-on-one or group settings. Individuals with high levels of Extraversion are con- fident, sociable, and tend thrive when interacting with others. Assessees with lower levels of Extraversion are unlikely to possess the necessary confidence and facility in working with others to effectively resolve the interpersonal demands presented in AC exercises. This leads to the fol- lowing hypotheses:

*Hypothesis 4*: Intelligence will be positively related to cognitively oriented broad dimensions.

*Hypothesis 5*: Dominance and Extraversion will be positively related to leadership-oriented broad dimensions.

*Hypothesis 6*: Conscientiousness and Extraversion will be positively

related to interpersonally oriented broad dimensions.

Although past criterion-related validity evidence suggests that exercise effects represent substantively meaningful variance rather than method bias (Lance et al., 2000; Lance, Foster et al., 2004), the nature and mean- ing of exercise effects remain unclear (Brannick, 2008; Lance, 2008b).The examination of the nomological network of individual differences sur- rounding AC exercises can be particularly informative to shedding light on the substantive meaning of exercise factors. Although somewhat lim- ited, existing research has investigated the individual difference corre- lates of exercise performance. For example, Lance et al. (2000) and the meta-analysis by Collins et al. (2003) supported the relationship between intelligence and performance in in-basket exercises and leaderless group discussion (LGD) exercises. Given that in-baskets emphasize problem solving and usually do not assess interpersonal skills, it is not surpris- ing that they are cognitively loaded. In addition, past leadership research consistently revealed intelligence as the strongest individual difference predictor of leader emergence in unstructured settings (Judge, Colbert,

& Ilies, 2004; Lord et al., 1986), supporting a link between intelligence

and LGD performance. Unfortunately, due to the historic interpretation of exercise effects as method bias, research investigating the nomological network of exercise effects is in its infancy and has not considered the association between personality and exercise performance. Accordingly, we do not make specific predictions regarding the personality predictors of exercise factors.

*Research Question 1*: Will individual differences be related to AC

exercise factors?

Finally, work performance and, to a lesser extent, AC research has investigated the nomological network of general performance.

Conceptually, more intelligent employees are expected to learn the specifics of their job and engage in superior decision making and problem solving, yielding increased effectiveness across performance dimensions. Similarly, hard workers with strong detail orientation and a high need for achievement should excel at all aspects of their jobs. In line with this, research from the work performance and AC domains support the over- lap between intelligence and Conscientiousness and general performance (e.g., Collins et al., 2003; Hoeft & Schuler, 2001).

*Hypothesis 7*: Intelligence and Conscientiousness will be positively related to general AC performance.

*Method*

To investigate our hypotheses and research question, we used four independent samples consisting of participants in different organizations, from different professions, and with ACs used for different purposes.

*Participants and Procedure*

*Sample 1 (Banking).* Participants were 1,075 applicants or internal candidates (67% male) for an entry-level management position in a bank- ing organization in Germany. Their mean age was 34 years. As part of the selection process, the participants completed an administrative AC that assessed nine dimensions: conflict management, cooperation, decisive- ness, drive for results, leading and motivating employees, interpersonal skills, persuasiveness, sales orientation, and service orientation (defini- tions of the dimensions for this and the following samples are provided in the Appendix). Between three and five of the nine dimensions were rated during each of the following seven exercises: an LGD, a presenta- tion of the assessees’ evaluation of another manager’s performance (PR1), a one-on-one role play meeting with a probationary subordinate (RP1), a one-on-one meeting with a subordinate suffering from recent unchar- acteristic performance deficiencies (RP2), an analysis and presentation of the results of a customer survey (PR2), a one-on-one meeting with a subordinate in need of developmental feedback (RP3), and a meeting with a customer (CMG). Participants were rated by three assessors in each exercise. These assessors participated in a 2-day training program that was based on a combination of behavior observation training and frame-of-reference training (cf. Woehr & Huffcutt, 1994).

*Sample 2 (Utility).* Data were collected for 88 participants in an AC designed to facilitate administrative decisions in a large utility firm in the United States. A majority of the participants were Caucasian (90%) men

(85%) with a mean age of 44, and the AC was used to select managers into middle-to-upper-middle-level management positions. AC exercises were a role-play meeting with a direct report making a personnel-related re- quest (RP1), a meeting with a tenured subordinate suffering performance deficiencies (RP2), and an in-basket exercise (IB) where assessees as- sumed the role of a newly hired general manager.2 Between three and six of the following dimensions were rated in each exercise: judgment, anal- ysis, decision making, sensitivity, leadership, and confrontation. At least two assessors who had received 20 hours of frame-of-reference training provided ratings for each participant on each exercise. Participant salary the year the AC was administered was subtracted from participant salary

4 years after the AC to give an indicator of manager success (cf. Jansen

& Stoop, 2001).

*Sample 3 (Beverage).* The participants were 359 applicants for a line- supervisor position in a large international beverage organization. They were primarily Caucasian (76%) men with a mean age of 36 years. The administrative AC consisted of five exercises: a meeting with a subordi- nate in regards to a request (RP1), a meeting with a subordinate to discuss performance issues (RP2), a group discussion with fellow supervisors to discuss a product quality problem (LGD1), a group task where partici- pants had to work cooperatively to build models (LGD2), and a production film (PRF) requiring the applicant to observe and evaluate the effective- ness of a production unit. Between three and five of the following seven dimensions were rated in each exercise: planning, production and con- trol, problem solving and judgment, oral comprehension, interpersonal relations, developing others, oral communication, and written communi- cation. In each exercise, candidates were assessed by three assessors who received approximately 3.5 days of rater training.

Three indicators of performance were available for this sample. Specif- ically, the assessees’ direct supervisors ranked assessees relative to other employees in similar positions on 18 performance dimensions. In addition, the supervisors completed single-item performance measures correspond- ing to the assessees’ overall performance and a ranking of the assessees’ overall performance. The performance measures were collected for re- search purposes only, and the supervisors were unaware of the results of the AC when making ratings. Furthermore, assessees completed a mea- sure of general mental ability (GMA), the Watson Glaser Critical Thinking Appraisal (Watson & Glaser, 1980).

*Sample 4 (Executive master’s of business administration).* This sam- ple consisted of 495 managers enrolled in an executive master’s of

2 Although we use the same acronyms for the same type of exercise across the four samples, the actual exercises were different.

business administration (EMBA) program at a large southeastern uni- versity in the United States. While enrolled in the EMBA program, the participants concurrently worked as managers in a diverse range of orga- nizations and industries. A majority of the participants were Caucasian (82%) men (68%) with a mean age of 44. AC dimensions rated were analysis, judgment, leadership, oral communication, persuasiveness, and sensitivity. These dimensions were observed in an LGD where participants assumed the role of a school board with the responsibility of allocating a large financial gift, a meeting with a subordinate in need of develop- mental feedback (RP1), and a meeting with a subordinate that pressed to have a request granted (RP2). All dimensions were rated in each of the exercises. At least two assessors who each received approximately

20 hours of frame-of-reference training rated each participant in each exercise.

Responsibility and sociability were assessed using the California Psy- chological Inventory (CPI, Gough & Bradley, 1996) and were used to approximate Conscientiousness and Extraversion, respectively, from the five-factor model. These constructs were chosen based on Barrick and Mount’s (1991) classification of constructs in the five-factor model. In ad- dition, based on prior research substantiating the role of dominance over and above Extraversion in predicting managerial performance (Hoffman, Woehr et al., 2010; Judge et al., 2002), dominance was also included as an external correlate. Finally, GMA was assessed using the Watson Glaser Critical Thinking Appraisal (Watson & Glaser, 1980).

*General Method*

*Components underlying AC ratings.* We used LISREL 8.7 (Jo¨ reskog

& So¨ rbom, 2004) to compare the different models of the structure of AC ratings. For a given dimension to be included, it had to be assessed in more than one AC exercise and be classifiable into a broad dimension factor. The same progression of models was used to test the structure of ACs across the four samples. We tested the four traditional models listed above and multiple versions of Model 5 that varied the nature of the

broad dimensions. For the purposes of model identification, Dimension ×

Exercise correlations were set to zero, the general performance factor

was uncorrelated with the other factors, and error terms for the different

PEDRs were assumed to be uncorrelated.

Consistent with past research (Gignac, 2005; Hoffman, Lance et al.,

2010; Lance, Lambert et al., 2004; Lance et al., 2000; Mulaik & Quar- tetti, 1997; Scullen et al., 2000), we specified general performance as a first-order factor rather than a second-order factor composed of latent dimension first-order factors. Thus, general performance included vari- ance common to all manifest ratings in all exercises. This approach was

desirable in the present context because (a) it has been recommended when one is interested in estimating relationships between latent factors and ex- ternal constructs (Gignac, 2006); (b) a general factor operationalized as a second-order factor would mathematically exclude exercise-related vari- ance in correlated trait-correlated method (CTCM) models, a substantively meaningful component of general performance (Lance et al., 2000); and (c) it is consistent with prior AC (Lance, Lambert et al., 2004; Lance

et al., 2000) and 360◦ feedback models (Hoffman et al., 2010; Scullen

et al., 2000), facilitating a comparison with prior research.

*Structure of broad dimension factors.* We tested a variety of com- peting broad dimensions structures. To determine the pattern of loadings of the narrow manifest dimensions on broad dimension factors, we drew from dimension structures proposed in both the AC and the more gen- eral managerial performance literatures. Specifically, we relied on three established taxonomies of the structure of managerial performance to sort narrow manifest dimensions into broad dimension factors. We chose these taxonomies because of their popularity, existing evidence supporting their validity across a variety of settings, and their applicability across a wide range of managerial jobs.

First, prior research in a variety of domains has supported a two-factor

structure of performance that roughly corresponds to task and interper- sonal performance (Blake & Mouton, 1964; Judge, Piccolo, & Ilies, 2004; Katz & Kahn, 1951; Stogdill, 1963). In addition, Shore et al. (1990) supported a PCDR model that specified broad “performance style” and “interpersonal style” factors. Next, based on a content analysis of the dimensionality of managerial performance, Borman and Brush (1993) proposed a structure consisting of three megadimensions (as well as a fourth, “other” category): communication and interpersonal facilitation, technical activities/mechanics of management, and leadership and su- pervision. Both meta-analytic evidence (Conway, 1999) and large-scale primary studies (Hoffman et al., 2010) have empirically supported this structure of managerial performance ratings. A similar structure has been supported in prior AC research (Kolk, Born, & Van der Flier, 2004; Kudisch & Hoffman, 2002) and applied uses of the AC method (Howard,

2008).

Finally, we used Arthur et al.’s (2003) seven-factor taxonomy that categorizes manifest dimensions into communication, consideration and awareness of others, organizing and planning, problem solving, influenc- ing others, drive, and tolerance for stress/uncertainty. In addition to Arthur et al.’s support for the criterion-related validity of these dimensions, re- cent meta-analyses have supported the structural validity of these seven dimensions (Bowler & Woehr, 2006) as well as their incremental validity beyond one another and intelligence and personality (Dilchert & Ones,

2009; Meriac et al., 2008).

