PERSONNEL PSYCHOLOGY

2009, 62, 735–765

**DISENTANGLING THE MEANING OF MULTISOURCE PERFORMANCE RATING SOURCE AND DIMENSION FACTORS**

BRIAN J. HOFFMAN University of Georgia

DAVID J. WOEHR University of Tennessee

We extend multisource performance rating (MSPR) construct validity research by examining the pattern of relationships between factor analyt- ically derived MSPR rating source and performance dimension factors and externally measured constructs (e.g., assessment center dimensions, personality constructs, and intelligence). The pattern of relationships among MSPR dimensions and external constructs provides modest con- struct validity evidence for the MSPR dimensions. In addition, MSPR source factors were differentially correlated with externally measured constructs, suggesting that MSPR source effects represent substantively meaningful source specific variance, as opposed to bias. These findings are discussed in the context of managerial skill diagnosis and the efficacy of collecting performance data from multiple sources.

Multisource performance ratings (MSPR) are increasingly common- place in organizations as a method of measuring employee work per- formance and as an input into employee development (Church & Allen,

1997). Typically, organizational constituents’ ratings are presented to the

target, separated by performance dimensions and the organizational level of the rater. The usefulness of MSPRs in developmental settings is predi- cated on the validity of the dimension ratings (London & Smither, 1995). To the extent that invalid MSPRs result in misdiagnosed performance issues, the utility of an MSPR system will be compromised. For exam- ple, a measurement system providing feedback that is not a reflection of managers’ underlying performance level will likely result in the feedback recipient devoting resources to improving aspects of performance that are

This manuscript is based on Brian Hoffman’s dissertation from the University of Tennessee and was the recipient of the Society for Human Resource Management Dis- sertation Grant Award. An earlier draft was presented at the 33rd Annual Conference of the Society for Industrial and Organizational Psychology, San Francisco, CA. The primary author would like to thank Mike Rush, Lowell Gartner, Lane Morris, and Katie Helland for their helpful comments on earlier drafts of this manuscript.

Correspondence and requests for reprints should be addressed to Brian J. Hoffman, De- partment of Psychology, The University of Georgia, Athens, GA 30602-3013; hoffmanb@ uga.edu.

 C 2009 Wiley Periodicals, Inc.

735

acceptable to the detriment of performance deficiencies that warrant atten- tion. Toward this end, it has been suggested that poor measurement may be one of a variety of factors contributing to MSPR’s modest impact on performance improvement (Seifert, Yukl, & McDonald, 2003; Smither, London, & Reilly, 2005).

Corresponding to the increasing popularity of MSPR systems, sub- stantial prior research has investigated the psychometric properties of multisource ratings (Conway, 1996; Conway & Huffcutt, 1997; Hoffman, Lance, Bynum, & Gentry, 2008; Mount, Judge, Scullen, Systma, & Hezlett, 1998; Scullen, Mount, & Goff, 2000). The preponderance of this research has incorporated an “internal” approach to accumulating psychometric evidence where the covariance of ratings made on the same instrument by different raters is examined. For example, research demon- strating the moderating role of rater level on the interrater reliability of performance ratings suggests that raters at the same organizational level provide more consistent ratings than raters from different organizational levels (Conway & Huffcutt, 1997). Next, substantial research has ap- plied confirmatory factor analysis of multitrait–multimethod (MTMM; Campbell & Fiske, 1959) matrices to investigate the internal structure of MSPRs (e.g., Conway, 1996; Hoffman, Lance, Bynum, & Gentry, 2008; Lance, Baxter, & Mahan, 2006; Woehr, Sheehan, & Bennett, 2005). This research has typically indicated that both dimension and source effects ac- count for meaningful variance in MSPRs and that source effects typically explain (much) more variance in ratings than do performance dimension effects. Despite the contribution of prior internal structure research, some have argued that this stream of research does not provide sufficient evi- dence to evaluate the construct-related validity of MSPRs (Borman, 1997; Lance, Hoffman, Gentry, & Baranik, 2008; Woehr, et al., 2005). That is, although source and dimension factors are consistently supported in internal structure research, the psychometric soundness of these system- atic rating effects has rarely been examined. For example, although rater source effects are common to MSPR systems, there is no consensus as to the meaning of these effects. From a traditional psychometric perspective, these effects represent unwanted method variance that should be reduced, but from a practical perspective, they may represent important differences in performance perceptions across rating groups (Borman, 1974; Lance, Hoffman, et al., 2008; Tornow, 1993).

One method of examining the meaning of MSPR source and dimension effects is a nomological network approach to construct validity (Cronbach

& Meehl, 1955; Lance, Hoffman, et al., 2008; Scullen, et al., 2000;

Woehr, et al., 2005). Using such an approach, MSPR dimensions should be more strongly related to similar than dissimilar constructs. To the extent that source effects reflect systematic, performance-related variance, they

should correlate in predictable ways with the other constructs pertinent to managerial performance, and to the extent that they represent random ef- fects they should not (Lance, Baranik, et al., 2008; Lance, Hoffman, et al.,

2008; Woehr, et al., 2005). This study integrates the internal structure and nomological network approaches to construct validity to facilitate infer- ences with respect to the psychometric properties of MSPRs. Specifically, the primary purpose of this study is to examine the meaning of MSPR source effects by providing an examination of the nomological network of individual differences surrounding MSPRs, in an effort to answer the question, “What are source effects measuring?” In addition, we investigate construct validity evidence for MSPR performance dimensions by exam- ining the relationship between MSPR dimension factors and externally measured constructs.

*Construct Validity of MSPRs*

Despite the usefulness of internal approaches for understanding the psychometric soundness of MSPRs (Conway, 1996; Conway & Huff- cutt, 1997; Hoffman, et al., 2008; Lance, Baranik, et al., 2008; Lance, Hoffman, et al., 2008; Scullen, et al., 2000; Woehr, et al., 2005), this approach does not provide the entire picture with respect to the construct validity of MSPRs. More concretely, although internal structure research clearly points to pervasive rater level and performance dimensions effects in MSPRs, this method does not allow for conclusions with respect to the meaning of these effects. An alternate approach to examining the construct validity of MSPRs is through an examination of the pattern of relations between external constructs and performance ratings (Cronbach & Meehl,

1955). Toward this end, some research has examined the relationship be-

tween MSPRs and externally measured constructs. Beehr, Ivanitskaya, Hansen, Erofeev, and Gudanowski (2001) investigated the relationships among a variety of individual differences (e.g., intelligence and commu- nication skills) and peer, supervisor, and self-ratings of three dimensions of performance. Similarly, Conway, Lombardo, and Sanders (2001) pro- vided a meta-analysis examining differential relationships of personality constructs with peer and subordinate ratings. In addition, Atkins and Wood (2002) examined the extent to which self, peer, supervisor, and subordinate ratings were differentially related to performance in different performance tasks (meetings with customers, a behavioral interview, etc.). Importantly, the vast majority of past MSPR research examining external correlates suffers from an important limitation: the failure to specify source and di- mension effects prior to examining relationships with external correlates. Consequently, previous research has typically confounded source and

dimension factors when examining the relationship between MSPRs and externally measured constructs.

Although recent MSPR research is unclear with respect to the con-

struct validity of performance ratings, early performance rating research suggests that rater general impression (e.g., halo) moderates the validity of selection measures such that selection measures are actually *more* valid at higher levels of halo in performance ratings (Nathan & Tippins, 1990). These findings provide indirect support for the meaningfulness of rater factors in performance ratings. Lance, et al. (2006) took an additional step in this direction by examining the correlates of peer and supervisor ratings of Air Force mechanic performance. Their results revealed that both peer and boss factors were correlated with external variables (e.g., a work sample focused on building a jet engine). However, the sample for this study was atypical to that of typical MSPR contexts (e.g., sample of nonmanagers) and did not examine the validity of subordinate ratings, In addition, Lance, et al. (2006) did not investigate the possibility of differential correlations among source factors; thus, their results do not provide an indication as to whether the meaning of source factors differs by source. Finally, this study did not investigate the external correlates of dimension factors. Given previous factor analytic research consistently supporting an MSPR consisting of both source and dimension effects and indications of the meaningfulness of rater factors (Lance, et al., 2006; Nathan & Tippins, 1990), any examination of the construct validity of MSPRs using a nomological network approach must first model these two variance components.

