

Singapore Management University Institutional Knowledge at Singapore Management University

Research Collection Lee Kong Chian School Of
Business

Lee Kong Chian School of Business

4-2004

Revised estimates of dimension and exercise variance components in assessment center postexercise dimension ratings

Charles E. LANCE
University of Georgia

Tracy A. LAMBERT
University of Georgia

Amanda G. GEWIN
University of Georgia

Filip LIEVENS
Singapore Management University, fili lievens@smu.edu.sg

James M. CONWAY
Central Connecticut State University
DOI: <https://doi.org/10.1037/0021-9010.89.2.377>

Follow this and additional works at: https://ink.library.smu.edu.sg/lkcsb_research

Part of the [Human Resources Management Commons](#), [Industrial and Organizational Psychology Commons](#), and the [Organizational Behavior and Theory Commons](#)

Citation

LANCE, Charles E.; LAMBERT, Tracy A.; GEWIN, Amanda G.; LIEVENS, Filip; and CONWAY, James M.. Revised estimates of dimension and exercise variance components in assessment center postexercise dimension ratings. (2004). *Journal of Applied Psychology*. 89, (2), 377-385. Research Collection Lee Kong Chian School Of Business.
Available at: https://ink.library.smu.edu.sg/lkcsb_research/5610

This Journal Article is brought to you for free and open access by the Lee Kong Chian School of Business at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection Lee Kong Chian School Of Business by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email libIR@smu.edu.sg.

Revised Estimates of Dimension and Exercise Variance Components in Assessment Center Postexercise Dimension Ratings

Charles E. Lance, Tracy A. Lambert, and
Amanda G. Gewin
The University of Georgia

Filip Lievens
Ghent University

James M. Conway
Central Connecticut State University

The authors reanalyzed assessment center (AC) multitrait-multimethod (MTMM) matrices containing correlations among postexercise dimension ratings (PEDRs) reported by F. Lievens and J. M. Conway (2001). Unlike F. Lievens and J. M. Conway, who used a correlated dimension-correlated uniqueness model, we used a different set of confirmatory-factor-analysis-based models (1-dimension-correlated Exercise and 1-dimension-correlated uniqueness models) to estimate dimension and exercise variance components in AC PEDRs. Results of reanalyses suggest that, consistent with previous narrative reviews, exercise variance components dominate over dimension variance components after all. Implications for AC construct validity and possible redirections of research on the validity of ACs are discussed.

Assessment centers (ACs) exhibit some of the highest criterion-related validities among alternative predictors of job performance (Gaugler, Rosenthal, Thornton, & Bentson, 1987; Schmidt & Hunter, 1998). Nevertheless, the construct validity of AC postexercise dimension ratings (PEDRs), that is, dimension ratings that are made at the completion of each exercise, has continued to be called into question for over 20 years (Lance, Newbolt, et al., 2000). Thus there is an apparent evidential dilemma that ACs demonstrate criterion-related validity but not construct validity (Lievens & Klimoski, 2001; Sackett & Tuzinski, 2001).

ACs were originally designed to assess candidate performance relating to various dimensions (e.g., Organizing & Planning, Perception, Oral and Written Communication) as they are assessed in various exercises (e.g., Leaderless Group Discussion, In-Basket, Role Play, etc.). Sackett and Dreher (1982) found, however, that factor analysis of PEDRs resulted in factors that reflected exercises, not the dimensions that were intended to be assessed. Subsequently, a number of additional studies have investigated the construct validity of PEDRs by using factor analysis, but only a few of these (e.g., Arthur, Woehr, & Maldegen, 2000) have supported a theoretical structure that includes both exercise and dimension factors. However, dozens of other studies have replicated Sackett and Dreher's (1982) basic findings that PEDRs substantially reflect the exercises in which they are assessed and not the

dimensions they are designed to assess (Howard, 1997; Lance, Newbolt, et al., 2000; Lievens & Klimoski, 2001; Sackett & Tuzinski, 2001). As Robertson, Gratton, and Sharpley (1987) stated "it is exercises (work samples) not dimensions that best represent the underlying structure of assessor ratings" (p. 193). Furthermore, these findings hold true despite various experimental efforts to increase dimension variance and decrease exercise variance in PEDRs (Lance, Newbolt, et al., 2000; Lievens, 1998; Woehr & Arthur, 1999). "This robust finding . . . has led many researchers to question the construct validity of assessment ratings" (Schleicher, Day, Mayes, & Riggio, 2002, p. 735). Howard (1997) summarized this state of affairs by stating "exercises and not dimensions are the currency of assessment centers" (p. 21).