TABLE 1

*Classification of Broad Managerial Skill Dimensions Based on Popular*

*Taxonomies*

|  |  |  |
| --- | --- | --- |
| Arthur et al. (2003) | Borman and Brush (1993) | Shore et al. (1990) |
| Problem solving | Technical activities/mechanics of | Performance style |
|  | management |  |
| Organizing and |  |  |
| planning |  |  |
| Drive | Other useful behavior |  |
| Tolerance for stress |  |  |
| and uncertainty |  |  |
| Influencing others | Leadership and supervision | Interpersonal style |
| Communication | Communication and interpersonal |  |
|  | facilitation |  |
| Consideration and |  |  |

awareness of others

To specify the dimension structure underlying AC ratings, the manifest dimensions from each of the four ACs were classified into broad dimen- sion factors using these three performance taxonomies. In many cases, a given AC did not assess all of the dimensions proposed by the broad dimension taxonomies. For instance, none of the four ACs in this study measured all of Arthur et al.’s (2003) seven broad dimensions. Similarly, the AC from Sample 3 did not assess any dimensions indicative of Arthur et al.’s influencing others. In order to examine as many relevant dimension structures as possible, we incorporated multiple aspects of the three dif- ferent taxonomies, when possible, into a single model (e.g., a model could have components of Arthur et al.’s and Borman and Brush’s taxonomies).

Table 1 summarizes the different dimension structures used for this

study. Not surprisingly, many of the various performance taxonomies specify similar underlying constructs labeled with different names (e.g., consideration and awareness of others versus communication and inter- personal facilitation). Specifically, multiple dimension taxonomies specify broad dimensions roughly corresponding to interpersonal skills, concep- tual/administrative skills, and leadership skills. For the sake of clarity, we adopt these terms in the remainder of this paper to refer to these three broad dimensions rather than referring to each dimension using the various different labels that have been applied by prior research.

*Relationships with external variables.* Once the best-fitting CFA

model was determined, we estimated the relationships among AC factors and externally measured variables in three of the four samples (Samples

2, 3, and 4). Consistent with Lance et al. (2000), we (a) fixed the AC

parameter values to the estimates provided in the initial CFAs, (b) added

each externally measured variable to the model, and (c) allowed LISREL to freely estimate latent factor correlations among AC factors and the externally measured constructs.

To determine the degree to which dimensions explain variance in success and performance beyond exercises and a general factor (Samples

2 and 3), we conducted regression analyses on the basis of the latent factor correlations among AC factors and the externally measured constructs. Specifically, we conducted hierarchical multiple regressions by entering AC factors sequentially to the model in three steps: (a) exercises factors, (b) general performance, and (c) broad dimension factors. This order of entry was chosen to evaluate whether broad dimensions explain meaningful amounts of criterion variance beyond exercises and general performance.

*Results*

Because the general method used was consistent across samples, we discuss the methods and results from all samples simultaneously. Specif- ically, first we describe the results of the confirmatory factor analyses (Hypothesis 1). Next, we present the relationships among AC factors of the supported AC model and externally measured constructs (Hypotheses

2 and 4 to 7 and Research Question 1). For the two samples with available criterion data, we report multiple regression analyses to investigate the incremental validity of broad dimensions beyond exercises and general performance (Hypothesis 3).

*Internal Structure of ACs*

Table 2 shows the results from the CFA analyses. Consistent with prior research of the traditional AC models (Models 1–4), the J-exercises + gen-

eral performance model (Model 4) returned the closest fitting admissible solution for all four samples. However, consistent with Hypothesis 1, an AC structure specifying broad dimension factors, multiple exercise fac- tors, and a general performance factor (Model 5) also provided a close fit to the data for all four samples. In addition, this structure fit the data (often substantially) better than the other models, and every variation of the hypothesized model fit the data significantly better than did the cor-

responding J-exercises + general performance model (Model 4). As the only exception, for Sample 2, Model 2 (6 dimensions + 3 exercises) pro-

vided a closer fit to the data. However, as is often the case with this model

(cf. Conway, 1996; Lance et al., 2007), it returned an inadmissible solu- tion in all four samples and thus was uninterpretable. Together, the support for the proposed parameterization of AC ratings across four independent samples provided strong support for Hypothesis 1. Given the consistent

PERSONNEL PSYCHOLOGY

366

TABLE 2

*Model Fit Statistics for the Structure of ACs Across Four Samples*

Model df *χ* 2 RMSEA RMSR TLI CFI  *χ* 2 vs. Model 4

Sample 1: (Banking, *N* = 1,075)

1. 9 Dimensionsa

2. 9 Dimensions + 7 exercisesa

3. 7 Exercises 356 1629.76 0*.*062 0*.*029 0.98 0.98

4. 7 Exercises + G 327 1159.53 0*.*051 0*.*024 0.98 0.99

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 5a. 3 Dimensions + 7 exercises + G | 295 | 659.08 | 0*.*035 | 0*.*021 | 0.99 | 0.99 | 500*.*45∗∗ |
| 5b. 2 Dimensions + 7 exercises + G | 297 | 766.87 | 0*.*040 | 0*.*021 | 0.99 | 0.99 | 392*.*66∗∗ |

Sample 2 (Utility, *N* = 88)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 1. 6 Dimensionsb | 75 | 217.26 | 0*.*14 | 0*.*12 | 0.72 | 0.80 |
| 2. 6 Dimensions + 3 exercisesb | 57 | 73.42 | 0*.*05 | 0*.*068 | 0.96 | 0.98 |
| 3. 3 Exercises | 87 | 275.86 | 0*.*17 | 0*.*13 | 0.68 | 0.74 |
| 4. 3 Exercises + G | 72 | 195.75 | 0*.*15 | 0*.*11 | 0.75 | 0.83 |
| 5a. 3 Dimensions + 3 exercises + G | 54 | 70.44 | 0*.*06 | 0*.*056 | 0.96 | 0.98 |
| 5b. 2 Dimensions + 3 exercises + Gb | 56 | 72.75 | 0*.*06 | 0*.*054 | 0.96 | 0.98 |

125*.*31∗∗

123*.*00∗∗

*continued*

367

BRIAN J. HOFFMAN ET AL.

TABLE 2 (continued)

Model df *χ* 2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| RMSEA | RMSR | TLI | CFI | *χ* 2 vs. Model 4 |

Sample 3 (Beverage, *N* = 359)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1. 6 Dimensionsb | 137 | 834.24 | 0*.*13 | 0*.*11 | 0.64 | 0.71 |  |
| 2. 6 Dimensions + 5 exercisesb | 108 | 112.60 | 0*.*010 | 0*.*037 | 1.00 | 1.00 |
| 3. 5 Exercises | 142 | 396.77 | 0*.*074 | 0*.*071 | 0.87 | 0.89 |
| 4. 5 Exercises + G | 123 | 234.91 | 0*.*050 | 0*.*049 | 0.94 | 0.95 |
| 5a. 4 Dimensions + 5 exercises + G | 98 | 93.03 | *<* 0*.*001 | 0*.*029 | 1.00 | 1.00 | 141*.*88∗∗ |
| 5b. 3 Dimensions + 5 exercises + G | 101 | 94.39 | *<* 0*.*001 | 0*.*030 | 1.00 | 1.00 | 140*.*52∗∗ |
| 5c. 2 Dimensions + 5 exercises + G | 103 | 113.02 | 0*.*016 | 0*.*030 | 0.99 | 1.00 | 121*.*89∗∗ |

Sample 4 (EMBA, *N* = 495)

1. 6 Dimensionsb 120 1275.71 0*.*160 0*.*11 0.75 0.81

2. 6 Dimensions + 3 exercisesa

3. 3 Exercises 132 543.60 0*.*086 0*.*062 0.92 0.93

4. 3 Exercises + G 114 303.23 0*.*058 0*.*046 0.96 0.97

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 5a. 4 Dimensions + 3 exercises + G | 90 | 154.30 | 0*.*037 | 0*.*030 | 0.98 | 0.99 | 148*.*93∗∗ |
| 5b. 3 Dimensions + 3 exercises + G | 93 | 157.93 | 0*.*037 | 0*.*030 | 0.98 | 0.99 | 145*.*29∗∗ |
| 5c. 3 Dimensions + 3 exercises + G | 93 | 161.64 | 0*.*038 | 0*.*030 | 0.98 | 0.99 | 141*.*59∗∗ |
| 5d. 2 Dimensions + 3 exercises + G | 95 | 187.44 | 0*.*044 | 0*.*032 | 0.97 | 0.98 | 115*.*79∗∗ |

*Note*. a Model did not converge; b Model converged to an inadmissible solution; ∗∗ *p <* 0.01.

support for the hypothesized structure, we retained this general model for the remaining broad dimension structure comparisons. Specifically, we tested different models that retained the general performance and exercise factors but varied the specification of the broad dimension factors based on the taxonomies of managerial performance.

Based on the components recommended by the aforementioned dimen- sion taxonomies, we tested two broad dimension structures with Sample

1. Model 5a specified three broad dimensions. Two of the dimensions

corresponded to the distinction between leadership skills (leading and motivating employees, persuasiveness, and decisiveness) and interper- sonal skills (cooperation, interpersonal skills, and conflict management) outlined in both Arthur et al. (2003) and Borman and Brush (1993). The third dimension, drive (sales orientation, service orientation, and drive for results), was drawn from Arthur et al. Second, Model 5b in- cluded an interpersonal (leadership and interpersonal dimensions from Model 5a loading on a single interpersonal factor) and drive factor. For Sample 1, Model 5a specifying drive, interpersonal, and leadership latent factors provided the closest fit to the data, and it fit the data significantly

better relative to the two dimension model (Model 5b),  *χ* 2 (2) = 107.79,

*p <* 0.01. Accordingly, Model 5a was retained as the most appropriate

representation of AC ratings in Sample 1.

The dimensions assessed in Sample 2 allowed for the specification of two broad dimension structures. Corresponding roughly to Borman and Brush’s (1993) three-factor taxonomy, the first-dimension struc- ture (Model 5a) specified leadership skills (leadership and confronta- tion), conceptual/administrative skills (analysis, judgment, and decision- making), and interpersonal skills (sensitivity). The second-dimension structure (Model 5b) parameterized a two-dimension solution, specify- ing interpersonal (leadership, confrontation, and sensitivity) and concep- tual/administrative (analysis, judgment, and decision making) factors. For Sample 2, both the three (Model 5a) and two (Model 5b) dimension struc- tures provided a close fit to the data; however, the two-dimension model converged to an inadmissible solution. Thus, the three-dimension model (Model 5a) specifying conceptual/administrative, interpersonal, and lead- ership factors was retained for subsequent analyses.

Based on the dimensions assessed for Sample 3, we specified three broad dimension structures. Model 5a specified four of Arthur et al.’s

(2003) seven factors: organizing and planning (production, planning, and control), problem solving (problem solving and judgment), consideration and awareness of others (interpersonal orientation and developing others), and communication (oral communication and written communication). The dimension oral comprehension could not be classified into any of the broad dimensions, so this dimension was omitted from the analyses. Next,

this model was amended to form three broad dimensions (Model 5b): con- ceptual/administrative, interpersonal, and communication factors. This model combines the production, planning, and control dimensions and the problem-solving and judgment dimensions to form a factor reflective of the broader conceptual/administrative skills factor. Finally, a two-factor model (Model 5c) consisting of interpersonal (communication and interpersonal dimensions set to load on a single factor) and conceptual/administrative factors was specified. For Sample 3, the four-factor model (Model 5a) pro- vided the closest fit to the data. However, the three-factor model (Model

5b) did not result in a significant decrement in model fit,  *χ* 2 (3) = 1.36,

ns. Alternately, Model 5b fit the data significantly better than the two- factor model (Model 5c),  *χ* 2 (2) = 18.63, *p <* 0.01. Thus, Model 5b

specifying communication, interpersonal, and administrative/conceptual factors was adopted for subsequent analyses.