*MSPR Dimension Effects*

Previous performance taxonomies (e.g., Borman & Brush, 1993; Hoffman, Blair, Meriac, & Woehr, 2007; Mann, 1965; Mintzberg, 1975; Smith, Organ, & Near, 1983), psychometric models of ratings (Kenny & Berman, 1980; King, Hunter, & Schmidt, 1980; Wherry, 1952; Wherry

& Bartlett, 1982), and performance evaluation instruments (Austin

& Villanova, 1992; DeVries, Morrison, Shullman, & Gerlach, 1986; Fleishman, 1957; McCauley & Lombardo, 1990) conceptualize work performance as consisting of multiple related yet distinct dimensions. Although performance dimensions can be operationalized in a variety of ways, conceptually, performance dimensions are typically viewed as the variance that is common across all raters’ ratings of a given aspect of performance (Hoffman, et al., 2008; Scullen, et al., 2000). Despite the fre- quent use of MSPRs to diagnose performance deficiencies, little research has focused on the extent to which performance dimension feedback

provides a valid approximation of an individual’s performance (Hoffman, Lance et al., 2008; Scullen, et al., 2000). Instead, it appears that the con- struct validity of MSPR dimensions is simply assumed (Arthur & Villado,

2008). However, research in this area suggests that this assumption may be unfounded. In particular, MSPR internal structure research (e.g., Conway, 1996; Hoffman, Lance, et al., 2008; Lance, Hoffman, et al., 2008; Scullen et al., 2000) consistently indicates performance dimension effects are weak relative to the standards set forth by Campbell and Fiske (1959) and more recently Widaman (1985). Thus, there are legitimate reasons to question the extent to which MSPRs accurately assess the dimensions they purport to assess.

Examining the construct validity of MSPR dimensions using an exter- nal approach to construct validity is consistent with traditional research using nomological network (Cronbach & Meehl, 1955) approaches to construct validation. Using this method, the covariance between MSPR dimensions and conceptually similar externally measured constructs is examined to draw inferences with respect to convergent validity, and the covariance with conceptually dissimilar external constructs is examined to draw inferences with respect to discriminant validity. To the extent that a given MSPR performance dimension (e.g., interpersonal performance) is more strongly related to conceptually similar constructs (e.g., empathy) than to conceptually dissimilar constructs (e.g., intelligence), construct validity evidence is provided for that performance dimension.

*Source Effects*

In the context of classical test theory and psychometric models of per- formance ratings (e.g., Kenny & Berman, 1980; King, et al., 1980; Wherry

& Bartlett, 1982), bias refers to systematic variance that is not a reflection

of the underlying construct (e.g., Lance, Baranik, et al., 2008). By exten- sion, method bias is broadly defined as systematic variance that is intro- duced by the method used to measure a construct (Podsakoff, MacKenzie, Podsakoff, & Lee, 2003). Lance, Hoffman, et al. (2008) observed that previous MSPR internal structure research has generally assumed that measurement method effects (e.g., source effects in this context) repre- sent contamination introduced by the method used to measure a trait and should be reduced as much as possible. Examples of this attribution are (a) Viswesvaran, Schmidt, and Ones’ (2005) suggestion that the “part of the overall impression that is in common with other raters *...* is considered true variance *...* . The part that is unique to that rater—the idiosyncratic part—is halo error” (p. 109), and (b) Wherry and Bartlett’s (1982) spec- ification of overall rater *bias* components in their psychometric theory of rating.

Based primarily on internal structure approach, the results of prior empirical research have also been interpreted as providing evidence that there are no meaningful cross-source differences in performance ratings and that nonoverlapping associated with individual raters should be con- ceptualized as error (Schmidt, Viswesvaran, & Ones, 2000). For example, in a meta-analysis of the moderators of cross-source agreement, Viswes- varan, Schmidt, and Ones (2002) presented evidence that cross-source differences in ratings are actually a function of rating difficulty, as op- posed to rater level effects. Based on these results, Viswesvaran, Schmidt, and Ones (2002) concluded that MSPR source effects are meaningless and in later work (e.g., Viswesvaran, Schmidt, & Ones, 2005) assumed the absence of source relevant variance when modeling the structure of performance ratings. Mount, et al. (1998) and Scullen, et al. (2000) ar- rived at similar conclusions using confirmatory factor analysis. Although Scullen et al., found stronger support for source effects than Mount et al., the importance ascribed to source effects in their study was minimal. Fi- nally, LeBreton, Burgess, Kaiser, Atchley, and James (2003) questioned the meaningfulness of cross-source differences in performance ratings in their examination of the interrater agreement observed in MSPRs.

On the other hand, the use of MSPRs in applied settings is predi-

cated on the assumption that raters from different levels provide unique perspectives on different aspects of ratee performance (Borman, 1997; Lance, Hoffman, et al., 2008; Tornow, 1993). If raters from different lev- els do not provide unique performance-relevant information, it would be difficult to justify the increased cost associated with having multiple raters from different organizational levels provide ratings. In support of these assumptions and in contrast to prior internal structure research, decades of interrater agreement research support higher levels of agreement within, relative to across sources (Conway & Huffcutt, 1997), and more recently, Hoffman, Lance, Bynum, and Gentry (2008) supported the presence and importance of source effects using second-order factor analysis of two large MSPR samples.

Toward this end, a variety of speculations have been forwarded as to the meaning of MSPR source effects. First, in order to investigate the pos-

sibility that source effects are attributable to different conceptualizations of performance by different sources, past research has applied invariance constraints to MSPRs. This research typically supports the invariance of MSPRs to the source providing the ratings, suggesting that source effects are not attributable to different conceptualizations of performance by dif- ferent sources (e.g., Campbell & Lee, 1988; Diefendorff, Silverman, & Greguras, 2005; Facteau & Craig, 2001; Woehr, et al., 2005). The oppor- tunity to observe hypothesis suggests that raters from different organiza- tional levels are privy to distinct aspects of performance (Borman, 1974;

Murphy & Cleveland, 1995). Subsequently, raters from different sources will rely on different types of performance information when making rat- ings. As a result of the reliance on different performance information, raters from different levels will have a low level of agreement yet still provide valid performance relevant information.

Next, when formulating their theory of the ecological nature of percep-

tion, Jones and Thibaut (1958) observed that in the context of evaluations, “the same information may be treated in different ways by the same ob- server” (p. 153) as a result of the influence that perceiver differences in the goals of an interpersonal interaction have on the perception of behavior. In essence, the same incident of an individual’s behavior could be perceived quite differently depending on the observer’s relationship and interaction goals with that individual (Lance, Hoffman, et al., 2008). Accordingly, it may be that although raters from different organizational levels observe very similar aspects of ratee behavior, their interaction goals with the ratee impact their perception and evaluation of the behavior. This process would be expected to result in a skew of rater overall impressions of the target toward the particular competencies valued by a particular rater (Beauvois

& Dubois, 2000; Hooijberg & Choi, 2000; Lance, Hoffman, Gentry, & Baranik, 2008; Lance & Woehr, 1989; Salam, Cox, & Sims, 1997).

Although the root cause of MSPR source effects differs based on the opportunity to observe and ecological validity explanation, both perspec- tives posit that: (a) different sources’ ratings should not be expected to converge entirely, and (b) variance specific to a given source reflects valid, performance-relevant information rather than bias. In a similar vein, Mur- phy and DeShon (2000) argued that collapsing all variance not shared across raters into error (or otherwise performance-irrelevant variance) is untenable and that this practice results in collapsing meaningful variance (information about a person’s performance) into error. Collapsing substan- tively meaningful variance into the error term can result in inappropriate inferences as to the construct validity of performance ratings. Consis- tent with this perspective, psychometric theory, prior performance rating research, and models of performance ratings assume that source effects represent bias, whereas current theory and applied uses of MSPRs posit the importance of source specific variance, and that existing empirical re- search has yielded somewhat inconsistent conclusions. Together, internal structure research has revealed much about the psychometric properties of MSPRs; however, traditional internal structure methods do not allow for firm conclusions as to whether source effects are meaningful at all, much less what these effects mean.

Although examining dimension effects using the nomological network approach is relatively straightforward, the examination of source effects is more ambiguous. With respect to source effects, there is little precedent for

postulating specific empirical relationships among the source effects and externally measured constructs. That is, it would be difficult to rationally compose a nomological network of similar and dissimilar constructs for source effects, in that the existing literature provides little indication of whether source effects are meaningful at all, much less what they mean. However, examining the covariance among source effects and external measures relevant to effective job performance can give an indication of the extent to which MSPR source effects are meaningful. Consistent with the both opportunity to observe and ecological validity explanations, to the degree that source effects reflect substantively meaningful variance, they should correlate with externally measured constructs, to the extent that they represent bias, they should not (Lance, Hoffman, et al., 2008; Woehr, et al., 2005).