However, findings from a recent large-scale quantitative review of multitrait-multimethod (MTMM) studies of the construct validity of PEDRs seem to indicate otherwise (Lievens & Conway, 2001). Although previous narrative reviews have concluded that exercise variance dominates over dimension variance in PEDRs, Lievens and Conway's (2001) quantitative review of 34 MTMM studies indicated that PEDRs reflect equal proportions of exercise and dimension variance, a more optimistic conclusion for the traditional AC community but one that conflicts with other reviews of this literature. How can this be? One possibility is that Lievens and Conway's quantitative review was more objective and consequently more accurate than previous qualitative reviews. A second possibility is that their findings were biased as the result of adopting a particular approach to the analysis of MTMM data, the correlated trait-correlated uniqueness (CTCU) model. In fact, Lievens and Conway noted that a "limitation of this . . . model is that if the restriction of independent methods does not hold, it suffers from a biasing effect" (p. 1212). The purpose of this study was to reanalyze the data reported by Lievens and Conway with the goal of determining whether findings based on the CTCU model may have led to upwardly biased estimates of proportions of

Charles E. Lance, Tracy A. Lambert, and Amanda G. Gewin, Department of Psychology, The University of Georgia; Filip Lievens, Department of Personnel Management, Ghent University, Belgium; James M. Conway, Department of Psychology, Central Connecticut State University.

This research was supported in part by Grant 10-21-RR274-221 from the National Institute on Aging to Charles E. Lance.

Correspondence concerning this article should be addressed to Charles E. Lance, Department of Psychology, The University of Georgia, Athens, GA 30602-3013. E-mail: clance@uga.edu

trait (i.e., dimension) variance components in PEDRs and, if so, to estimate the extent of this bias.

In the remainder of this article, we first summarize and critique the Lievens and Conway (2001) quantitative review. Second, we point to specific sources of bias in the estimation of trait and method variance components under the CTCU model. Finally, we reanalyze data reported by Lievens and Conway and, in doing so, show that analyses based on a different set of models than those reported by Lievens and Conway result in estimated proportions of trait (i.e., Dimension) and method (i.e., Exercise) variance represented in PEDRs that are consistent with previous narrative reviews and that suggest that Exercise effects do in fact dominate over Dimension effects in AC PEDRs.

The Lievens and Conway (2001) Review

Lievens and Conway (2001) assembled an impressive database consisting of 34 MTMM matrices of AC ratings that (a) were composed of correlations among PEDRs, (b) included at least three Dimensions and two Exercises, (c) in which assessors rotated across AC Exercises, (d) were positive definite, and (e) had “sample size. . . greater than 50 and greater than the number of parameters estimated” (p. 1206). Lievens and Conway fit a number of confirmatory factor analysis (CFA) models to each MTMM matrix, including (a) a correlated-Dimensions 0-Exercise (CD0E) model, (b) a 0-Dimension correlated-Exercises (0DCE) model, (c) a correlated-Dimension correlated-Exercises (CDCE) model, and (d) a 1-Dimension correlated-Exercise model (1DCE) model. We note here that the majority of previous CFAs of PEDR MTMM matrices have supported either the 0DCE model, indicating strong cross-situational (i.e., cross-Exercise) specificity in AC performance, or a 1DCE model, indicating cross-situational consistency (convergent validity) in assessing a “general trait or person factor” (Schleicher et al., 2002, p. 735) plus cross-situational (i.e., cross-Exercise) specificity in AC performance, although overall goodness-of-fit for these models is not always strong. Lievens and Conway also fit two other models to each PEDR MTMM matrix: Browne’s (1984) Direct Product (DP) model and Marsh’s (1989) correlated-dimension correlated-uniqueness (CDCU) model in which exercise effects are modeled as covariances between the uniquenesses of PEDRs measured in the same exercise. For each model fit, the model was judged to be appropriate if it returned a convergent and admissible solution and if it provided acceptable model goodness-of-fit indices. Lievens and Conway found that 53% of the CDCU models resulted in proper and well-fitting solutions. For the other models, these percentages were as follows: DP, 52%; CD0E, 3%; 0DCE, 24%; 1DCE, 29%; and CDCE, 9%. Primarily on the basis of these results, Lievens and Conway chose the CDCU model to conduct their quantitative review. We note here that their criteria for model selection—model fit, convergence, and admissibility—are important and appropriate. However, we also note that selection of the two CFA models that have found empirical support in previous literature, the 0DCE and 1DCE models, would also have resulted in 53% of the MTMM matrices being retained. We return to this issue later.