The dimensions assessed in Sample 4 allowed us to test four com- peting models of the dimension structure underlying this AC. Based on Arthur et al.’s (2003) framework, Model 5a consisted of communication (oral communication), interpersonal (sensitivity), leadership (leadership and persuasiveness), and conceptual/administrative (analysis and judg- ment) factors. Next, based on Borman and Brush’s (1993) taxonomy, oral communication and sensitivity were collapsed to form a single in- terpersonal factor, resulting in a three-factor model (Model 5b): inter- personal (oral communication and sensitivity), leadership (leadership and persuasiveness), and conceptual/administrative (analysis and judgment) factors. Model 5c also specified three factors: communication, concep- tual/administrative, and interpersonal skills (interpersonal and leadership factors from Model 5a set to load on a single factor). Finally, we specified a two-factor model (Model 5d) consisting of interpersonal (interpersonal and communication load on a single factor) and conceptual/administrative (analysis and judgment) factors.

In Sample 4, the four-factor model (Model 5a) specifying commu-

nication, interpersonal, leadership, and conceptual/administrative factors was the closest approximation of the data. Although this model repre- sented a marginally significant improvement in fit relative to Model 5c,

*χ* 2 (3) = 7.34, *ns*, the four-dimension model (Model 5a) did not fit the

data significantly more closely than the other three-factor model specify-

ing interpersonal (with the correlation between communication and inter- personal factors set to unity), leadership, and conceptual/administrative

factors (Model 5b),  *χ* 2 (3) = 3.63, *ns*. Because they are more parsimo-

nious, Model 5b and 5c were retained as more appropriate than Model

5a. On the other hand, Model 5b fit the data very slightly better than Model 5c. Furthermore, because Model 5b directly corresponds to an ex- isting performance taxonomy (Borman & Brush, 1993), this model was

retained. Finally, Model 5b fit the data significantly more closely than did the two-dimension structure (Model 5d),  *χ* 2 (2) = 29.51, *p <* 0.01.

Based on these results, Model 5b was retained for further analyses.

In summary, across each of the four samples, the proposed AC struc- ture (Model 5) provided a better representation of AC ratings than more traditional models, regardless of the underlying dimension specification parameterized. Further, only 1 of the 13 models paramaterizing broad dimensions, exercises, and general performance resulted in an improper solution, whereas all 4 models conforming to the traditional MTMM CFA model (Model 2) yielded improper solutions. Thus, our results highlight subtle differences in the underlying dimension structure across the ACs but more markedly substantiate the appropriateness of the hypothesized three component structure across the four samples. Together, these results strongly support Hypothesis 1.

As a next step, we determined the extent to which the different com- ponents account for variance in PEDRs. Tables 3–6 present standard- ized parameter estimates and the proportion of variance accounted for by broad dimensions, exercises, general performance, and the unique- nesses. And Tables 7–10 present the latent factor correlations for each of the four samples. To determine the proportion of variance accounted for, the completely standardized path loadings were squared, and the grand mean for each component was computed. Across the four samples, cross- situationally consistent performance accounted for 27% of the variance in AC ratings, with broad dimensions accounting for an average of 13% of the variance and general performance accounting for an average of 14% of the variance. On the other hand, exercise-specific variance accounted for an average of 41% of the variance in AC ratings. In short, exercise-specific as well as cross-situationally consistent performance explains variance in AC ratings. Consistent with prior research (Woehr & Arthur, 2003), the relative importance of these components was highly variable across ACs.

*Criterion-Related Validity of AC Factors*

*Salary growth.* We investigated the relationship between salary growth and the latent AC factors in Sample 2. Because available data did not allow us to estimate the reliability of the single indicator of salary growth, the error term for salary growth was set to 0 and the factor load- ing to 1. Thus, this procedure conservatively assumes perfect reliability of salary growth. As shown in Table 8, of the seven latent AC factors, leadership was the only significant correlate of salary growth.

Next, we conducted multiple hierarchical regression analyses based on the latent factor correlation matrix by entering all of the exercises into the regression in the first step, general performance in the second step,

TABLE 3

*Standardized Parameter Estimates for Sample 1 (Banking, N* = *1,075)*

PEDR LGD PR1 PR2 RP1 RP2 RP3 CMG INTP LEAD Drive G Unique

LGD\_CO 0.80∗ 0.40∗ 0.20∗ 0.16

LGD\_PE 0.90∗ 0.15∗ 0.11∗ 0.15

LGD\_DR 0.90∗ 0.19∗ 0.13∗ 0.15

LGD\_IS 0.80∗ 0.37∗ 0.21∗ 0.17

PR1\_LM 0.86∗ 0.18∗ 0.18∗ 0.19

PR1\_CO 0.92∗ 0.09∗ 0.11∗ 0.13

PR1\_D 0.67∗ 0.49∗ 0.27∗ 0.24

PR1\_PE 0.71∗ 0.57∗ 0.31∗ 0.07

PR2\_PE 0.87∗ 0.17∗ 0.16∗ 0.19

PR2\_SO 0.88∗ 0.20∗ 0.23∗ 0.13

PR2\_SA 0.88∗ 0.21∗ 0.20∗ 0.16

RP1\_LM 0.81∗ 0.11∗ 0.41∗ 0.16

RP1\_CM 0.86∗ 0.15∗ 0.15∗ 0.22

RP1\_D 0.83∗ 0.20∗ 0.15∗ 0.25

RP1\_PE 0.86∗ 0.10∗ 0.35∗ 0.13

RP1\_IS 0.69∗ 0.04 0.57∗ 0.20

RP2\_LM 0.90∗ 0.12∗ 0.22∗ 0.13

RP2\_CM 0.89∗ 0.10∗ 0.15∗ 0.17

RP2\_CO 0.80∗ 0.03 0.27∗ 0.28

RP2\_DR 0.83∗ 0.21∗ 0.17∗ 0.24

RP3\_LM 0.87∗ 0.13∗ 0.25∗ 0.16

RP3\_CM 0.83∗ 0.20∗ 0.07 0.27

RP3\_DR 0.84∗ 0.24∗ 0.14∗ 0.22

RP3\_SO 0.79∗ 0.10∗ 0.33∗ 0.25

RP3\_SA 0.76∗ 0.19∗ 0.16∗ 0.36

CMG\_D 0.77∗ 0.11∗ 0.04 0.40

CMG\_SO 0.85∗ 0.06 0.21∗ 0.23

CMG\_SA 0.91∗ 0.11∗ 0.07 0.16

CMG\_IS 0.76∗ 0.12∗ 0.39∗ 0.26

*Note.* LGD = leaderless group discussion; PR = presentation; RP = role play; CMG = customer meeting; INTP = interpersonal skills; LEAD = leadership skills; G = general performance; CO = cooperation; PE = persuasiveness; DR = drive for results; IS = interpersonal skills; LM = leading and motivating employees; D = decisiveness; SO = service orientation; SA = sales orientation; CM = conflict management. ∗ *p <* 0.05.

and all broad dimension factors in the third step (Table 11). As pointed out by an anonymous reviewer, the strong negative correlation between interpersonal and leadership skills (cf. Table 8) has the potential to bias regression results. Given that the direction of this relationship was likely an artifact of the MTMM-based analyses stemming from the small size of this sample and the relatively small number of manifest indicators (Marsh, Hau, Balla, & Grayson, 1998), we removed the dimension inter- personal skills from the regression analyses. Although neither exercises nor general performance explained significant variance in salary growth, broad dimensions explained significant incremental variance in salary

growth beyond exercise and general performance factors,  *R*2 = .06, *p <*

0.05. Together, the three systematic variance components had a moderate

TABLE 4

*Standardized Parameter Estimates for Sample 2 (Utility, N* = *88)*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| PEDR | IB | RP1 | RP2 | ADMIN | LEAD | INTP G Unique | | |
| IB\_AN | 0.61∗ |  |  | 0*.*32∗ |  |  | 0*.*28∗ | 0.45 |
| IB\_JD | 0.98∗ |  |  | 0*.*09 |  |  | −0*.*12 | 0.01 |
| IB\_DM | 0.63∗ |  |  | 0*.*23 |  |  | −0*.*09 | 0.54 |
| RP1\_AN |  | −0*.*11 |  | 0*.*06 |  |  | −0*.*84∗ | 0.29 |
| RP1\_JD |  | −0*.*21 |  | −0*.*40∗ |  |  | −0*.*78∗ | 0.18 |
| RP1\_DM |  | 0*.*61∗ |  | −0*.*46∗ |  |  | −0*.*14 | 0.39 |
| RP1\_CNF |  | 0*.*31∗ |  |  | 0*.*57∗ |  | −0*.*53∗ | 0.29 |
| RP1\_SN |  | −0*.*18 |  |  |  | 0.65∗ | −0*.*49∗ | 0.30 |
| RP1\_LD |  | 0*.*60∗ |  |  | −0*.*38∗ |  | −0*.*70∗ | 0.01 |
| RP2\_AN |  |  | 0*.*23 | 0*.*09 |  |  | −0*.*51∗ | 0.58 |
| RP2\_JD |  |  | 0*.*18 | −0*.*16 |  |  | −0*.*46∗ | 0.53 |
| RP2\_DM |  |  | 0*.*77∗ | −0*.*26∗ |  |  | −0*.*14 | 0.33 |
| RP2\_CNF |  |  | 0*.*43∗ |  | 0*.*44∗ |  | −0*.*33∗ | 0.51 |
| RP2\_SN |  |  | 0*.*11 |  |  | 0.69∗ | −0*.*22 | 0.47 |
| RP2\_LD |  |  | 0*.*58∗ |  | −0*.*25∗ |  | −0*.*57∗ | 0.28 |

*Note.* IB = in-basket; RP = role play; ADMIN = conceptual/administrative skills; LEAD = leadership skills; INTP = interpersonal skills; G = general performance; AN = analysis; JD = judgment; DM = decision-making; CNF = confrontation; SN = sensitivity; LD = leadership. ∗ *p <* 0.05.

albeit nonsignificant relationship with salary, *R* = 0.29, *ns*. Based on the correlation and regression analyses, broad dimension factors appear to be important predictors of managerial success as operationalized by growth in salary 4 years after the completion of the AC.

*Performance.* For Sample 3, higher-level managers’ ratings of the participants’ performance on three different performance variables were used to investigate the criterion-related validity of each systematic vari- ance component. The 18 dimensional percentile rankings were set to load on a single latent performance factor. For the two single-item performance indicators, their respective loadings were set based on the internal consis- tency reliability estimate from Viswesvaran, Ones, and Schmidt’s (1996)

meta-analysis of the reliability of performance ratings (*α* = 0.86). As

shown in Table 9, one of the five exercise factors (LGD1) was significantly

related to all three performance variables. Furthermore, another exercise factor (PRF) was a significant correlate of the dimensional percentile ratings, but the remaining correlations between exercise factors and per- formance were nonsignificant. Concerning the validity of AC dimensions, the communication factor was a significant correlate of the percentile and global rating, and the interpersonal dimension was inexplicably negatively related to the percentile rating. The remaining correlations between di- mensions and performance ratings were nonsignificant. Finally, general performance was unrelated to the three criterion variables.