In general, three broad patterns of results may emerge with re-

spect to the relationship between source effects and external measures of individual differences. First, MSPR source effects may be unrelated to externally measured constructs. This would provide evidence that source effects do not represent substantively meaningful source-specific, performance-relevant variance and that it is appropriate to interpret vari- ance attributable to the source providing the rating as a form of bias or otherwise performance-irrelevant variance (e.g., Conway, 1996; Viswes- varan, Schmidt, & Ones, 2002; Viswesvaran, Schmidt, & Ones, 2005; Wherry, 1952). Second, MSPR source effects may be differentially re- lated to external measures of individual differences. For example, a latent factor representing peer ratings may be more strongly related to external measures indicative of interpersonal skills than are other source effects (e.g., supervisor and subordinates). This pattern of results would sup- port the argument that interpreting all variance not shared across raters as performance irrelevant is inappropriate and consistent with theoretical explanations for the presence of source effects, that raters from differ- ent levels tap different aspects of performance-relevant variance. Finally, source effects may display equivalent relationships with external measures of individual differences. To illustrate, external measures indicative of in- terpersonal skills may relate similarly to each of the source factors. This would indicate that although source effects are substantively meaningful, they provide the same type of information across sources.

*Externally Measured Constructs*

In this study, we include two methods of measuring external con- structs: a managerial assessment center (AC) and paper-and-pencil mea- sures of personality and cognitive ability. Each of the three measurement methodologies incorporated in this study assess constructs that can be classified on conceptual basis as relevant to managerial performance. AC

research has indicated convergence between AC dimensions and con- ceptually similar constructs assessed using paper and pencil measures of cognitive ability and personality (Meriac, Hoffman, Woehr, & Fleisher,

2008; Shore, Thornton, & Shore, 1990). Similarly, MSPR research has indicated some degree of convergence between MSPRs and conceptually similar personality constructs (Conway, 2000; Conway, et al., 2001). To- gether, this research underscores the assumptions that (a) the constructs assessed using performance ratings, paper-and-pencil measures, and AC dimensions can be categorized on the basis of conceptual similarity, and (b) construct validity evidence can be evaluated by comparing the rela- tionship between conceptually similar and conceptually dissimilar con- structs across the three measurement methodologies incorporated in this study.

In sum, the primary goal of this study was to combine internal and

external approaches to validation in examining the psychometric proper- ties of MSPRs. To investigate the extent to which MSPR source factors represent substantively meaningful variance, or bias, we investigate the degree of convergence among performance-related constructs assessed using three different methodologies—multisource performance ratings, managerial assessment center dimension ratings, and paper-and-pencil measures of personality and intelligence. We also examine the extent to which MSPR dimension factors relate to constructs assessed external to the performance rating instrument.

*Method*

*Participants and Procedure*

Participants were 404 managers enrolled in an Executive Masters of Business Administration (EMBA) program at a large southeastern univer- sity. Although enrolled in the EMBA program, participants concurrently worked as managers in a diverse set of organizations and industries. The participants were largely Caucasian (79%) and male (76%). Participants reported a mean age of 41 years, an average of 11.7 years of managerial experience, and supervised 10 direct reports on average.

Before beginning the EMBA program, the managers completed the

Watson-Glaser Critical Thinking Appraisal (CTA; Watson & Glaser,

1980), the California Psychological Inventory (CPI; Gough & Bradley,

1996), and participated in an AC. For the MSPR instrument, participants were mailed multisource rating forms to be completed by their coworkers prior to beginning the EMBA program. The participants were instructed to distribute the MSPR forms to their coworkers. To ensure anonymity, those completing the surveys were instructed to mail the MSPR forms directly to the university upon completion.

*Measures*

For the AC, two trained assessors provided dimension ratings on an

11-point behaviorally anchored rating scale. In the AC, four exercises including two one-on-one role plays, a leaderless group discussion, and an in-basket were used to evaluate the following behavioral dimensions: analysis, judgment, communication skills, planning and organizing, de- cision making, initiative, influencing others, sensitivity, persuasiveness, team building, and coaching. In the role plays, participants must suc- cessfully resolve a potentially volatile meeting with a direct report, the leaderless group discussion centers around the allocation of a large finan- cial gift, and the in-basket requires participants to assume the role of a high level manager facing a transition. Consistent with recent conceptualiza- tions of AC performance, post consensus dimension ratings were used to operationalize AC dimension performance (Arthur, Day, & Woehr, 2008; Howard, 2008). Post consensus dimension ratings (AKA across exercise dimension ratings) are dimension-level ratings generated on the basis of a consensus discussion at the completion of all exercises and represent a richer, more integrative view of assessee performance than can be gleaned from observing performance in a single exercise (Howard, 2008). Recent research supports the value of operationalizing AC performance in this way (Arthur, Day, McNelly, & Edens, 2003; Meriac, et al., 2008).

Personality constructs were assessed using the 20 folk scales of the

California Psychological Inventory. The CPI manual reports acceptable reliabilities for each of the folk scales and correlations of each among conceptually similar constructs, supporting the construct psychometric soundness of the scales of the CPI. Cognitive ability was assessed using the Watson Glaser Critical Thinking Appraisal. This 80-item instrument is designed to measure critical thinking skills and has frequently been used in research as a measure of general mental ability (GMA).

The MSPR instrument assessed 12 dimensions of managerial perfor-

mance including idealized influence, inspirational motivation, intellectual stimulation, analysis, judgment and decision making, planning and orga- nizing, team building, sensitivity, confrontation effectiveness, communi- cating skills, initiative, and performance management. Supervisor (*N* =

404), subordinate (*N* = 1,255), and peer (*N* = 1,236) responses were used in this study.

*Conceptual Similarity Among Constructs*

Consistent with Murphy’s (1989) suggestions, we followed a deduc- tive approach to specifying an underlying performance model. Consistent with Borman and Brush’s (1993) taxonomy of managerial performance,

TABLE 1

*Classification of Managerial Skills*

|  |  |  |
| --- | --- | --- |
| Technical skills | Interpersonal skills | Leadership skills |
| Judgment & decision makinga | Confrontation effectivenessa | Idealized influencea |
| Decision makingb | Communication skillsc | Inspirational motivationa |
| Judgmentb | Team buildingc | Intellectual stimulationa |
| Analysisc | Sensitivityc | Performance managementa |
| Planning and organizingc |  | Influencing othersb |
|  |  | Persuasivenessb |
|  |  | Coachingb |

a Measured using MSPRs.

b Measured by the AC.

c Measured using both AC and MSPRs.

we conceptualized managerial performance as consisting of three broad di- mensions: technical activities/mechanics of management, communication and interpersonal facilitation, and leadership and supervision (the fourth dimension, “useful personal behaviors and skills” essentially represents an “other” category and was not incorporated in this study). Borman and Brush’s taxonomy includes a variety of subordinate dimensions and cat- egorizes them with respect to the three broad dimensions they represent. Recent meta-analyses (Conway, 1999) and large-scale primary studies (Hoffman, Lance, et al., 2008) have empirically supported this taxonomy. The 12 subscales measured by the MSPR instrument and the 11 AC di- mensions were conceptually sorted by four subject matter experts (SMEs) into one of the three broad performance factors specified by Borman and Brush.

In addition, intelligence and each of the 20 CPI scales were classified

as conceptually similar/dissimilar to each of the three broad dimensions of managerial performance described by Borman and Brush. This de- ductive approach of conceptual classification and empirical evaluation is consistent with previous job performance research that has specified an a priori performance model based on SME classification to evaluate the construct-related validity of managerial skills (e.g., Hoffman, Lance, et al., 2008; Mount, et al., 1998; Scullen, et al., 2000; Shore, et al., 1990). Of the managerial skill dimensions, initiative was the only dimension that was not categorized into the same broad performance dimension by all four SMEs. Table 1 presents the results of the SME classification of the managerial constructs assessed by the AC and MSPRs. Although the four SMEs were experts in managerial assessment, only three of the four had expertise with the CPI and CTA. Accordingly, only these three SMEs were asked to categorize the paper-and-pencil assessed constructs on the basis

of conceptual similarity. Of the 21 constructs measured with the paper and pencil instruments, 5 were deemed to be conceptually similar to one or more of the three managerial performance dimensions by all three SMEs. Specifically, tolerance, empathy, and sociability were expected to be cor- relates of interpersonal skills, dominance and sociability were expected to be correlates of leadership, and GMA was expected to be a correlate of technical administrative skills.