Having selected the CDCU model as the analytic model of choice, Lievens and Conway (2001) estimated Dimension variance components as PEDRs’ squared loadings on Dimension factors and the Exercise variance components using a residualized CFA technique described by Scullen (1999) and found overall that the

proportions of Dimension and Exercise variance in PEDRs were equal (34%) across the studies analyzed. We note that Lievens and Conway based their analyses on all CDCU solutions (admissible as well as inadmissible), although they commented that their “conclusions did not change much when only the 18 admissible [CDCU] solutions (those free of improper estimates) were considered” (p. 1208). They also investigated the effects of various AC design factors on proportions of Dimension and Exercise variance in PEDRs, but we are not concerned about these issues here. Our concern is that the CDCU model that they chose as their analytic approach likely provided biased estimates of trait (i.e., dimension) and method (i.e., exercise) variance components.

Bias in CFA Parameter Estimates Under the CDCU Model

Lance, Noble, and Scullen (2002) recently provided a critique of correlated trait-correlated method (CTCM, analogous to the CDCE model) and correlated trait-correlated uniqueness (CTCU, analogous to the CDCU model) models for MTMM data. Lance et al. (2002) noted that the CTCU model has gained popularity among researchers, because it does, in fact, return convergent and proper solutions for MTMM data more often than does the CTCM model. However, beyond this advantage, Lance et al. (2002) identified several, and in some cases severe, shortcomings of the CTCU model. The two most relevant shortcomings here are (a) the necessary assumption of orthogonal Method (i.e., exercise) effects, and (b) the potential for upward bias in Trait (i.e., dimension) factor loadings under the CTCU model. The following section explains these problems in some detail.

Under the CTCM model, monotrait–heteromethod (MTHM), heterotrait–heteromethod (HTHM), and heterotrait–monomethod (HTMM) correlations are modeled as functions of estimated model parameters, respectively, as follows:

$$\text{MTHM} = \lambda_{Tij}\lambda_{Tij'} + \lambda_{Mij}\lambda_{Mij'}\phi_{MjMj'}, \quad (1a)$$

$$\text{HTHM} = \lambda_{Tij}\lambda_{Ti'j'}\phi_{TiTi'} + \lambda_{Mij}\lambda_{Mi'j'}\phi_{MjMj'}, \quad (1b)$$

$$\text{HTMM} = \lambda_{Tij}\lambda_{Ti'j'}\phi_{TiTi'} + \lambda_{Mij}\lambda_{Mi'j'}, \quad (1c)$$

where λ_{Tij} refers to the standardized loading of the ij th Trait-Method Unit (TMU, i.e., a measure of the i th Trait as measured by the j th measurement Method) on the i th latent Trait factor, λ_{Mij} is the ij th TMU’s loading on the j th Method factor, $\phi_{TiTi'}$ refers to the correlation between different Trait factors, and $\phi_{MjMj'}$ refers to the correlation between different Method factors. Under the CTCU model these correlations are modeled as follows:

$$\text{MTHM} = \lambda_{Tij}\lambda_{Tij'}, \quad (2a)$$

$$\text{HTHM} = \lambda_{Tij}\lambda_{Ti'j'}\phi_{TiTi'}, \quad (2b)$$

$$\text{HTMM} = \lambda_{Tij}\lambda_{Ti'j'}\phi_{TiTi'} + \theta_{\delta(ij,i'j')}. \quad (2c)$$