The multiple regression analyses (Table 11) revealed that exercise factors explained significant variance in each of the three criterion

373

BRIAN J. HOFFMAN ET AL.

TABLE 5

*Standardized Parameter Estimates for Sample 3 (Beverage, N* = *359)*

PEDR LGD1 LGD2 PRF RP1 RP2 COM INTP ADMIN G Unique

LGD1\_PSJ 0.77∗ −0*.*09 0*.*14∗ 0.38

LGD1\_DO 0.69∗ 0*.*01 0*.*10 0.52

LGD1\_PPC 0.79∗ −0*.*05 0*.*10 0.36

LGD2\_PSJ 0*.*51∗ 0*.*44∗ 0*.*02 0.54

LGD2\_PPC 0*.*50∗ 0*.*69∗ 0*.*18∗ 0.25

LGD2\_WC −0*.*03 0*.*49∗ 0*.*62∗ 0.37

LGD2\_IR 0*.*51∗ 0*.*45∗ 0*.*24∗ 0.48

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| LGD2\_OC 0*.*49∗ −0*.*18∗  PRF\_PSJ 0.67∗  PRF\_DO 0.60∗ −0*.*13∗ | | | | | |  | 0*.*48∗ | 0.50 |
| −0*.*06 | 0*.*27∗ | 0.47 |
|  | 0*.*27∗ | 0.55 |
| PRF\_PPC | 0.47∗ |  |  |  |  | 0*.*03 | 0*.*22∗ | 0.73 |
| PRF\_WC | 0.31∗ |  |  | 0*.*32∗ |  |  | 0*.*51∗ | 0.54 |
| PRF\_IR | 0.51∗ |  |  |  | 0*.*00 |  | 0*.*15∗ | 0.71 |
| RP1\_DO |  | 0.63∗ |  |  | 0*.*14 |  | 0*.*11 | 0.58 |
| RP1\_IR |  | 0.55∗ |  |  | 0*.*46∗ |  | 0*.*13∗ | 0.47 |
| RP1\_OC |  | 0.60∗ |  | 0*.*35∗ |  |  | 0*.*36∗ | 0.39 |
| RP2\_DO |  |  | 0.92∗ |  | −0*.*08 |  | 0*.*07 | 0.15 |
| RP2\_IR |  |  | 0.59∗ |  | 0*.*25∗ |  | 0*.*12 | 0.58 |
| RP2\_OC |  |  | 0.41∗ | 0*.*47∗ |  |  | 0*.*57∗ | 0.28 |

*Note*. LGD = leaderless group discussion; PRF = performance production; RP = role play; COM = communication skills; INTP = interpersonal skills; ADMIN = administrative/conceptual skills; G = general performance; PSJ = problem solving & judgment; DO = developing others; PPC = planning, production, & control; WC = written communication; IR = interpersonal relations; OC = oral communication. ∗ *p <* 0.05.

TABLE 6

*Standardized Parameter Estimates for Sample 4 (EMBA, N* = *495)*

PEDR LGD RP1 RP2 INTP LEAD ADMIN G Unique

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| LGD\_OC 0.36∗ −0*.*04 | | | | | | | 0*.*29∗ | 0.78 |
| LGD\_AN | 0.65∗ |  |  |  |  | 0.22 | 0*.*19∗ | 0.50 |
| LGD\_JD | 0.71∗ |  |  |  |  | 0.25∗ | 0*.*12 | 0.41 |
| LGD\_LD | 0.54∗ |  |  |  | 0.37∗ |  | 0*.*19∗ | 0.54 |
| LGD\_CNF | 0.56∗ |  |  |  | 0.15∗ |  | 0*.*21∗ | 0.62 |
| LGD\_SN | 0.33∗ |  |  | 0*.*25∗ |  |  | 0*.*25∗ | 0.77 |
| RP1\_OC |  | 0.41∗ |  | −0*.*26∗ |  |  | 0*.*35∗ | 0.64 |
| RP1\_AN |  | 0.47∗ |  |  |  | 0.61∗ | 0*.*11 | 0.40 |
| RP1\_JD |  | 0.66∗ |  |  |  | 0.35∗ | 0*.*20∗ | 0.41 |
| RP1\_LD |  | 0.72∗ |  |  | 0.27∗ |  | 0*.*33∗ | 0.29 |
| RP1\_CNF |  | 0.53∗ |  |  | 0.33∗ |  | 0*.*11 | 0.60 |
| RP1\_SN |  | 0.69∗ |  | 0*.*12 |  |  | 0*.*34∗ | 0.40 |
| RP2\_OC |  |  | −0*.*25∗ | −0*.*38∗ |  |  | 0*.*59∗ | 0.45 |
| RP2\_AN |  |  | 0*.*21 |  |  | 0.53∗ | 0*.*59∗ | 0.33 |
| RP2\_JD |  |  | 0*.*21 |  |  | 0.49∗ | 0*.*47∗ | 0.27 |
| RP2\_LD |  |  | −0*.*26∗ |  | 0.40∗ |  | 0*.*72∗ | 0.26 |
| RP2\_CNF |  |  | 0*.*16 |  | 0.27∗ |  | 0*.*53∗ | 0.47 |
| RP2\_SENS |  |  | −0*.*45∗ | 0*.*27∗ |  |  | 0*.*60∗ | 0.37 |

*Note.* LGD = leaderless group discussion; RP = role play; INTP = interpersonal skills; LEAD = leadership skills; ADMIN = administrative/conceptual skills; OC = oral commu- nication; AN = analysis; JD = judgment; LD = leadership; CNF = confrontation skills; SN = sensitivity; G = general performance, ∗ *p <* 0.05.

variables. In support of Hypothesis 3, dimensions explained incremental variance beyond exercises and general performance in the three outcomes. Together, exercises, broad dimensions, and general performance explained significant variance in the different criteria, with multiple correlations ranging from 0.27 to 0.33. In concert with the results from Sample 2, these results suggest that exercises and dimensions contribute to the criterion- related validity of AC ratings, whereas general performance is unrelated to criterion variables. Thus, partial support was found for Hypothesis 2. In addition, dimensions explained variance in managerial performance and salary increase beyond exercise and general performance factors, support- ing Hypothesis 3.3

3 We also examined whether exercises explain variance beyond dimensions by repeating all regression analyses with dimensions entered in the final step and exercises the first. Exercises explained significant variance beyond dimensions and general performance in the three criterion variables of Sample 3 but did not explain variance in salary growth beyond the other facets in Sample 2. A summary of these analyses is available from the first author upon request.

375

BRIAN J. HOFFMAN ET AL.

TABLE 7

*Latent Factor Correlations for Sample 1 (Banking)*

1 2 3 4 5 6 7 8 9 10 11

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1. LGD  2. PR1 | 1  0*.*24∗∗ | 1 |  |  |  |  |  | | | | |
| 3. PR2 | 0*.*39∗∗ | 0*.*35∗∗ | 1 |  |  |  |
| 4. RP1 | 0*.*31∗∗ | 0*.*28∗∗ | 0*.*43∗∗ | 1 |  |  |
| 5. RP2 | 0*.*18∗∗ | 0*.*33∗∗ | 0*.*27∗∗ | 0*.*27∗∗ | 1 |  |
| 6. RP3 | 0*.*30∗∗ | 0*.*45∗∗ | 0*.*34∗∗ | 0*.*30∗∗ | 0*.*46∗∗ | 1 |
| 7. CM | 0*.*28∗∗ | 0*.*35∗∗ | 0*.*43∗∗ | 0*.*44∗∗ | 0*.*27∗∗ | 0*.*35∗∗ | 1 |  |  |  |  |
| 8. Interpersonal | −0*.*05 | −0*.*05 | −0*.*07 | −0*.*06 | −0*.*17∗∗ | −0*.*16∗∗ | 05 | 1 |  |  |  |
| 9. Leadership | −0*.*01 | 0*.*01 | 0*.*01 | 0*.*01 | 0*.*06 | 0*.*04 | 0*.*03 | 0*.*33∗∗ | 1 |  |  |
| 10. Drive | −0*.*05 | 0*.*09 | −0*.*01 | 0*.*00 | −0*.*05 | −0*.*01 | −0*.*02 | 0*.*39∗∗ | 0*.*86∗∗ | 1 |  |
| 11. General performance | 0 | −0*.*01 | −0*.*01 | 0*.*00 | −0*.*04 | −0*.*03 | 0*.*00 | 0*.*12 | −0*.*04 | 0*.*03 | 1 |

*Note.* LGD = leaderless group discussion; PR = presentation; RP = role play; CM = customer meeting. ∗ *p <* 0.05, ∗∗ *p <* 0.01.

TABLE 8

*Latent Factor Correlations for Sample 2 (Utility)*

1 2 3 4 5 6 7 8

1. IB 1

2. RP1 −0*.*08 1

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 3. RP2 | 0*.*08 | 0*.*67∗∗ | 1 |  | | | |
| 4. Administrative/conceptual | 0*.*01 | −0*.*03 | 0*.*00 | 1 | | | |
| 5. Leadership | 0*.*12 | 0*.*03 | 0*.*08 | 0*.*01 1 | | | |
| 6. Interpersonal | 0*.*01 | 0*.*00 | −0*.*15 | −0*.*19 | −0*.*85∗∗ | 1 |  |
| 7. General performance | 0*.*01 | 0*.*00 | 0*.*01 | 0*.*00 | −0*.*01 | 0*.*00 | 1 |
| 8. Salary growth | 0*.*01 | −0*.*02 | 0*.*08 | 0*.*01 | 0*.*26∗ | −0*.*15 | 0*.*10 1 |

*Note*. IB = in-basket exercise; RP = role play. ∗ *p <* 0.05, ∗∗ *p <* 0.01.

*Nomological Network of AC Factors*

*Individual differences.* We investigated the nomological network of individual differences surrounding the systematic variance components associated with AC ratings in Samples 3 (Table 9) and 4 (Table 10). In partial support of Hypothesis 4, GMA was a significant correlate of conceptual/administrative skills in Sample 4 but not in Sample 3. Al- though dominance was significantly related to leadership, Extraversion was not, providing partial support for Hypothesis 5. Finally, Hypothesis

6 was partially supported, as Conscientiousness was significantly related to interpersonal skills, but unexpectedly Extraversion was not.

In addition, a few individual difference constructs were significant,

albeit modest correlates of AC exercise factors in Samples 3 and 4, pro- viding affirmative evidence for Research Question 1. Extraversion and Conscientiousness were weakly related to RP1 in Sample 4. GMA was related to the LGD1 factor in Sample 3 and the LGD factor in Sample 4. GMA was also significantly related to the PRF factor in Sample 3. Finally, Hypothesis 7 was partially supported: Although general AC performance was unrelated to GMA in Sample 4, general performance was significantly related to Conscientiousness in Sample 4 and GMA in Sample 3.