*Dimensionality of Managerial Performance*

Although we followed Murphy’s (1989) recommendation to rely on a deductive approach to the specification of the performance domain, it is also important to empirically examine the appropriateness of the a priori structure relative to other possible performance structures. Accordingly, in addition to comparing models with different trait X method parame- terizations, we also tested a variety of competing performance models to ensure the adequacy of the Borman and Brush taxonomy for the dimension structure of both the MSPRs and the AC. Informed by Guilford’s (1954) general performance factor, Kenny and Berman’s (1980) true correlation, Cooper’s (1981) true halo, and Viswesvaran, Schmidt, and Ones’s (2005) actual correlation, we specified a general factor where all raters’ ratings of all dimensions are parameterized as loading on a single general per- formance factor (MSPR Model 4; AC Model 1). Next, prior research has consistently supported a two-factor structure of performance that roughly corresponds to task and interpersonal performance dimensions in both managerial (Blake & Mouton, 1964; Fleishman, 1957; Judge, Piccolo,

& Ilies, 2004; Katz & Kahn, 1951; Stogdill, 1963) and nonmanagerial

(Hoffman, et al., 2007; Kram, 1985; Noe, 1988; Smith, Organ, & Near,

1983) contexts. This model was parameterized by setting the leadership and interpersonal latent factor correlation to unity, resulting in an inter- personal skills factor and a technical activities/mechanics of management factor (MSPR Model 5; AC Model 2). Next, based on the suggestion of an anonymous reviewer, we specified a model that matched our SME based three-factor specification, except team building loaded on the leadership as opposed to the interpersonal factor (MSPR Model 7; AC Model 4). Third, we specified a performance model consisting of administrative, technical, and interpersonal skills based on early (Katz, 1974; Mann, 1965) and more recent (Mount, et al., 1998; Scullen, et al., 2000) conceptualiza- tions of managerial performance. This model collapsed interpersonal and leadership into an “interpersonal” factor and modeled a technical factor that was the same as the a priori conceptualization, except planning and organizing was removed from the technical factor and set to load on an administrative factor (MSPR Model 8; AC Model 5).

*Results*

*Preliminary Statistical Analyses*

Before examining the substantive research questions, we exam- ined the factor structure of the MSPR instrument and the AC ratings. The multitrait–multimethod variance–covariance matrix among subscales measured by each source providing ratings on the MSPR instrument served as the input into LISREL version 8.5 (Joreskog & Sorbom, 1996). Al- though prior research supports an MSPR model consisting of source and dimension factors, multiple models potentially characterizing the MSPR data were evaluated to ensure that the source and dimension model pro- vided the most appropriate representation of the data in this study. The classification agreed upon by the SMEs was used to determine which MSPR subscales were specified to load on each of the three performance factors. An evaluation of relevant model fit indices provided the basis for model evaluation.

The results of the set of confirmatory factor analyses examining the structure of the MSPRs are presented in Table 2. The first model specified a general performance factor across sources (one factor model) and no source factors. This model specified a path loading for the three sources’ ratings on all subscales on a single performance factor, did not specify source factors, and did not provide an adequate fit with the data (Model

1; *χ* 2 = 15,484.99; RMSEA = .27; SRMSR = .18; NNFI = .76; CFI =

.77). Second, a source-only model with three latent factors representing each of the three sources was examined (supervisor, peer, and subordinate latent factors). In essence, this model is interpreted as a general factor for each source, with no factors corresponding to variance shared across sources (e.g., no “trait” factors in MTMM parlance). The three-source factor model did not fit the data well (Model 2; *χ* 2 = 3,923.63; RMSEA =

.13; SRMSR = .06; NNFI = .91; CFI = .92). The third model specified

three dimension latent factors consistent with the a priori dimension struc- ture and no source factors. The dimension-only model also demonstrated inadequate fit with the data (Model 3; *χ* 2 = 15,494.18; RMSEA = .28; SRMSR = .19; NNFI = .76; CFI = .78).

Finally, a six-factor correlated trait-correlated method model com- posed of three source (supervisor, peer, and subordinate) and three dimen-

sion (technical/administrative, leadership, and interpersonal skills) latent factors was examined (MSPR Model 3), with each of the MSPR sub- scales set to load on a respective broad performance dimension. And, all of the subscales rated by a single source load on a source factor, resulting in three latent source factors. Consistent with convention when model- ing multitrait–multimethod data, the correlated trait-correlated method

TABLE 2

*Model Fit Statistics for Multisource Performance Ratings and Assessment*

*Center Dimensions*

|  |  |  |  |
| --- | --- | --- | --- |
| *df χ* 2 RMSEA | SRMSR | TLI | CFI |
| Models of MSPRs |  |  |  |  |  |  |
| 1a. One general | 495 | 15,484.99 | .27 | .18 | .76 | .77 |
| 1b. Three sources | 492 | 3,923.63 | .13 | .06 | .91 | .92 |
| 1c. Three dimensions | 492 | 15,494.18 | .28 | .19 | .76 | .78 |
| 1d. Three sources + one dimension | 459 | 2,412.71 | .10 | .05 | .94 | .95 |
| 2. Three sources + two dimensions | 458 | 2,070.12 | .09 | .08 | .94 | .95 |
| 3. Three sources + Three dimensions | 456 | 1,738.17 | .08 | .06 | .95 | .96 |
| 4. Three sources + Three dimensions | 456 | 1,894.21 | .09 | .09 | .95 | .95 |
| 5. Three sources + Three dimensions | 456 | 1,983.94 | .09 | .07 | .95 | .95 |
| Models of AC dimensions |  |  |  |  |  |  |
| 1. One dimension | 35 | 301.58 | .14 | .08 | .85 | .88 |
| 2. Two dimensions | 34 | 61.80 | .05 | .04 | .98 | .98 |
| 3. Three dimensions | 32 | 55.89 | .04 | .04 | .98 | .99 |
| 4. Three dimensions | 32 | 60.47 | .05 | .04 | .98 | .98 |
| 5. Three dimensions | 33 | 60.56 | .05 | .04 | .98 | .98 |
| Full model | 970 | 2,693.82 | .07 | .05 | .91 | .92 |

*Note*. For the AC and MSPR models that varied the dimension specifications, Model

2 = interpersonal and technical factors; Model 3 = technical, interpersonal, and leadership

factors; Model 4 = technical, interpersonal, and leadership factors, with team-building set

to load on leadership; Model 5 = interpersonal, technical, and administrative factors.

approach was used where correlations between dimensions and sources were set to zero (Lance, Noble, & Scullen, 2002). Given prior MSPR re- search supporting a model composed of source and dimensions factors, the six-factor model was expected to provide the closest fit to the data. Inspec- tion of Table 2 indicates that this model provides an acceptable fit to the data and the best fit of the five models tested (*χ* 2 = 1,738.17; RMSEA =

.08; SRMSR = .06; NNFI = .95; CFI = .96). To more closely examine

the appropriateness of our a priori performance model, we compared this model to a variety of models that varied the underlying dimension struc- ture. As presented in Table 2, the a priori dimension structure provided a closer fit to the MSPR data relative to the other dimension structures spec- ified, providing empirical support for the SMEs’ conceptual classification of the MSPR subscales based on Borman and Brush’s taxonomy.

Next, we examined the factor structure of the AC ratings (Table 2). Here again, the SMEs’ classification was used to specify which of the pat- tern of loadings of AC dimension ratings onto the broad latent dimension factors. The proposed three-factor model of managerial skills was tested with each of the post consensus dimension ratings set to load on one of the three broad dimensions of managerial skills as classified by the SMEs.

Results of the CFA indicate that this model fit the AC data well (Model 3; *χ* 2 = 55.89; RMSEA = .04; SRMSR = .04; NNFI = .98; CFI = .99). In addition, consistent with the results from the CFA of MSPRs, the a priori dimension structure fit the data more closely than did any of the alternate AC dimension structures tested.

To examine the nomological network of the MSPR dimension and source effects, the AC dimensions, CPI folk scales, and CTA were simul- taneously entered into LISREL 8.5. For the paper-and-pencil instruments, each of the constructs judged to be conceptually similar to at least one of the three performance domains served as a single manifest indicator of a latent factor in this model with the factor loading set to the square root of the reliability of each of the relevant scales. The full model (a model composed of latent factors representing three MSPR source latent factors, three MSPR performance dimension latent factors, three AC di- mension latent factors, and the five constructs measured by the paper and pencil instruments) provided an adequate fit to the data (*χ* 2 = 2,693.82; RMSEA = .07; SRMSR = .05; NNFI = .91; CFI = .92).

*Construct-Related Validity of MSPR Dimensions*

Table 3 presents the latent factor correlations between MSPR dimen- sion factors and external constructs aligned with respect to conceptual similarity. In order to determine whether the correlations between MSPR dimensions and similar constructs differed from the mean correlation be- tween MSPR dimensions and dissimilar constructs, we used the approach recommended by Meng, Rosenthall, and Rubin (1992) for constructing confidence intervals about composite correlations among dependent cor- relations. Results indicate that the MSPR technical/administrative factor was not related to either of the conceptually similar externally measured constructs. However, the MSPR interpersonal factor was more strongly related to both AC interpersonal performance (difference in *r* between similar and dissimilar constructs = .17) and empathy (difference in *r* with dissimilar = .09) than to dissimilar constructs. On the other hand, the relationship between MSPR interpersonal performance and tolerance and sociability did not differ from the relation of MSPR interpersonal and dissimilar external constructs (difference in *r* with dissimilar = −.02 for both). Finally, the MSPR leadership factor was more strongly related to AC leadership (difference in *r* with dissimilar = .10), dominance (difference in *r* with dissimilar = .08), and sociability (difference in *r* with dissimi- lar = .09) than to the dissimilar externally measured constructs. Overall, however, the mean correlation between MSPR dimensions and similar constructs (mean *r* = .09) was weak based on the guidelines forwarded by Cohen (1988) and is indicative of somewhat weak convergent validity.