Comparing Equations 1c and 2c, one can see that the CTCM and CTCU models account for common Method effects in HTMM correlations by alternative parameterizations of the same covariance component as follows: (a) the CTCM model uses the common causal effect of the measurement Method shared in common (i.e., $\lambda_{Mij}\lambda_{Mi'j'}$ in Equation 1c) and (b) the CTCU model uses the covariance between the uniquenesses of TMUs that share the same

measurement Method (i.e., $\theta_{\delta(ij,i'j)}$ in Equation 2c). However, under the CTCM model, Method effects appear also for MTHM ($\lambda_{Mij}\lambda_{Mij'}\phi_{MjMj'}$ from Equation 1a) and HTHM ($\lambda_{Mij}\lambda_{Mi'j'}\phi_{MjMj'}$ from Equation 1b) correlations. That is, some portion of the MTHM and HTHM correlations are assumed to be due to effects of correlated Methods on TMUs. Note however, that these components do not appear for the MTHM and HTHM correlations under the CTCU model (Equations 2a and 2c). In special cases in which one or both of the $\lambda_{Mij}s = 0$ or $\phi_{MjMj'} = 0$, Equations 1a and 1b reduce to their corresponding Equations 2a and 2b. If, however, all $\lambda_{Mij}s \neq 0$ and $\phi_{MjMj'} \neq 0$, that is, if TMUs' loadings on Method factors are nonzero and correlations between different Methods are nonzero, then Equations 2a and 2b are misspecified by omitting relevant variance components. In the present context, this would be the case in which Exercise factor loadings are nonzero (a very common occurrence) as are Exercise factor correlations (also a common occurrence as we show later). If Exercise factor loadings and Exercise factor correlations are nonzero, there are sources of covariance in the MTHM and HTHM correlations that are not accounted for in Equations 2a and 2b. As a result, and as Conway, Lievens, Scullen, and Lance (2002) showed empirically, the omission of these effects in the CTCU model results in upwardly biased estimates of the $\lambda_{Tij}s$ (leading to inflated estimates of convergent validity) and the $\phi_{TjTj'}s$ (leading to underestimated estimates of discriminant validity), and when the omitted λ_{Mij} and $\phi_{MjMj'}$ effects are large, the bias incurred in the estimation of the $\lambda_{Tij}s$ and $\phi_{TjTj'}s$ is substantial. This is a direct consequence of the type of model misspecification known as the *unmeasured variables problem* (James, 1980) introduced by the CTCU model (Lance et al., 2002).

Reanalysis of the Lievens and Conway (2001) Database

Studies Included

We included all studies analyzed previously by Lievens and Conway (2001), plus five others that met their criteria for inclusion that were not available to them previously (matrices reported by Lance et al. 2000, Lance, Foster, Gentry, & Thorensen, 2004).

Models Tested

We used CFA to estimate three correlated exercise models: (a) a CDCE model, which includes both correlated Dimension and correlated Exercise factors; (b) a 1DCE model, which posits convergent validity in the assessment of a general performance factor across exercises, plus Exercise factors; and (c) a 0DCE model, which posits only Exercise factors, that is, a model that specifies only cross-situationally (i.e., cross-exercise) specific performance factors. As noted earlier, the 1DCE and 0DCE models have received the majority of empirical support among AC construct validity studies using CFA of PEDRs. We also estimated two models in which exercise effects were modeled as covariances among uniquenesses for PEDRs measured in the same Exercise: (a) a CDCU model, which corresponds to the CDCE model, and (b) a IDCU model, which corresponds to the 1DCE model.