*Discussion*

Decades of research have questioned the appropriateness of inter- preting ACs on the basis of cross-situationally consistent performance dimensions, and more recently, it has been proposed that performance dimensions be abandoned entirely in favor of task-based design, scoring, and interpretation of ACs (Jackson et al., 2005; Lance, 2008a,b; Lance, Foster et al., 2004; Lance et al., 2000). This study reevaluated the as-

377

BRIAN J. HOFFMAN ET AL.

TABLE 9

*Latent Factor Correlations for Sample 3 (Beverage)*

1 2 3 4 5 6 7 8 9 10 11 12 13

10. GMA 0*.*27∗∗ 0*.*06 0*.*21∗∗ 0*.*09 0*.*02 0*.*14 0*.*09 −0*.*03 0*.*19∗∗ 1

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1. LGD1  2. LGD2 | 1  0*.*35∗∗ | 1 |  |  |  |  | | | |
| 3. PRF | 0*.*20∗∗ | 0*.*18∗ | 1 |  |  |
| 4. RP1 | 0*.*17∗ | 0*.*08 | 0*.*14∗ | 1 |  |
| 5. RP2 | 0*.*19∗∗ | 0*.*39∗∗ | 0*.*19∗∗ | 0*.*41∗∗ | 1 |
| 6. Communication | −0*.*07 | 0*.*02 | 0*.*03 | −0*.*12 | 0*.*08 | 1 |  |  |  |
| 7. Interpersonal | −0*.*04 | 0*.*01 | 0*.*01 | −0*.*03 | 0*.*01 | −0*.*01 | 1 |  |  |
| 8. Administrative/conceptual | 0*.*02 | 0*.*03 | −0*.*08 | 0*.*00 | 0*.*03 | 0*.*20∗∗ | −0*.*03 | 1 |  |
| 9. General performance | 0*.*00 | −0*.*01 | 0*.*00 | 0*.*00 | −0*.*01 | −0*.*01 | 0*.*01 | 0*.*01 | 1 |

11. Percentile rating 0*.*16∗ 0*.*08 0*.*10 0*.*01 0*.*06 0*.*22∗∗ −0*.*14 0*.*13 0*.*02 0*.*02 1

12. Global rating 0*.*18∗∗ 0*.*11 0*.*11 −0*.*03 0*.*08 0*.*20∗∗ −0*.*07 0*.*03 0*.*04 0*.*04 0*.*86∗∗ 1

13. Dimensional percentile rating 0*.*15∗ 0*.*14 0*.*15∗ 0*.*05 0*.*09 0*.*12 −0*.*09 0*.*01 0*.*07 0*.*07 0*.*65∗∗ 0*.*85∗∗ 1

*Note.* LGD = leaderless group discussion; PRF = performance production, RP = role play. ∗ *p <* 0.05, ∗∗ *p <* 0.01.

TABLE 10

*Latent Factor Correlations for Sample 4 (EMBA)*

1. RP1 1

1 2 3 4 5 6 7 8 9 10 11

2. RP2 −0*.*34∗∗ 1

PERSONNEL PSYCHOLOGY

378

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 3. LGD  4. Interpersonal | 0*.*20∗∗  0*.*00 | −0*.*27∗∗  0*.*03 | 1  −0*.*12 | 1 | 1 | | | | | |
| 5. Administrative/conceptual | 0*.*01 | 0*.*01 | 0*.*02 | 0*.*01 |
| 6. Leadership | −0*.*11 | 0*.*04 | 0*.*05 | −0*.*09 | 0.72∗∗ | 1 | | | | |
| 7. General performance 0*.*01 0*.*00 −0*.*01 0*.*00 0.00 0*.*01 1 | | | | | | | | | | |
| 8. Dominance | 0*.*09 | −0*.*06 | 0*.*02 | −0*.*03 | 0.08 | 0*.*18∗∗ | 0.16∗ | 1 |  |  |
| 9. Extraversion | 0*.*13∗ | −0*.*10 | −0*.*09 | −0*.*05 | 0.04 | 0*.*09 | 0.10 | 0.60∗∗ | 1 |  |
| 10. Conscientiousness | 0*.*14∗ | −0*.*04 | 0*.*06 | 0*.*18∗ | 0.05 | −0*.*06 | 0.13∗ | 0.37∗∗ | 0.23∗∗ | 1 |
| 11. GMA | 0*.*07 | −0*.*05 | 0*.*18∗∗ | 0*.*02 | 0.16∗ | −0*.*06 | 0.10 | 0.12∗ | 0.10 | 0.26∗∗ |

1

Note. RP = role play, LGD = leaderless group discussion; GMA = general mental ability. ∗ *p <* 0.05, ∗∗ *p <* 0.01.

TABLE 11

*Salary Growth and Effectiveness Regressed on AC Factors for Samples 2 (N* = *88) and 3 (N* = *359)*

*β R*2  *R*2

Salary—Sample 2

Step 1 0*.*01 0*.*01

In-basket 0*.*01

Role play 1 −0*.*04

Role play 2 0*.*10

Step 2 0*.*02 0*.*01

General performance 0*.*10

Step 3 0*.*08 0*.*06∗

Conceptual/administrative 0*.*11

Leadership 0*.*26∗

Percentile rating—Sample 3

Step 1 0*.*03∗ 0*.*03∗

Leaderless group discussion 1 0*.*15∗∗

Production film 0*.*07

Role play 1 *<*0*.*01

Role play 2 *<*0*.*01

Leaderless group discussion 2 0*.*01

Step 2 0*.*03∗ *<*0*.*01

General performance 0*.*02

Step 3 0*.*11∗∗ 0*.*08∗∗

Communication 0*.*21∗∗

Interpersonal −0*.*13∗

Conceptual/administrative 0*.*09

Global rating—Sample 3

Step 1 0*.*05∗∗ 0*.*05∗∗

Leaderless group discussion 1 0*.*17∗∗

Production film 0*.*07

Role play 1 −0*.*06

Role play 2 0*.*04

Leaderless group discussion 2 0*.*03

Step 2 0*.*05∗∗ *<*0*.*01

General performance 0*.*04

Step 3 0*.*09∗∗ 0*.*04∗∗

Communication 0*.*20∗∗

Interpersonal −0*.*07

Administrative/conceptual −0*.*01

Dimensional percentile rating—Sample 3

Step 1 0*.*04∗ 0*.*04∗

Leaderless group discussion 1 0*.*10

Production film 0*.*11∗

Role play 1 0*.*02

Role play 2 *<*0*.*01

Leaderless group discussion 2 0*.*08

*continued*

TABLE 11 (continued)

*β R*2  *R*2

Step 2 0*.*05∗ 0*.*01

General performance 0*.*07

Step 3 0*.*07∗ 0*.*02∗

Communication 0*.*13∗∗

Interpersonal −0*.*09

Administrative/conceptual −0*.*02

*Note*. ∗ *p <* 0.05, ∗∗ *p <* 0.01.

sumptions and conclusions of prior AC research by investigating a novel structure of AC ratings and supplementing this analysis with an exami- nation of the nomological network of empirically supported AC factors in order to provide a more holistic view of the validity of ACs (Messick,

1995). In contrast to prior factor analytic work that has consistently failed to support the presence of dimensions (cf., Lievens, 2009), across four independent samples our results revealed that AC ratings are most appro- priately characterized by a structure consisting of broad dimension factors, exercise factors, and a general performance factor. In addition, both the rel- ative variance explained by these factors and the criterion-related validity results point to the importance of exercises and dimensions in under- standing the “active ingredient” of ACs. Finally, the correlations between these variance components and individual differences afford preliminary inferences with respect to the construct validity of broad dimensions and general performance as well as the meaning of the controversial exercise effects.

*Structure of AC Ratings*

By testing a model that includes exercises, general performance, and broad performance dimensions, this study arrived at markedly different conclusions than those of prior internal structure research. The appropri- ateness of this structure is bolstered by supportive evidence stemming from four samples. Of note, this model did not suffer from the improper solutions that have plagued the default correlated-trait correlated-method parameterizations of the AC structure (Lance et al., 2007). The consistent superiority of the three-component model relative to prevailing models in the AC literature provides compelling evidence for the appropriate- ness of this AC structure and, by extension, the presence of performance dimensions in the AC architecture.

The primary divergence between our model and prior AC models is the specification of broad dimensions. The broad dimension approach

recognizes that many dimensions assessed in a given AC exercise are ac- tually indicators of broader dimension factors by setting multiple similar PEDRs to load on the same broad latent dimension factor. By specifying broader dimensions consisting of multiple conceptually similar mani- fest dimensions, the resulting dimension factors are more distinguishable, yielding a higher likelihood of model convergence. Indeed, this approach is common when examining the structure of summary PCDRs (e.g., Arthur et al., 2003; Bray, 1982; Shore et al., 1990; Thornton & Byham, 1982), as well as when investigating the structure of the relevant criterion do- main (Hoffman, Lance et al., 2010). Accordingly, this study contributes to the literature by providing evidence of the usefulness of this approach in understanding the structure of AC within-exercise dimension ratings and signals the need for additional research investigating the appropriateness of this structure.

Similarly, although new to the study of the structure of ACs, the proposed three-component AC structure closely approximates the struc- ture previously found to characterize managerial performance ratings (Hoffman, Lance et al., 2010; Scullen et al., 2000). Thus, a primary the- oretical contribution of this study is the integration of general managerial performance research with AC research in the development of a more the- oretically informative model of the structure of AC ratings. In addition, the relative proportion of variance explained by dimension factors, exercise factors, and a general performance factor across ACs in this study was quite consistent with the magnitude of their analogs in measures of man- agerial performance (Hoffman, Lance et al., 2010; Scullen et al., 2000). The consistency between the AC structure and measures of the targeted criterion domain therefore supports the validity of ACs and underscores the generalizability of our a priori three-component model. In addition to the consistency in the basic components of the internal structure, the content of the broad dimensions was consistent with those proposed in performance taxonomies (Arthur et al., 2003; Borman & Brush, 1993). For instance, consistent with the managerial performance models, all four ACs included broad dimensions indicative of interpersonal skills and some variant of task skills. In addition, three of the four included a broad di- mension indicative of leadership. It should be noted that many of the dimension factors were very strongly intercorrelated, indicating the po- tential for a lack of discriminant validity. However, the reduction in the number of factors resulted in significantly worse model fit, supporting the discriminability of these dimensions. In addition, the magnitude of the correlations is consistent with those stemming from on-the-job measures of these dimensions (Hoffman & Woehr, 2009), supporting the generaliz- ability of our findings.

Despite these central similarities, aspects of the dimension structure varied based on the content of the AC. Similar to Campbell et al.’s (1990) performance model, not all components of a given performance taxonomy are expected to be present in a given job. Instead, the various performance dimensions should be weighted, depending on the nature of the job. For instance, it is instructive that although a broad leadership dimension factor was supported in three of the four ACs, no manifest dimensions indica- tive of Arthur et al.’s “influencing others” were assessed in Sample 3. As suggested by Thornton and Byham (1982), when applied to lower levels of management, ACs are most appropriately focused on the as- sessment of more basic dimensions, such as communicating with others, rather than higher-level dimensions such as strategic planning. In line with this, Sample 1 did not include multiple dimensions indicative of concep- tual/administrative skills. Because this AC was designed for an entry-level management position where customer contact is a defining characteristic of the job, this AC focused on service and sales orientation. Thus, despite support for a similar underlying model, the makeup of the dimensions varied in accordance with the nature of the focal criterion space.