TABLE 3

*Difference in Relationship Among MSPR Dimensions and Similar Versus*

*Dissimilar Constructs*

|  |  |  |
| --- | --- | --- |
|  | Mean *r* with dissimilar and |  |
| Contrastsa | *r* | (difference in *r*) | *Z* | 95% CI*Z* |
| MSPR technical |  | −*.*06b |  |  |
| AC technical | −*.*06 | (*.*00) | *.*04 | −.512 – .590 |
| Intelligence | *.*10 | (*.*16) | 3*.*54 | 3.10 – 4.21 |
| MSPR interpersonal |  | *.*04b |  |  |
| AC interpersonal | *.*21 | (*.*17) | 3*.*56 | 3.12 – 3.95 |
| Sociability | *.*02 | (−*.*02) | −*.*31 | −.700 – .080 |
| Empathy | *.*13 | (*.*09) | 1*.*91 | 1.52 – 2.30 |
| ToleranceMSPR leadership | *.*02 | (−*.*02)*.*06b | −*.*31 | −.700 – .080 |
| AC leadership | *.*16 | (*.*10) | 2*.*04 | 1.60 – 2.49 |
| Dominance | *.*14 | (*.*08) | 1*.*59 | 1.15 – 2.04 |
| Sociability | *.*15 | (*.*09) | 1*.*81 | 1.37 – 2.26 |

a We compared each correlation between similar constructs with the mean of the corre- lations between dissimilar constructs using the formulae presented by Meng, Rosenthall, and Rubin (1992).

b Mean correlation between the MSPR dimension and dissimilar external constructs. Value in parentheses is the difference in correlation between the similar construct listed and the mean of the correlation between that MSPR dimension and dissimilar constructs.

Although the correlations between MSPR dimensions and dissimilar con- structs were weak as well (mean *r* = .01), because the mean correlation with similar constructs was weak as well, discriminant validity evidence becomes difficult to interpret. Together, these findings provide modest convergent and discriminant validity evidence for the MSPR dimension factors.

*The Meaning of MSPR Source Factors*

The pattern of latent factor correlations between external constructs and each of the three latent source factors were used to draw inferences regarding the extent to which source effects represent substantively mean- ingful source-specific, performance-relevant variance or bias (Table 4). The manager latent source factor was weak-moderately related to one of the externally measured constructs indicative of technical skills (for AC technical skills, *r* = .19), weakly related with two of the constructs indicative of interpersonal skills (for AC interpersonal skills, *r* = .13; for CPI sociability, *r* = −.13), and weakly related to one of the constructs indicative of leadership skills (for AC leadership skills, *r* = .13). The peer latent source factor was a weak correlate of one of the constructs

751

HOFFMAN AND WOEHR

TABLE 4

*Correlations Among Latent Factors*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| 1 | MSPR leadership | 1*.*0 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 2 | MSPR interpersonal | *.*58 | 1*.*0 |  |  |  |  |  |  |  |  |  |  |  |  |
| 3 | MSPR technical | *.*10 | *.*41 | 1*.*0 |  |  |  |  |  |  |  |  |  |  |  |
| 4 | Manager | – | – | – | 1*.*0 |  |  |  |  |  |  |  |  |  |  |
| 5 | Peer | – | – | – | *.*38 | 1*.*0 |  |  |  |  |  |  |  |  |  |
| 6 | Subordinate | – | – | – | *.*26 | *.*37 | 1*.*0 |  |  |  |  |  |  |  |  |
| 7 | AC technical | *.*10 | *.*04 | −*.*06 | *.*19 | *.*13 | *.*17 | 1*.*0 |  |  |  |  |  |  |  |
| 8 | AC interpersonal | *.*20 | *.*21 | −*.*07 | *.*13 | *.*23 | *.*06 | *.*53 | 1*.*0 |  |  |  |  |  |  |
| 9 | AC leadership | *.*16 | *.*06 | −*.*09 | *.*13 | *.*16 | *.*29 | *.*62 | *.*85 | 1*.*0 |  |  |  |  |  |
| 10 | Dominance | *.*14 | *.*11 | −*.*11 | *.*02 | −*.*02 | *.*08 | *.*13 | *.*21 | *.*34 | 1*.*0 |  |  |  |  |
| 11 | Sociability | *.*15 | *.*02 | −*.*04 | −*.*13 | −*.*06 | *.*02 | *.*06 | *.*17 | *.*18 | *.*64 | 1*.*0 |  |  |  |
| 12 | Empathy | *.*12 | *.*13 | −*.*08 | −*.*05 | *.*05 | *.*05 | *.*15 | *.*29 | *.*16 | *.*42 | *.*67 | 1*.*0 |  |  |
| 13 | Tolerance | −*.*08 | *.*02 | *.*02 | *.*07 | −*.*02 | *.*01 | *.*10 | *.*24 | *.*15 | *.*19 | *.*24 | *.*45 | 1*.*0 |  |
| 14 | Intelligence | −*.*02 | −*.*07 | *.*10 | *.*05 | *.*02 | *.*04 | *.*35 | *.*18 | *.*20 | *.*11 | *.*10 | *.*21 | *.*35 | 1*.*0 |

*Note*. 95% confidence interval = *r* ± .10.

indicative of technical skills (for AC technical/administrative skills, *r* =

.13), a weak-moderate correlate of one construct indicative of interper- sonal skills (for AC interpersonal skills, *r* = .23), and a weak correlate of one construct indicative of leadership skills (for AC leadership skills, *r* = .16). Finally, the subordinate latent source factor was a weak corre- late of one of the constructs indicative of technical/administrative skills (for AC technical/administrative skills, *r* = .17), unrelated to all of the constructs indicative of interpersonal skills, and a moderate correlate of one of the constructs indicative of leadership skills (for AC leadership skills, *r* = .29). These correlations between MSPR source factors and externally measured constructs suggest that MSPR source factors reflect substantively meaningful variance as opposed to performance irrelevant bias. However, these results do not provide an indication of the extent to which the performance relevant variance represented in source factors is *source specific*.

To determine whether the substantively meaningful variance repre- sented in the source factors (e.g., the variance the source factors share

with the external constructs) is source specific, the next set of analyses examined the extent to which a given externally measured construct was more strongly related to one source factor than the other two latent source factors. For these analyses, only externally measured constructs that had confidence intervals that did not overlap with zero were included. If a construct (e.g., dominance) is not correlated with any of the three latent source factors, it is of little value to ask whether the correlation between that construct and the source factors differs significantly. The four exter- nally measured constructs that had nonzero relationships with the latent source factors were AC technical/administrative skills, AC interpersonal skills, AC leadership skills, and sociability. To determine whether the cor- relations among each externally measured construct and source factors differed from one another, we built confidence intervals (CI) around the difference in correlations using the approach recommended by Meng, et al. (1992) for comparing dependent correlations (Table 5).

Results indicated that the correlation between AC techni- cal/administrative skills and the manager and peer (difference in *r* =

.06), manager and subordinate (difference in *r* = .02), and peer and sub- ordinate (difference in *r* = .04) source factors did not differ. In contrast, the AC interpersonal skills factor was more strongly related to the peer source factor than to the subordinate (difference in *r* = .17) source factor. However, the correlation between AC interpersonal skills and manager and subordinate source factors (difference in *r* = .07) and the correla- tion between AC interpersonal skills and the peer and boss source factors (difference in *r* = .10) did not differ. In sum, the peer source factor was more strongly related to AC interpersonal skills than was the subordinate

TABLE 5

*Differences in Correlations With Source Factors*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| External |  |  | Manager | Manager vs. | Peer vs. |
| construct | Boss | Peer | Subordinate vs. peer | subordinate | subordinate |
| AC technical | .19 | .13 | .17 .06a | .02 | .04 |

(−.05–.17)b (−.10–.14) (−.07–.15)

AC interpersonal .13 .23 .06 .10 .07 .17 (−01–.21) (−.04–.18) (.06 –.28)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| AC leadership | .13 | .16 | .29 | .06 | .16 | .13 |
|  |  |  |  | (−.05–.17) | (.05–.27) | (.03–.23) |
| Sociability | −.13 | −.06 | .02 | .07(−.04–.18) | .15(.03 –.27) | .08(−.03–.19) |

a Difference in correlation between the two sources listed.

b We compared the intercorrelation with source factors using Meng et al.’s formula for

comparing dependent correlations.

*Note.* Values in parentheses are confidence internal about the difference in correlations

between those sources and the external correlate.

source factor. The AC leadership skills factor was more strongly related to the subordinate source factor than to the peer (difference in *r* = .13) or manager (difference in *r* = .16) source factors. However, the correla- tions between the AC leadership skills and peer and manager latent source factors did not differ significantly (difference in *r* = .06). These results suggest that the subordinate source factor represents leadership skills to a greater extent than do peer and supervisor source factors.