Model Selection

We required that models retained for the analysis of dimension and exercise variance components yield proper solutions. Obtain-

ing a proper solution is one key requirement in evaluating model fit—improper solutions usually indicate that the model being fit is inconsistent with the data (Marsh, 1994). Improper solutions were those that contained estimated factor correlations or standardized factor loadings greater than 1.0 in absolute value or negative unique variances. We also considered model fit in selecting models for further analysis, but large differences in sample size (N s ranged from 59 to 1170), matrix size (from 3D2E to 10D8E matrices), and model parsimony (e.g., 1DCE model df ranged from 11 to 690) complicated matters. Basically, we considered those models that returned admissible solutions as defining the set of plausible models for each study and relied on (a) the difference in chi-square ($\Delta\chi^2$) and (b) the difference in comparative fit index (ΔCFI , based on Bentler's [1990] CFI; Cheung & Rensvold, 2001) tests to determine the model that best fit the data. Statistically significant $\Delta\chi^2$ values and ΔCFI s greater than .01 were taken as indicating significant differences in model fit.

Table 1 summarizes results relevant for model selection. For example, the row for the Bobrow and Leonards (1997) data set indicates that improper estimates were obtained in the Λ , Φ , and Θ_{δ} matrices for the CDCE model and in the Φ matrix for the CDCU model. Other entries in the Bobrow and Leonards (1997) row present model degrees of freedom, chi-square, and CFI values for models that returned proper solutions (e.g., $\chi^2[225, N = 196] = 540.72$; CFI = .91 for the 1DCE model). For data sets in which more than one CE model returned admissible solutions, we conducted $\Delta\chi^2$ and ΔCFI tests to determine which model best fit the data because the 0DCE model is nested within the 1DCE model, which is in turn nested within the CDCE model. For example, because the 1DCE model fit the data better than did the 0DCE model for the Arthur (2000) data set, $\Delta\chi^2(12, N = 149) = 230.31$, $p < .001$; $\Delta CFI = .25$, and because the CDCE model fit even better than did the 1DCE model, $\Delta\chi^2(6) = 90.61$, $p < .001$; $\Delta CFI = .07$, we selected the CDCE model and its corresponding CDCU model for subsequent analysis of Dimension and Exercise variance components.

As Table 1 shows, the CDCE model fared poorly, returning inadmissible solutions for all but two data sets (Arthur (2000) and Veldman-1 (1994); the 1 indicates that this was the first of more than one data set reported by Veldman). These results are common among tests of AC construct validity and are consistent with some literature that indicates that the CTCM model for MTMM data can suffer empirical identification problems (Brannick & Spector, 1990; Kenny & Kashy, 1992; Marsh & Bailey, 1991). However, these findings are also consistent with previous reviews of AC construct validity studies that indicate that PEDRs generally do not demonstrate significant convergent and discriminant validity in assessing distinct dimensions. Consistent with previous reviews, Table 1 indicates that, among the CE models, either a 1DCE or a 0DCE model most often resulted in an admissible solution. For data sets in which both 1DCE and 0DCE models returned admissible solutions, the 1DCE model always provided a better fit to the data, both in terms of the $\Delta\chi^2$ and ΔCFI tests. Eighty-five percent of the admissible 1DCE models met traditional standards for a well-fitting model according to the CFI index (CFI > .90) and 60% met Hu and Bentler's (1999) more stringent criterion (CFI > .95), indicating that, generally speaking, the 1DCE model fit the data sets analyzed well.

Data sets for which only the 0DCE model returned an admissible solution were not selected for further analysis of dimension and