In addition, this study contributes to the literature by providing re- vised estimates of the proportion of variance accounted for by situation-

ally specific and cross-situationally consistent aspects of performance in AC ratings. Lance, Lambert et al. (2004) found that exercise factors were approximately four times as large as a cross-situationally consistent vari- ance, whereas Bowler and Woehr (2006) found slightly larger exercise than cross-situationally consistent effects in their meta-analysis. Our re- sults are closer to those presented in Bowler and Woehr’s meta-analysis, with cross-situationally consistent aspects of performance explaining an average of 27% and exercise specific variance accounting for an average of 41% of the variance in AC ratings across the four samples. However, because our study (a) supported the presence of dimensions without the use of post hoc parameter constraints, (b) simultaneously parameterized the effect of dimensions, exercises, and general performance based on a theoretically justified structure of performance, and (c) isolated a better- fitting model than those tested in prior research, the estimates provided here likely reflect a more accurate approximation of the relative impor- tance of the variance components in ACs than those stemming from earlier work.

*Nomological Network of AC Ratings*

Following from a unitarian perspective on validity (American Ed- ucational Research Association, 1999; Landy, 1986; Messick, 1995), we investigated the relationship between AC structural components and

indices of performance, success, and individual differences to draw inferences with respect to the validity of ACs. Given their use in decision- making contexts, the degree to which ACs predict organizationally rel- evant outcomes is arguably the most important criterion when judging their value to organizations (Jones & Klimoski, 2008). Our results sup- port the criterion-related validity of ACs in two independent samples with exercises and broad dimensions consistently explaining practically mean- ingful proportions of the variance in outcomes. Interestingly, the overall criterion-related validity of ACs was highly consistent with the magni- tudes reported in meta-analyses of both the overall assessment rating and PCDRs (Arthur et al., 2003; Gaugler et al., 1987; Hermelin et al., 2007; Meriac et al., 2008). However, this study extends prior criterion-related validity research by investigating the relative validity of empirically sup- ported aspects of ACs. In doing so our results help to paint a picture of the “active ingredient” of ACs and answer recently posed questions with respect to the centrality of dimensions to the AC method. Finally, these results are supplemented by an investigation of the convergence between individual differences and AC ratings.

*Dimensions.* In contrast to prior internal structure research question- ing the value of dimensions in the design and interpretation of ACs, broad

dimension factors explained similar proportions of variance in effective- ness and success criteria relative to exercises and general performance in two independent samples, *and* they explained incremental variance be- yond exercises and general performance. Based on the regression results, exercises explained an average of 3% of the variance in outcomes, and broad dimensions explained an average of 5% of the variance in criterion variables beyond exercise factors and general performance. Although the internal structure results indicated that exercise effects explained more variance in AC ratings than either broad dimensions or general perfor- mance, dimensions appear to be a critical component of the criterion- related validity of ACs and, accordingly, should remain a central aspect of the AC method.

The relationships among individual differences and broad dimension factors also provided a measure of support for the validity of AC dimen-

sions. For instance, consistent with our expectations and prior research (Hoffman, Woehr et al., 2010; Judge et al., 2002; Lord et al., 1986), dom- inance is among the most robust personality predictors of leadership. On the other hand and in contrast to expectations, Extraversion was not sig- nificantly related to leadership or interpersonal skills. It is possible that operationalizing Extraversion using only sociability precludes detecting relationships in this context. Specifically, high need for affiliation has been proposed to be detrimental in leadership roles due to the associated desire to maintain close relationships at the possible expense of conveying bad

news, confronting others, and focusing on the tasks that need to get done. Because sociability does not capture the more agentic aspects of Extraver- sion (dominance, confidence), it is possible that more sociable assessees were simply too “nice” to complete some of the more unpleasant tasks necessary for effective AC performance. On the other hand, Conscien- tiousness was a significant correlate of interpersonal skills, a finding that likely reflects the importance of treating others consistently and personal responsibility central to maintaining productive working relationships. Next, as expected, GMA was the strongest individual difference corre- late of conceptual/administrative skills in Sample 4 but was unrelated to conceptual/administrative skills in Sample 3. A potential reason for this could be that the AC in Sample 4 was designed to simulate a higher level of management and Sample 3 simulated first-level supervisor. Thus, our findings might reflect the increased importance of intelligence at higher levels of management.

Although the magnitude of the trait-AC dimensions correlations was somewhat weak, it was consistent with a recent meta-analysis of the cor-

relation between personality and AC ratings (Meriac et al., 2008) and trait–behavior correlations in the broader leadership literature (Hoffman, Woehr et al., 2010). In addition, some evidence for discriminant validity is provided by the nonsignificant correlations between the individual dif- ferences and the remaining AC dimensions. For instance, intelligence was unrelated to the more interpersonally oriented dimensions in both ACs. Similarly, personality was unrelated to more task-oriented dimensions in Sample 4. Together, the internal structure, criterion-related validity, and, to some extent, the observed correlations with individual differences support the validity of broad dimensions.

*Exercises.* Despite clear evidence for the validity of AC dimensions,

our results also underscore the importance of exercise-specific variance for the criterion-related validity of ACs, with exercise factors explaining

1% to 5% of the variance in criterion variables. These results are consis-

tent with Lance’s (2008a; Lance et al., 2000) interpretation of exercise effects as valid indicators of situationally specific variance, central to the interpretation of ACs. Together, our results indicate that both exercises and dimensions are responsible for the criterion-related validity of ACs.

This study further sheds light on the validity of AC exercise factors by painting an initial picture of the nomological network of individual differ-

ences of AC exercises. First, GMA was related to the LGD in Sample 4 and one of the two LGDs in Sample 3, suggesting that overall performance in the LGD is a function of candidates’ intelligence, irrespective of the particular dimension being rated. The relationship between intelligence and performance in LGDs has been previously documented (e.g., Collins et al., 2003; Kiessling & Kalish, 1961) and is likely attributable to the

role of intelligence in emergent leadership (Lord et al., 1986). From this perspective, candidates with higher levels of intelligence are more likely to be familiar with the exercise materials and to speak up in the group setting, in turn emerging as leaders.

Next, in Sample 3, Extraversion was weakly correlated with perfor-

mance in RP1. This exercise required participants to successfully resolve a meeting with a concerned subordinate. In doing so, assessees must put the role player at ease and speak openly to control the direction of the meeting—behaviors presumed to underscore the correlation between Ex- traversion and leadership (Judge et al., 2002). Conscientiousness was also significantly and weakly related to performance in RP1. The linkage be- tween Conscientiousness and leadership has been proposed to rest in the tendency for conscientious leaders to provide specific developmental feed- back (Judge, Piccolo, & Kosalka, 2009). Given that this role play required participants to coach the role player on ways to improve, it is not surpris- ing that conscientious participants performed better across dimensions in this exercise.

Although the relationship between individual differences and exercise factors gives some insight into the meaning of AC exercise effects, the magnitude of the relationships was weak. The weak relationships are po- tentially a function of differences in the nature of underlying constructs assessed. Specifically, exercise effects reflect situationally specific perfor- mance, whereas traits are expected to impact behavior across situations. From this perspective, it should not be surprising that stable individual differences were not particularly strongly related to situationally specific aspects of behavior. Perhaps including additional constructs in the nomo- logical network of exercise factors would provide more favorable results.

For example, 360◦ feedback research shows that raters from different

sources provide unique, performance-relevant information (for a review of this research, see Lance, Hoffman, Gentry, & Baranik, 2008). An in- teresting avenue would be to evaluate whether exercise and source factors

converge when the level of the 360◦ feedback rater and the situation

presented in the exercise correspond.

*General AC performance.* Interestingly, general AC performance was not a significant predictor of managerial performance or success in ei- ther of the two samples with available data. The divergence between our findings and those of prior research (e.g., Lance, Foster et al., 2004) is potentially attributable to the omission of dimension effects in prior AC models. In other words, research that does not model dimensions runs the risk of misattributing dimensional variance to variance due to the general performance factor. When broad dimensions are modeled, the importance of the general performance factor is less pronounced and, possibly, the covariance with external measures is reduced.

The relationship between general performance and individual differ- ences helps to shed light on the meaning of the general performance factor. As predicted, the general performance factor was significantly, al- beit weakly, related to Conscientiousness in Sample 4 and to GMA in Sample 3. In addition, although not hypothesized, the significant correla- tion between general performance and dominance is not surprising given that the AC was used as a tool for the assessment of skills relevant to the leadership domain. Thus, despite the weak criterion-related validity of general AC performance, the convergence with theoretically relevant individual differences supports the validity of the general factor.

*Limitations*

Recent research has questioned the value of relying on MTMM-based analyses of PEDRs for AC research (Arthur et al., 2008; Howard, 2008). As PEDRs are the unit of measurement that is compared when undertaking consensus discussions, combining ratings, and understanding differences in candidates’ performance across exercises, we believe a complete aban- donment of research examining PEDRs to be premature (Lance, 2008a). Instead, our study shows that with the application of a construct-oriented approach, analyses of PEDRs can provide important information about the psychometric properties of ACs. In addition, the primary method- ological criticism with these analyses stems from the frequent improper solutions (Lance et al., 2007); thus, a primary contribution of this study is the elucidation of a model that circumvents the limitations associated with traditional parameterizations. In any case, until there is a richer un- derstanding with respect to the information that informs the rest of the assessment process (PEDRs), a strict reliance on the summary PCDRs is not a panacea.

Next, although we found strong support for the three-component model across the four samples, the broad dimension structure varied across the four ACs, leading to questions as to the generalizability of our findings. However, given that the four ACs actually targeted different constructs, were tailored to different professions, were intended for different levels of management (Mumford, Campion, & Morgeson, 2007), and incorporated different exercises, it should come as no surprise that a variety of differ- ent dimension structures emerged (Thornton & Byham, 1982). Indeed, just as the Minnesota Multiphasic Personality Inventory, Neuroticism- Extraversion-Openness Personality Inventory, and the Thematic Apper- ception Test would not be expected to measure the same underlying constructs simply because they are all personality scales, it is similarly untenable to assume that all ACs measure the same underlying constructs regardless of the content of the specific AC (Arthur & Villado, 2008).

Another potential limitation concerns the interpretation of the broad dimension factors. Specifically, these are often defined based on a limited number of manifest dimensions. Thus, there might be a concern whether these broad dimensions reflect a representative sample of the construct domain in each case. However, as a given AC dimension captures vari- ous behavioral indicators, the broad dimensions reflect the commonalities among a relatively large range of behavioral manifestations. In addition, the a priori specification of multiple competing performance models, the associated structural validity evidence, and the support for the broad di- mensions’ nomological network substantiates the validity of the broad dimensions. Consequently, the support for different broad dimensions across ACs is not a serious limitation to the generalizability of our results.