Finally, sociability was more strongly related to the manager (*r* =

−.13) source factor than to the subordinate source factor (difference in *r* = .14). However, the correlation between sociability and the manager and peer source factors (difference in *r* = .07), and the correlation be- tween sociability and the peer and subordinate (difference in *r* = .08) source factors did not differ. Overall, these findings suggest that all three source factors tap AC technical skills to a similar extent, the peer source factor reflects rate interpersonal skills to a greater extent than does the sub- ordinate source factor, and the subordinate source factor is indicative of leadership skills to a greater extent than either the supervisor or peers source factors. Thus, not only do MSPR source factors reflect meaningful variance, the meaning of MSPR source factors depends upon the source providing the ratings.

*Discussion*

Despite increased focus on the psychometric properties of MSPRs, there is no consensus with respect to the meaning of source effects and the

construct validity of performance dimensions. This study contributes to the literature by supplementing the traditional internal structure approach to examining the construct validity of MSPRs (Campbell & Fiske, 1959) with a nomological network approach (Cronbach & Meehl, 1955) by ex- amining the pattern of relationships between MSPR source and dimension factors and external measures of individual differences. Of central interest were the results demonstrating that MSPR source factors differentially correlated with externally measured constructs, providing evidence that MSPR source factors represent substantively meaningful source specific variance, as opposed to method bias. In addition, based on the corre- lations between MSPR performance dimension factors and similar and dissimilar external constructs, somewhat weak support was provided for the construct-related validity of MSPR dimensions. These results are dis- cussed in the context of two overarching questions: (a) What do MSPR source factors represent? and (b) What is the construct validity evidence for MSPR dimensions?

*The Meaning of MSPR Source Effects*

By integrating internal and external methods of construct validation to examine the meaning of MSPR source effects, this study arrived at markedly different conclusions than research relying strictly on inter- nal structure and interrater agreement based approaches (e.g., Conway,

1996; Conway & Huffcutt, 1997; LeBreton, et al., 2003; Mount, et al.,

1998; Scullen, et al., 2000; Viswesvaran, Schmidt, & Ones, 2002). Re- sults revealed that the three MSPR source factors were weak-moderately related to externally measured constructs and that externally measured constructs differentially correlated with the peer and subordinate source factors. Together, these results suggest that (a) MSPR source factors repre- sent substantively meaningful variance, and (b) the meaning of the MSPR source factor differs depending on the source providing the ratings. An inspection of the differential correlations among source factors and ex- ternal constructs reveals a few interesting patterns. First, source factors typically did not correlate substantially with the paper-and-pencil assessed constructs (mean *r* = .01). However, AC dimensions were consistently related to source factors (mean *r* = .17). These findings suggest that each source’s overall impression is formed based on a combination of their coworker’s technical/administrative skills, interpersonal skills, and leadership skills.

The pattern of relationships between the three source factors and exter-

nally measured constructs facilitates inferences with respect to the mean- ing of these effects. First, the AC interpersonal skills factor was more strongly related to the peer than the subordinate source factor. Thus, peer

ratings are contingent on an individual’s interpersonal skills to a greater extent than are the general impressions of subordinates. Next, the AC lead- ership skills factor was more strongly related to the subordinate source factor than to either the peer or supervisor source factors; thus, subordi- nate general impressions tend to be based on the ratee’s leadership skills to a greater extent than do peer and supervisor general impressions. Fi- nally, the AC technical/administrative performance factor was related to each of the three latent source factors; however, none of these correlations differed from one source to another, suggesting that a manager’s tech- nical/administrative performance is of equal importance to raters from different organizational levels.

Although our findings clearly indicate that source effects reflect sub- stantively meaningful variance, our results are consistent with multiple hypotheses regarding the substantive meaning of source effects and con- sequently do not shed light on *why* source effects are present in ratings. Consistent with our results, subordinates likely have a greater opportu- nity to observe leadership skills than do peers or supervisors. Similarly, peer relationships are typically characterized by high levels of interdepen- dence and large amounts of time spent together relative to other working relationships, affording peers substantial opportunities to observe their coworkers’ interpersonal skills (Landsberger, 1961). Thus, our finding that peer ratings reflected interpersonal skills to a greater extent than did other source factors is also consistent with the opportunity to observe hypothesis.

On the other hand, our results are also consistent with the ecological

validity hypotheses. From this perspective, to the extent that raters from different sources have different interaction goals with the ratee, they may be expected to perceive the same behavior differently (Jones & Thibaut,

1958). Hooijberg and Choi (2000) demonstrated that in the eyes of one’s peers, the most important criterion for effectiveness is the extent to which a manager facilitates group processes. Further, their results indicated that subordinates valued leadership roles more so than any other managerial role when evaluating a manager’s effectiveness. It is noteworthy that the competencies that are most valued by peers and subordinates in Hooi- jberg and Choi were also the competencies most relied upon by these raters when making ratings in this study. Thus, the results of our study are consistent with the ecological perspective to perception as it is rea- sonable to suspect that followers’ interaction goals with their boss would center on leadership, and peers would greatly value getting along with coworkers. Although we did not directly test the opportunity to observe or ecological perspectives, attempting to tease apart these competing hy- potheses for *why* source effects are present represents an interesting area for future research. In any event, what seems clear based on our results

is that source effects reflect meaningful variance and warrant additional investigation.

*Construct Validity of MSPR Dimensions*

The results provide modest support for the construct validity of MSPR dimensions. That is, the results of the CFA clearly point to the presence and importance of modeling dimension effects in MSPRs. In addition, there was some degree of convergence between MSPR dimensions and similar external constructs, and MSPR dimensions were typically unre- lated to dissimilar external constructs. However, the correlations between MSPR dimensions and similar external constructs were sufficiently weak to warrant concern with respect to their construct validity. Together, al- though the results of the internal structure research point to the presence of dimensions, the nomological network results were less positive, yielding modest support for the validity of MSPR dimensions.

It is noteworthy that in the typical MSPR application, feedback re- cipients are presented with specific “behaviors” (items) separated by the source providing the ratings, as opposed to more general aggregates of items into composite dimensions (Heslin & Latham, 2004; Kaiser & Craig,

2005; London & Smither, 1995). The typical practice of presenting feed- back at the item-level does not preclude the need for confidence in the

accuracy of the feedback being presented. Indeed, the aggregation of individual items into performance dimensions should reduce the error as- sociated with the measurement (increase reliability) and afford a greater opportunity for validity, compared to examining the validity of individ- ual items (Nunally & Bernstein, 1994). After all, if the more reliable performance dimension summaries are not valid representations of ratee performance, one would hardly expect less reliable individual items to be. Thus, despite the reliance on item-level feedback in applied settings, an understanding of the construct validity of the performance dimensions being rated is critical to the appropriate use of MSPRs.

Clearly, the modest construct validity evidence demonstrated for the

MSPR dimensions is somewhat troubling for organizational scientists interested in using MSPR in developmental settings. Specifically, these results suggest that giving performance feedback to managers on the basis of the dimensions assessed using this tool may at best be unhelpful and at worst counterproductive to managerial development. MSPRs have become a staple of management development as the process provides the foundation on which subsequent development plans and skill building effort is built. Clearly, a closer look at the validity of MSPR *dimensions* is a critical, if often ignored, area for research.

Toward this end, this study incorporated a construct-oriented approach (Arthur & Villado, 2008) to validation of constructs assessed in the man- agerial performance domain by examining the validity of our measure using multiple bases of evidence, including (a) content validity on the ba- sis of the SME classification, (b) structural validity evidence provided by the CFA-based dimension structure comparisons, (c) the convergent and discriminant validity evidence provided by the correlations with external constructs, and (d) the consistency in the dimension structure across two divergent methodologies. Together, the reliance on a construct-oriented approach to validation using multiple different evidentiary bases strength- ens our conclusions with respect to the structure of managerial perfor- mance. Nevertheless, the same dimension structure supported here will not necessarily generalize to other contexts using different measures of managerial performance, as evidenced by support for a wide variety of managerial performance taxonomies in prior research (e.g., Mount, et al.,

1998; Scullen, et al., 2000). We echo recent recommendations for man- agerial performance researchers to subject their measures to a rigorous

psychometric evaluation focusing on the underlying constructs prior to drawing substantive conclusions (Arthur, et al., 2008; Arthur & Villado,

2008; Hoffman, et al., 2008; Meriac, et al., 2008). Alternately, by demon-

strating that source effects contain substantively meaningful portions of variance, our findings contrast recent arguments that “method” of mea- surement effects (e.g., source effects, in this context) are not appropriate foci for organization researchers (Arthur & Villado, 2008).

*Limitations and Avenues for Future Research*

Despite these contributions to the MSPR literature, this study is not without limitations. First, as with many performance appraisal instru- ments, the MSPR instrument used in this study is a “one of a kind” instrument specific to the leadership development program from which the sample was drawn. Thus, the generalizability of the results is a cause for concern. However, given the multiple bases of validity evidence pre- sented, the support of an a priori model that is consistent with theoretical performance models (Borman & Brush, 1993), and the consistency of the internal structure results with prior internal structure research, some con- fidence is lent to the generalizability of our results. Still, the replication of this study with other MSPR instruments is clearly warranted.

Next, the impact of individual raters was not explicitly modeled in the above structural models. However, the same supervisor, peer, or sub-

ordinate never provided ratings for more than one target in this sample. Thus, the potentially confounding effect of rater level and specific indi- vidual rater was eliminated (Woehr, et al., 2005). Essentially, the source

effect components represent effects associated with different supervisors, peers, or subordinates. Thus, common within-source variance cannot be attributed to a common rater providing ratings. Consequently, confound- ing source by specific rater was not an issue in this study. In addition, the findings that rater level effects differentially correlated with external mea- sures in meaningful ways supports the meaningfulness of our approach to modeling the structure of MSPRs.

Despite the interesting results afforded by examining the covariance

among individual difference constructs and MSPRs, this approach is not without limitations. In particular, this approach closely mirrors methods used in traditional criterion-related validity settings, where one is val- idating measures of individual difference constructs. Accordingly, it is reasonable to question the extent to which our approach affords infer- ences with respect to the construct-related validity of MSPRs, as opposed to the criterion-related validity of the external individual difference con- structs. One potential alternative to the individual difference approach taken here would be an examination of the relationship between source and dimension factors and objective performance indices. However, the construct-oriented approach to validation (Arthur & Villado, 2008) we used facilitates important inferences that could not be made by relying strictly on objective performance measures. For example, it is quite diffi- cult to devise objective performance indices for managerial work, much less multiple objective measures that would be expected to differentially correlate with different sources’ ratings (Austin & Villanova, 1992). Fi- nally, although valuable, a construct validation approach using objective performance indices would give little indication of *what* MSPR source effects are actually measuring and the degree to which MSPR dimen- sions reflect the competencies they purport to assess. Thus, in accordance with recent thought (Landy, 1986; Messick, 1995) and policy suggestions (AERA, APA, & NCME 1999) emphasizing a Unitarian approach to val- idation, the approach to construct validation used here provides a useful step in facilitating inferences of the appropriate interpretation of MSPRs.

At the same time, the degree to which the external measures that we used to validate MSPRs are construct valid is an area of concern. Given

that AC dimension ratings were used as a criterion with which to compare the construct validity of MSPRs, it is important that the AC dimensions provide a valid approximation of the competencies they purport to assess. Extensive research has documented issues with the construct validity of AC ratings when using traditional MTMM designs (Bowler & Woehr,

2006; Lance, 2008; Lievens & Conway, 2001). On the other hand, recent

research has pointed to flaws in historical (Howard, 2008; Rupp, Thornton,

& Gibbons, 2008), theoretical (Arthur, Day, & Woehr, 2008), and empiri- cal (Lance, Woehr, & Meade, 2007) foundation on which calls to abandon

dimensions are based. In fact, some have recommended relying solely on across-exercise dimension ratings (as we did in this study) as opposed to the post exercise dimension ratings typical to traditional AC internal structure research (Arthur, et al., 2008). In addition, empirical evidence substantiates the value of operationalizing AC performance using post- consensus dimension ratings (Arthur, et al., 2003; Meriac, et al., 2008; Shore, et al., 1990). Further, supplemental analyses revealed that (a) AC dimensions related twice as strong to conceptually similar (mean *r* = .22) than to dissimilar (mean *r* = .11) external constructs, and (b) AC dimen- sions differentially correlated with MSPR source factors, supporting the convergent validity of the AC dimensions. Accordingly, based on current thinking in the AC literature as well as the results of this study, the use of AC dimensions to validate MSPRs is not a serious limitation.

Interestingly, the AC skill dimensions converged with similar MSPR

ratings to a greater extent than did paper-and-pencil constructs. Specif- ically, the mean correlation between AC dimensions and MSPR ratings was larger than that between MSPR ratings and paper-and-pencil assessed constructs. The degree of psychological and physical fidelity (Ployhart, Schneider, & Schmitt, 2006) between the AC and MSPR may account for this difference. In other words, the AC assessed functionally similar constructs to the MSPRs (high psychological fidelity) in a context closely aligned to the performance context (high physical fidelity), whereas the paper-and-pencil measures assessed different constructs judged to be re- lated to managerial skills using a method with a relatively low level of fidelity. Together, it is not surprising that the display of a behavior in a simulation of the work environment such as an AC (e.g., a sample of behavior) is more closely linked to subsequent work behaviors than is a general tendency to behave in a certain way, such as paper-and-pencil assessed constructs (e.g., a sign; Wernimont & Campbell, 1968).

Finally, although we found evidence for meaningful differences in relations between MSPR ratings and external constructs, the magnitude

of these differences was typically not particularly large. For instance, the largest correlation between MSPR ratings and the external constructs was .29, and the largest difference in correlations between similar and dissimilar constructs with MSPR dimensions was .16. Although the cor- relations between individual differences and performance ratings reported here were modest, the magnitudes of individual difference–performance associations found in this study are consistent with prior research (Arthur, et al., 2003; Barrick & Mount, 1991; Meriac, et al., 2008; Schmidt & Hunter, 1998). In addition, the magnitude of difference in relationships reported here are in line with other established findings in the organi- zational literature such as differential correlations between attitudes and task performance versus citizenship behaviors (Hoffman, et al., 2007) and

differential correlations between narrow personality constructs and similar bandwidth behavioral dimensions (Hogan & Holland, 2003).

*Summary and Conclusions*

This study represents a first step at disentangling the meaning of MSPR source and dimension effects by providing an empirical exami- nation of the construct validity of empirically supported MSPR variance components. Based on our results and consistent with practitioners’ uses of MSPRs, source effects do indeed appear to reflect performance rele- vant variance. Further, our study contributes to the literature by painting an initial picture of what these source effects mean. On the other hand, the validity evidence for MSPR dimensions was modest, at best. Ad- ditional research investigating the psychometric properties and theoreti- cal underpinnings of empirically supported performance rating factors is needed.

REFERENCES

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

Arthur W, Day EA, Woehr DJ. (2008). Mend it, don’t end it: An alternate view of as- sessment center construct-related validity evidence. *Industrial and Organizational Psychology: Perspectives on Research and Practice, 1*, 105–111.

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

125–154.

Arthur W, 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.

Atkins PWB, Wood RE. (2002). Self- versus others’ ratings as predictors of assessment center ratings: Validation evidence for 360-degree feedback programs. PERSONNEL PSYCHOLOGY, *55*, 871–904.

Austin JT, Villanova P. (1992). The criterion problem: 1917–1992. *Journal of Applied*

*Psychology, 77*, 836–874.

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

Beauvois JL, Dubois N. (2000). Affordances in social judgment: Experimental proof of why it is a mistake to ignore how others behave toward the target and look solely at how the target behaves. *Swiss Journal of Psychology, 59*, 16–33.

Beehr TA, Ivanitskaya L, Hansen CP, Erofeev D, Gudanowski DM. (2001). Evaluation of

360 degree feedback evaluations: Relationships with each other and with perfor- mance and selection predictors. *Journal of Organizational Behavior, 22*, 775–788.

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

Borman WC. (1974). The rating of individuals in organizations: An alternative approach.

*Organizational Behavior and Human Performance, 12*, 105–124.

Borman WC. (1997). 360 ratings: An analysis of assumptions and research agenda for evaluating their validity. *Human Resource Management Review, 7*, 299–

315.

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

Bowler MC, 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.

Campbell DT, Fiske DW. (1959). Convergent and discriminant validation by the multitrait- multimethod matrix. *Psychological Bulletin, 56*, 81–105.

Campbell DJ, Lee C. (1988). Self-appraisal in performance evaluation: Development versus evaluation. *Academy of Management Review, 13*, 302–314.

Church AH, Allen DW. (1997). Advancing the state of the art of 360-degree feedback.

*Group and Organization Management, 22*, 149–161.

Cohen J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erbaum.

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.

Conway JM. (2000). Managerial performance development constructs and personality correlates. *Human Performance, 13*, 23–46.

Conway JM, Huffcutt AI. (1997). Psychometric properties of multisource performance ratings: A meta-analysis of subordinate, supervisor, peer, and self-ratings. *Human Performance, 10*, 331–360.

Conway JM, Lombardo K, Sanders KC. (2001). A meta-analysis of incremental validity and nomological networks for subordinate and peer ratings. *Human Performance,*

*14*, 267–303.