Similarly, our findings of substantial variance in the relative impor- tance of the three AC components across the four ACs is consistent with

prior AC research indicating that the relative magnitude of dimension and exercise factors differs substantially across ACs (Woehr & Arthur, 2003). One possible explanation for this effect in this study is the number of exercises used in each AC. There appears to be a linear trend between the number of exercises and the magnitude of the exercise effects. In the ACs with more exercises, the exercise effects were greater (and di- mension effects weaker), and in the ACs with fewer exercises, exercise effects were less pronounced (and dimension effects more pronounced). Perhaps the content of the exercises in ACs with more exercises is more circumspect. That is, by including more tasks, the information obtained in each task may be more specific. On the other hand, in ACs with fewer exercises, each exercise needs to pick up richer behavioral information. Consistent with recent suggestions to implement task-based approaches to AC design (e.g., Jackson et al., 2005), ACs with more exercises may actually be designed (at least implicitly) to assess more specific aspects of performance—resulting in stronger exercise effects. It is also possible that the exercises in lower-level ACs, as with Samples 1 and 3, assess more homogenous behavioral content, whereas ACs for higher levels, as with Samples 2 and 4 assess more heterogeneous behavioral content due to the high complexity inherent in upper levels of management. In any case, future research investigating these possibilities and other potential moderators of the magnitude of exercise, broad dimension, and general performance factors is needed.

*Practical Implications*

In substantiating the value of the dimensions as items approach, our results point to a potential limitation of previous AC research and practice. Specifically, the dimensions assessed in ACs are not typically subjected to

rigorous psychometric evaluation. Accordingly, our study also contributes to the literature by demonstrating a technique to evaluate the degree to which espoused constructs are actually being measured in ACs and, in do- ing so, supports the continued reliance of dimensions in applied settings. Although there is natural hesitancy by practitioners to reduce the number of dimensions in ACs, it is important to acknowledge that we do not advo- cate redesigning ACs to measure a few broad dimensions. Instead, as is the case with popular multisource feedback instruments (cf., McCauley et al.,

1989), the dimensions commonly assessed in ACs should be organized around empirically supported broad dimensions when interpreting AC performance, making administrative decisions, and relaying feedback.

In contrast to the common practice of assuming the validity of the un- derlying dimensions (see Arthur et al., 2008) in the face of limited evidence (Lance, 2008a), our approach allows for a much improved interpretation of ACs by providing an empirical foundation for their interpretation. Specifi- cally, traditional AC models rarely converge in CFAs, making it difficult to undertake traditional scale redesign procedures; however, given the con- vergent and admissible model presented here, our model is ideally suited to allow for the diagnosis of potential measurement problems. For in- stance, with variance due to exercises and general performance removed, some of the interpersonal manifest dimensions did not load significantly on their respective broad dimension factor. This finding is consistent with findings that interpersonally oriented performance dimensions are charac- terized by reduced levels of cross-method convergence (Bowler & Woehr,

2006) and lower reliability, in general (Viswesvaran, Schmidt, & Ones,

2002). Despite these benefits of the proposed model, the findings of non- significant and/or negative factor loadings are troubling. It is possible that this indicates that our a priori dimension classifications were in error or, more likely, that too little dimensional variance remained after removing variance due to exercises and general performance. When faced with sim- ilar findings, AC designers have two options: (a) redesign the exercise to more effectively elicit behavior relevant to the dimension (Lievens, Tett, & Schleicher, 2009) or (b) reevaluate whether the dimension can be effectively rated in the exercise and, if not, remove the dimension.

A second key implication is that exercise-specific variance should be considered valid information in decision-making and feedback contexts. Although our results imply the importance of considering exercises when providing developmental feedback, little if any research has investigated the effectiveness of exercise-oriented feedback. Given that exercise factors account for substantial variance in AC ratings, it seems justifiable to place more weight on exercise performance in delivering feedback than is often the case, and accordingly, research determining the appropriate methods of delivering such feedback is needed.

*Conclusion*

Despite recent assertions to the contrary, the results presented here support a multifaceted interpretation of ACs based on exercises, general performance, and broad dimensions. Based on these results, we believe that both dimension- and exercise-based interpretations have value and that arguments to focus exclusively on one aspect of AC performance at the expense of the others are counterproductive. We therefore encourage future AC research to devote energy toward further articulating a multi- faceted view of AC performance.

REFERENCES

American Educational Research Association, American Psychological Association, & Na- tional Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Associa- tion.

Anderson N, Lievens F, van Dam K, Born M. (2006). A construct-driven investigation of gender differences in a leadership-role assessment center. *Journal of Applied Psychology*, *91*, 555–566.

Arthur W, Jr., Day EA, McNelly, TL, Edens P. S. (2003). A meta-analysis of the criterion- related validity of assessment center dimensions. PERSONNEL PSYCHOLOGY, *56*,

125–154.

Arthur W, Jr., Day EA, Woehr DJ (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*, 105–111.

Arthur W, Jr., Villado AJ (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection research and practice. *Journal of Applied Psychology*, *93*, 435–442.

Austin JT, Crespin TR. (2006). Problems of criteria in industrial and organisational psy- chology: progress problems and prospects, In Bennett W, Lance CE, Woehr DJ (Eds.), *Performance measurement: current perspectives and future challenges,* (pp.

9–48). Mahwah, NJ: Erlbaum.

Avolio B, Bass B. (2004). *Manual for the multifactor leadership questionnaire*. Menlo

Park, CA: Mind Garden.

Barrick MR, Mount MK. (1991). The Big Five personality dimensions and job performance: A meta-analysis. PERSONNEL PSYCHOLOGY, *44*, 1–26.

Blake RR, Mouton JS (1964). *The managerial grid*. Houston, TX: Gulf.

Borman WC, Brush DH. (1993). More progress toward a taxonomy of managerial perfor- mance requirements. *Human Performance*, *6*, 1–21.

Bowler M, Woehr DJ. (2006). A meta-analytic evaluation of the impact of dimension and exercise factors on assessment center ratings. *Journal of Applied Psychology*, *91*,

1114–1124.

Brannick MT. (2008). Back to basics of test construction and scoring. *Industrial and*

*Organizational Psychology, 1*, 131–133.

Bray DW. (1982). The assessment center and the study of lives. *American Psychologist*,

*37*, 180–189.

Campbell CH, Ford P, Rumsey MG, Pulakos ED, Borman WC, Felker DB, de

Vera MV, Riegelhaupt BJ. (1990). Development of multiple OB performance

measures in a representative sample of jobs. PERSONNEL PSYCHOLOGY, *43*, 277–

300.

Collins JM, Schmitt FL, Sanchez-Ku M, Thomas L, McDaniel, MA, Le H. (2003).

Can basic individual differences shed light on the construct meaning of assess- ment center evaluations? *International Journal of Selection and Assessment*, *11*,

17–29.

Conway JM (1996). Analysis and design of multitrait-multirater performance appraisal studies. *Journal of Management*, *22*, 139–162.

Conway JM (1999). Distinguishing contextual performance from task performance for managerial jobs. *Journal of Applied Psychology*, *84*, 3–13.

Cronbach LJ, Meehl PE (1955). Construct validity in psychological tests. *Psychological*

*Bulletin*, *52*, 281–302.

Dilchert S, Ones DS. (2009). Assessment center dimensions: Individual differences corre- lates and meta-analytic incremental validity. *International Journal of Selection and Assessment, 17*, 254–270.

Gaugler BB, Rosenthal DB, Thornton GC, III, Bentson C. (1987). Meta-analysis of assess- ment center validity. *Journal of Applied Psychology*, *72*, 493–511.

Gignac GE. (2005). Revisiting the factor structure of the WAIS-R. *Assessment*, *12*, 320–329. Gignac GE. (2006). Evaluating subtest “*g*” saturation levels via the single trait-correlated uniqueness (STCU) SEM approach: Evidence in favor of crystallized subtests as the

best indicators of “*g*.” *Intelligence*, *34*, 29–46.

Gough HG, Bradley P. (1996). *California psychological inventory manual*. Palo Alto, CA: Counseling Psychologist’s Press.

Guilford JP. (1954). *Psychometric methods* (2nd ed.). New York, NY: McGraw-Hill. Hermelin E, Lievens F, Robertson IT. (2007). The validity of assessment centres for

the prediction of supervisory performance ratings: A meta-analysis. *International*

*Journal of Selection and Assessment*, *15*, 405–411.

Hoeft S, Schuler H. (2001). The conceptual basis of assessment centre ratings. *International*

*Journal of Selection and Assessment*, *9*, 114–123.

Hoffman BJ, Blair C, Meriac J, Woehr DJ. (2007). Expanding the criterion domain? A

meta-analysis of the OCB literature. *Journal of Applied Psychology*, *92*, 555–566. Hoffman BJ, Lance CE, Bynum BH, Gentry WA. (2010). Rater source effects are alive and

well after all. PERSONNEL PSYCHOLOGY, *63*, 119–151.

Hoffman BJ, Woehr DJ. (2009). Disentangling the meaning of multisource feedback source and dimension factors. PERSONNEL PSYCHOLOGY, *62*, 735–765.

Hoffman BJ, Woehr DJ, Maldegan R, Lyons B. (2010). Great man or great myth?

A meta-analysis of the relationship between individual difference and ef- fective leadership. *Journal of Occupational and Organisational Psychology*. doi:10.1348/096317909X485207

Howard A. (2008). Making assessment centers work the way they are supposed to. *Industrial and Organizational Psychology: Perspectives on Research and Practice*, *1*, 98–104. Jackson DJR, Stillman JA, Atkins SG. (2005). Rating tasks versus dimensions in assessment

centers: A psychometric comparison. *Human Performance*, *18*, 213–241.

Jansen PGW, Stoop BAM. (2001). The dynamics of assessment center validity: Results of a 7-year study. *Journal of Applied Psychology*, *86*, 741–753.

Jones RG, Klimoski RJ. (2008). Narrow standards for efficacy and the research playground: Why either-or conclusions do not help. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *1*, 137–139.

Jo¨ reskog K, So¨ rbom D. (2004). *LISREL 8.70*. Chicago, IL: Scientific Software International. Judge TA Bono JE, Ilies R, Gerhardt MW. (2002). Personality and leadership: A qualitative

and quantitative review. *Journal of Applied Psychology*, *87*, 765–780.

Judge TA, Colbert AE, Ilies R. (2004). Intelligence and leadership: A quantitative review and test of theoretical propositions. *Journal of Applied Psychology*, *89*, 542–552.

Judge TA, Piccolo RF. (2004). Transformational and transactional leadership: A meta- analytic test of their relative validity. *Journal of Applied Psychology*, *89*, 755–768.

Judge TA, Piccolo RF, Ilies R. (2004). The forgotten ones? The validity of consideration and initiating structure in leadership research. *Journal of Applied Psychology*, *89*,

36–51.

Judge TA, Piccolo RF, Kosalka T. (2009). The bright and dark sides of leader traits: A

review and theoretical extension of the leader trait paradigm. *Leadership Quarterly*,

*20*, 855–875.

Katz D, Kahn RL. (1951). Human organization and worker motivation. In Tripp LR (Ed.), *Industrial productivity* (pp. 146–171). Madison, WI: Industrial Relations Research Association.

Kiessling RJ, Kalish RA. (1961). Correlates of success in leaderless group discussions.

*Journal of Social Psychology*, *54*, 359–365.

King LM, Hunter JE, Schmidt FL. (1980). Halo in a multidimensional forced-choice performance evaluation scale. *Journal of Applied Psychology*, *65*, 507–516.

Kleinmann M, Kuptsch C, Ko¨ ller O. (1996). Transparency: A necessary requirement for the construct validity of assessment centres. *Applied Psychology: An International Review*, *45*, 67–84.