Cooper WH. (1981). Ubiquitous halo. *Psychological Bulletin, 90*, 218–244.

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

*Bulletin, 52*, 281–302.

DeVries DL, Morrison AM, Shullman SL, Gerlach ML. (1986). *Performance appraisal on the line*. Greensboro, NC: Center for Creative Leadership.

Diefendorff JM, Silverman SB, Greguras GJ. (2005). Measurement equivalence and mul- tisource ratings for non-managerial positions: Recommendations for research and practice. *Journal of Business and Psychology, 19*, 399–425.

Facteau JD, Craig SB. (2001). Are performance ratings from different ratings sources comparable? *Journal of Applied Psychology, 86*, 215–227.

Fleishman EA. (1957). The leader opinion questionnaire. In Stogdill RM, Coons AE (Eds.), *Leader behavior: Its description and measurement* (pp. 120–133). Columbus, OH: Bureau of Business Research.

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: McGraw-Hill.

Heslin PA, Latham GP. (2004). The effect of upward feedback on managerial behavior.

*Applied Psychology: An International Review, 53*, 23–37.

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. (2008, April). Rater source effects are alive and well after all. Paper presented at the 33rd Annual Conference of the Society for Industrial and Organizational Psychology, San Francisco, CA.

Hogan J, Holland B. (2003). Using theory to evaluate personality and job-performance relations: A socioanalytic perspective. *Journal of Applied Psychology, 88*, 100–112. Hooijberg R, Choi J. (2000). Which leadership roles matter to whom? An examination of rater effects on perceptions of effectiveness. *Leadership Quarterly, 11*, 341–

364.

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.

Jones EE, Thibaut W. (1958). Interaction goals as bases of inference in interpersonal perception. In Tagiuri R, Petrullo L (Eds.), *Person perception and interpersonal behavior* (pp. 151–178), Palo Alto, CA: Stanford Univ. Press.

Joreskog KG, Sorbom D. (1996). *LISREL 8.5: User’s reference guide*. Chicago: Scientific

Software.

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.

Kaiser RB, Craig SB. (2005). Building a better mouse trap: Item characteristics associated with rating discrepancies in 360-degree feedback. *Consulting Psychology Journal: Practice and Research, 57*, 235–245.

Katz RL. (1974). Skills of an effective administrator. *Harvard Business Review, 52*, 90–102. 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.

Kenny DA, Berman JS. (1980). Statistical approaches to the correction of correlational bias. *Psychological Bulletin, 88*, 288–295.

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

Kram KE. (1985). *Mentoring at work*. Glenview, IL: Scott, Foresman and Company. Lance CE. (2008). Why assessment centers do not work the way they’re supposed to.

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

Lance CE, Baranik LE, Lau AR, Scharlau EA. (2008). If it ain’t trait it must be method: (Mis)application of the multitrait-multimethod methodology in organizational re- search. In Lance CE, Vandenberg RJ (Eds.), *Statistical and methodological myths and urban legends: Received doctrine, verity, and fable in organizational and social research*. Mahwah, NJ: Erlbaum.

Lance CE, Baxter D, Mahan RP. (2006). Evaluation of alternative perspectives on source effects in multisource performance measures. In Bennett W, Lance CE, Woehr DJ (Eds.), *Performance measurement: Current perspectives and future challenges*. Mahwah, NJ: Erlbaum.

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, Noble CL, Scullen SE. (2002). A critique of the correlated trait–correlated method (CTCM) and correlated uniqueness (CU) models for multitrait–multimethod (MTMM) data. *Psychological Methods, 7*, 228–244.

Lance CE, Woehr DJ. (1989). The validity of performance judgments: Normative accu- racy model versus ecological perspectives. In Ray DF (Ed.), *Southern Manage- ment Association Proceedings* (pp. 115–117). Oxford, MS: Southern Management Association.

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.

Landsberger HA. (1961). The horizontal dimension in bureaucracy. *Administrative Science*

*Quarterly, 6*, 298–333.

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

LeBreton JM, Burgess JD, Kaiser RB, Atchley EK, James LR. (2003). The restriction of variance hypothesis and interrater reliability and agreement: Are ratings from multiple sources really dissimilar? *Organizational Research Methods, 6*, 80–128.

Lievens F, Conway JM. (2001). Dimension and exercise variance in assessment center scores: A large-scale evaluation of multitrait–multimethod studies. *Journal of Ap- plied Psychology, 86*, 1202–1222.

London M, Smither JW. (1995). Can multisource feedback change perceptions of goal ac- complishment, self-evaluations, and performance-related outcomes? Theory-based applications and directions for future research. PERSONNEL PSYCHOLOGY, *48*,

803–839.

Mann FC. (1965). Toward an understanding of the leadership role in formal organizations.

In Dubin R, Homans GC, Mann FC, Miller DC (Eds.), *Leadership and productivity*

(pp. 68–77). San Francisco: Chandler.

McCauley, C, Lombardo M. (1990). BENCHMARKS R : An instrument for diagnosing managerial strengths and weaknesses. In Clark KE, Clark MB (Eds.), *Measures of leadership* (pp. 535–545). West Orange, NJ: Leadership Library of America.

Meng XL, Rosenthal R, Rubin DB. (1992). Comparing correlated correlation coefficients.

*Psychological Bulletin, 111*, 172–175.

Meriac JP, Hoffman BJ, Woehr DJ, Fleischer 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.

Mintzberg H. (1975). The manager’s job: Folklore and fact. *Harvard Business Review, 53*,

49–61.

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.

Murphy KR. (1989). Dimensions of job performance. In Dillon RF, Pellegrino JW (Eds.), *Testing: Theoretical and applied perspectives* (pp. 218–247). New York: Praeger.

Murphy KR, Cleveland JN. (1995). *Understanding performance appraisal: Social, orga- nizational, and goal based approaches*. Thousand Oaks, CA: Sage.

Murphy KR, DeShon, R. (2000). Interrater correlations do not estimate the reliability of job performance ratings. PERSONNEL PSYCHOLOGY, *53*, 873–900.

Nathan BR, Tippins N. (1990). The consequences of halo ‘error’ in performance ratings: A field study of the moderating effect of halo on test validation results. *Journal of Applied Psychology, 75*, 290–296.

Noe RA. (1988). An investigation of the determinants of successful assigned mentoring relationships. PERSONNEL PSYCHOLOGY, 41, 457–479.

Nunnally JC, Bernstein IH. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Ployhart RE, Schneider B, Schmitt N. (2006). *Staffing organizations: Contemporary prac- tice and theory* (3rd ed.). Mahwah, NJ: Erlbaum.

Podsakoff PM, MacKenzie SB, Podsakoff NP, Lee J. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*, 879–903.

Rupp DE, Thornton GC, Gibbons AM. (2008). The construct validity of the assessment center method and usefulness of dimensions as focal constructs. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 116–120.

Salam S, Cox J, Sims H. (1997). In the eye of the beholder: How leadership relates to

360-degree performance ratings. *Group & Organizational Management, 22*, 185–

209.

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, Viswesvaran C, Ones DS. (2000). Reliability is not validity and validity is not reliability. PERSONNEL PSYCHOLOGY, *53*, 901–912.

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

Seifert CF, Yukl G, McDonald RA. (2003). Effects of multisource feedback and a feedback facilitator on the influence behavior of managers toward subordinates. *Journal of Applied Psychology, 88*, 561–569.

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.

Smither JW, London M, Reilly RR. (2005). Does performance improve following mul- tisource feedback? A theoretical model, meta-analysis, and review of empirical findings. PERSONNEL PSYCHOLOGY, *58*, 33–66.

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

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

Tornow WW. (1993). Perceptions of reality: Is multi-perspective measurement a means or an end? *Human Resource Management, 32*, 221–230.

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 performance? A meta-analytic framework for disentangling substantive and error influences. *Journal of Applied Psychology, 90*, 108–131.

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

New York: Psychological Corporation.

Wernimoont PF, Campbell JP. (1968). Signs, samples, and criteria. *Journal of Applied*

*Psychology, 52*, 372–376.

Wherry RJ. (1952). *The control of bias in ratings: VIII. A theory of rating* (PRB report No. 922, Contract No. DA-49-083 OSA69, Department of the Army). Columbus, OH: Ohio State University Research Foundation.

Wherry RJ, Bartlett CJ. (1982). The control of bias in ratings: A theory of rating.

PERSONNEL PSYCHOLOGY, *35*, 521–551.

Widaman KF. (1985). Hierarchically nested covariance structure models for multitrait- multimethod data. *Applied Psychological Measurement, 9*, 1–26.

Woehr DJ, Sheehan MK, Bennett W. (2005). Assessing measurement equivalence across ratings sources: A multitrait–multirater approach. *Journal of Applied Psychology,*

*90*, 592–600.