Kolk NJ, Born MP, Van Der Flier H. (2004). A triadic approach to the construct validity of the assessment center: The effect of categorizing dimensions into a feeling, thinking, and power taxonomy. *European Journal of Psychological Assessment*, 20, 149–156.

Kudisch J, Hoffman BJ. (2002, April). *Expanding the nomological network of assessment center dimensions using external measures of cognitive ability and personality*. Paper presented at the 17th Annual Conference of the Society for Industrial and Organizational Psychology, Toronto, Canada.

Lance CE. (2008a). Why assessment centers do not work the way they’re supposed to.

*Industrial and Organizational Psychology*, *1*, 87–100.

Lance CE. (2008b). Where have we been, how did we get there, and where shall we go?

*Industrial and Organizational Psychology: Perspectives on Science and Practice*,

*1*, 140–146.

Lance CE, Foster MR, Gentry WA, Thoresen JD. (2004). Assessor cognitive processes in an operational assessment center. *Journal of Applied Psychology*, 89, 22–35.

Lance CE, Hoffman BJ, Gentry W, Baranik LE. (2008). Rater source factors represent important subcomponents of the criterion construct space, not rater bias. *Human Resource Management Review*, 18, 223–232.

Lance CE, Lambert TA, Gewin AG, Lievens F, Conway JM. (2004). Revised estimates of dimension and exercise variance components in assessment center postexercise dimension ratings. *Journal of Applied Psychology*, *89*, 377–385.

Lance CE, Newbolt WH, Gatewood RD, Foster MR, French NR, Smith DE. (2000).

Assessment center exercise factors represent cross-situational specificity, not method bias. *Human Performance*, *13*, 323–353.

Lance CE, Woehr DJ, Meade AW. (2007). Case study: A Monte Carlo investigation of assessment center construct validity models. *Organizational Research Methods*, *10*,

430–448.

Landy FJ. (1986). Stamp collecting versus science: Validation as hypothesis testing. *Amer- ican Psychologist*, *41*, 1183–1192.

LePine JA, Van Dyne L. (2001). Voice and cooperative behavior as contrasting forms of contextual performance: Evidence of differential relationships with big five per- sonality characteristics and cognitive ability. *Journal of Applied Psychology*, *86*,

326–336.

Lievens F. (2008). What does exercise-based assessment really mean? *Industrial and Or- ganizational Psychology: Perspectives on Science and Practice*, *1*, 112–115.

Lievens F. (2009). Assessment centres: A tale about dimensions, exercises, and dancing bears. *European Journal of Work and Organizational Psychology*, *18*, 102–121.

Lievens F, Tett RP, Schleicher DJ. (2009). Assessment centers at the crossroads: Toward a reconceptualization of assessment center exercises. In Martocchio JJ, Liao H (Eds.), *Research in personnel and human resources management* (pp. 99–152), Bingley, UK: JAI Press.

Lord RG, DeVader C.L, Alliger GM. (1986). A meta-analysis of the relation between per- sonality traits and leadership perceptions: An application of validity generalization procedures. *Journal of Applied Psychology*, *71*, 402–409.

Marsh HW, Hau K, Balla JR, Grayson D. (1998). Is more ever too much? The number of indicators per factor in factor analysis. *Multivariate Behavioral Research*, *33*, 181-

220.

McCauley C, Lombardo M, Usher C. (1989). Diagnosing management development needs: An instrument based on how managers develop. *Journal of Management*, *15*, 389–

403.

Melchers KG, Henggeler C, Kleinmann M. (2007). Do within-dimension ratings in assess- ment centers really lead to improved construct validity? A meta-analytic reassess- ment. *Zeitschrift fu¨ r Personalpsychologie*, *6*, 141–149.

Melchers KG, Ko¨ nig CJ. (2008). It is not yet time to dismiss dimensions in assessment centers. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *1*, 125–127.

Meriac JP, Hoffman BJ, Woehr DJ, Fleisher M. (2008). Further evidence for the validity of assessment center dimensions: A meta-analysis of the incremental criterion-related validity of assessment center dimension ratings. *Journal of Applied Psychology*, *93*,

1042–1052.

Messick S. (1995). Validity of psychological assessment: Validation of inferences from per- sons’ responses and performance as scientific inquiry into score meaning. *American Psychologist*, *50*, 741–749.

Mount MK, Judge TA, Scullen SE, Sytsma MR, Hezlett SA. (1998). Trait, rater, and level effects in 360-degree performance ratings. PERSONNEL PSYCHOLOGY, *51*,

557–576.

Mulaik SA, Quartetti DA. (1997). First order or higher order general factor? *Structural*

*Equation Modeling*, *4*, 193–211.

Mumford TV, Campion MA, Morgeson FP. (2007). The leadership skills strataplex: Lead- ership skill requirements across organizational levels. *Leadership Quarterly*, *18*,

154–166.

OSS Assessment Staff (1948). *Assessment of men*. New York, NY: Rinehart.

Sackett PR, Dreher GF. (1982). Constructs and assessment center dimensions: Some trou- bling empirical findings. *Journal of Applied Psychology*, *67*, 401–410.

Schmidt FL, Hunter JE. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262–274.

Schmidt FL, Kaplan LB. (1971). Composite vs. multiple criteria: A review and resolution of the controversy. PERSONNEL PSYCHOLOGY, *24*, 419–434.

Scullen SE, Mount MK, Goff M. (2000). Understanding the latent structure of job perfor- mance ratings. *Journal of Applied Psychology*, *85*, 956–970.

Shore TH, Thornton GC, Shore LM. (1990). Construct validity of two categories of assess- ment center dimension ratings. PERSONNEL PSYCHOLOGY, *43*, 101–116.

Smith CA, Organ DW, Near JP. (1983). Organizational citizenship behavior: Its nature and antecedents. *Journal of Applied Psychology*, *68*, 653–663.

Stogdill RM. (1963). *Manual for the leader behavior description questionnaire, form XII*.

Columbus, OH: Bureau of Business Research, Ohio State University.

Thornton GC, Byham WC. (1982), *Assessment centres and managerial performance*, New

York, NY: Academic Press.

Thornton GC, Gibbons AM. (2009). Validity of assessment centers for personnel selection.

*Human Resource Management Review*, *19*, 169–187.

Viswesvaran C, Ones DS, Schmidt FL. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology*, *81*, 557–574.

Viswesvaran C, Schmidt FL, Ones DS. (2002). The moderating influence of job performance dimension on convergence of supervisory and peer ratings of job performance: Unconfounding construct-level convergence and rating difficulty. *Journal of Applied Psychology*, *87*, 245–354.

Viswesvaran C, Schmidt FL, Ones DS. (2005). Is there a general factor in ratings of job performance? A meta-analytic framework for disentangling substantive and error influences. *Journal of Applied Psychology*, *90*, 108–13.

Watson G, Glaser E. (1980). *Manual for the Watson Glaser critical thinking appraisal*.

New York, NY: Psychological Corporation.

Woehr DJ, Arthur W, Jr. (2003). The construct-related validity of assessment center rat- ings: A review and meta-analysis of the role of methodological factors. *Journal of Management*, *29*, 231–258.

Woehr DJ, Huffcutt AI. (1994). Rater training for performance appraisal: A quantitative review, *Journal of Occupational & Organizational Psychology 67*, 189–205.

Wong C, Law KS, Huang G. (2008). On the importance of conducting construct-level analysis for multidimensional constructs in theory development and testing. *Journal of Management, 34*, 744–764.

Yukl G. (2006). *Leadership in organizations* (6th ed.). Upper Saddle River, NJ: Pearson

Education.

APPENDIX

Assessment Center Dimension Definitions

Sample 1

*Conflict management*

Recognize potential conflicts in time and address them, do not avoid conflicts when own (appropriate) interests are concerned, constructive handling of conflicts, search good compromises, good negotiating skills

*Cooperation*

Promote cooperation with and between others as well as understanding for others’ points of view, facilitation/mediation skills, is able to promote constructive cooperation between others, understand diverging interests and is able to mediate between them

*Decisiveness*

Reach well-grounded decisions on the basis of available information, do not abandon well-grounded decisions, do not unnecessarily delay decision making

*Drive for results*

Act on the basis of self-set goals or goals provided by the task/situation, weight costs and benefits and search pragmatic solutions, introduce solutions to tasks/problems and push their implementation efficiently, direct the course of action toward a goal

*Leading and motivating employees*

Goal setting and monitoring, motivation of employees and delegation of tasks and responsibilities, acknowledge good performance, point out problems, promote personal initiative

*Persuasiveness*

Convince others with good arguments, respond to arguments/objections from others and is able to invalidate them, is well able of expressing himself, can stand and overcome opposition

*Sales orientation*

Search and take advantage of sales chances, ask for, discern and is responsive to customer needs, act to come to concrete results/deals, show potential advantages/benefits to customers, is competitive

*Service orientation*

Adjusts his/her behavior according to the needs of internal and external customers, regard himself/herself as a service provider, strive for optimal results/satisfaction of customer needs (without leaving sight for business objectives)

*Interpersonal skills*

Attract the attention of others, approach others actively, can easily come into contact even with different people, successfully build and maintain rapport to others

Sample 2

*Analysis*

Recognize major issues, show familiarity with the materials, probe for additional information, use available data, and integrate information across sources

*Judgment*

Provide sound rationale for decisions, forward alternative courses of action, and consider implications before taking action

*Decision-making*

Consistently make clear decisions and provide detailed action plans to ensure decisions are carried out properly

*Sensitivity*

Remain polite and attentive, distribute appropriate praise toward the role player, work to build rapport, and display empathy

*Leadership*

Is able to control the meeting, pursue an agenda, solicit input, motivate, and display influence during the meeting

*Confrontation*

Remain tactful and address potentially volatile topics and defend own and others’

perspectives

Sample 3

*Oral communication*

Ability to convey essential information effectively to others through verbal communications. This includes the ability to convey technical information orally to less technically versed employees

*Written communication*

Ability to communicate essential information effectively to others in clear and concise written form

*Interpersonal relations*

Ability to initiate and maintain cooperative and effective relationships. This includes sensitivity to cultural diversity

*Developing others*

Ability to instruct and motivate others to improve their performance. This includes sensitivity providing feedback and establishing long-term developmental plans

*Problem solving and judgment*

Ability to make cost-effective decisions that impact quality and/or productivity. This includes securing information and recognizing opportunities to generate new ideas and solutions

*Planning, production, and control*

Ability to plan and monitor activities and projects to ensure the desired outcome is achieved

Sample 4

*Analysis*

Recognize major issues, show familiarity with the materials, probe for additional information, use available data, and integrate information across sources

*Judgment*

Provide sound rationale for decisions, forward alternative courses of action, and consider implications before taking action

*Oral communication*

Speak in a way that was clear and understandable, use appropriate volume and tone, maintain eye contact, use a minimal amount of filler language, (e.g., “um,” “like”), and appropriately utilized expressive and/or enthusiastic language

*Sensitivity*

Remain polite and attentive, distribute appropriate praise toward the role player, work to build rapport, and display empathy

*Leadership*

Is able to control the meeting, pursue an agenda, solicit input, motivate, and display influence during the meeting

*Confrontation*

Tactfully and willingly address areas of conflict, correct misgivings, refute opposing viewpoints, and defend his/her own perspective

Copyright of Personnel Psychology is the property of Wiley-Blackwell and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.