Table 1
Model Convergence, Admissibility, and Goodness-of-Fit for Models Tested

Study	CDCE	1DCE	0DCE	CDCU	1DCU
Arthur et al. (2000)	33/13.51/1.00	39/104.12/.93	51/334.43/.68	30/16.08/1.00	36/273.19/.73
Becker (1990)					
1	$\Lambda, \Phi, \Theta_{\delta}$	Λ, Θ_{δ}	113/154.58/.95	Λ, Θ_{δ}	Λ, Θ_{δ}
2	Φ	96/114.49/.96	113/156.40/.90	Φ	91/116.90/.94
Bobrow & Leonards (1997)	$\Lambda, \Phi, \Theta_{\delta}$	225/540.72/.91	249/769.64/.86	Φ	168/210.56/.99
Bycio et al. (1987)	Φ	690/2270.54/.97	730/2788.44/.96	Φ	600/1194.92/.99
Chorvat (1994)	Θ_{δ}	397/968.80/.65	428/1160.26/.55	226/315.12/.97	321/855.19/.67
Donahue et al. (1997)	Φ	487/736.55/.95	521/843.83/.93	Φ	399/513.69/.98
Fleenor (1996)	$\Lambda, \Phi, \Theta_{\delta}$	Λ, Θ_{δ}	168/330.30/.74	Φ	148/420.86.65
Fredricks (1989)	Λ, Θ_{δ}	72/141.35/.93	87/232.52/.85	Λ, Θ_{δ}	60/106.67/.95
Harris et al. (1993)	Φ	218/796.74/.91	242/1265.41/.85	182/343.49/.98	203/841.53/.91
Joyce et al. (1994)					
1	$\Lambda, \Phi, \Theta_{\delta}$	$\Lambda, \Phi, \Theta_{\delta}$	$\Lambda, \Phi, \Theta_{\delta}$	$\Lambda, \Phi, \Theta_{\delta}$	69/116.43/.85
2	$\Lambda, \Phi, \Theta_{\delta}$	$\Lambda, \Phi, \Theta_{\delta}$	62/81.69/.82	Φ	42/46.57/.96
Kleinmann et al. (1994)					
1	$\Lambda, \Phi, \Theta_{\delta}$	Φ	Φ	15/10.04/1.00	18/20.34.99
2	$\Lambda, \Phi, \Theta_{\delta}$	Φ	Φ	$\Lambda, \Phi, \Theta_{\delta}$	Λ, Θ_{δ}
Kleinmann et al. (1996)	$\Lambda, \Phi, \Theta_{\delta}$	Λ, Θ_{δ}	24/46.61/.91	15/12.41/1.00	18/27.84/.96
Kleinmann (1997)					
4	$\Lambda, \Phi, \Theta_{\delta}$	15/11.28/1.00	24/63.71/.89	15/11.64/1.00	18/18.46/1.00
5	$-df$	Θ_{δ}	8/12.17/.99	0/.16/1.00	3/5.88/.99
6	$-df$	Λ, Θ_{δ}	8/9.55/.99	0/.83/1.00	3/4.39/.99
Kolk et al. (2000)					
1	$-df$	Λ, Θ_{δ}	Λ, Θ_{δ}	0/0	3/8.39/.96
2	$-df$	Λ, Θ_{δ}	Λ, Θ_{δ}	Φ	3/3.61/1.00
Kudisch et al. (1997)	$\Lambda, \Phi, \Theta_{\delta}$	127/273.78/.85	146/395.04/.74	$\Lambda, \Phi, \Theta_{\delta}$	115/284.27/.82
Lance, Newbolt, et al. (2000)					
1	Φ	140/281.82/.95	160/356.24/.93	Φ	135/260.06/.96
2	Φ	158/312.12/.90	179/473.07/.81	Φ	152/323.91/.98
3	Φ	72/132.52/.95	87/208.26/.91	Φ	57/80.95/.98
Lance et al. (2004)					
4	Φ	72/291.71/.95	87/510.61/.91	Φ	57/95.54/.99
5	Φ	72/151.62/.98	87/372.00/.92	Φ	57/101.66/.99
Lievens & Van Keer (1999)	$\Lambda, \Phi, \Theta_{\delta}$	65/96.45/.98	80/154.38/.95	Φ	73/143.51/.95
Parker (1992)	$\Lambda, \Phi, \Theta_{\delta}$	$\Lambda, \Phi, \Theta_{\delta}$	402/1141.36/.70	$\Lambda, \Phi, \Theta_{\delta}$	Λ, Θ_{δ}
Robie (2000)	$\Lambda, \Phi, \Theta_{\delta}$	11/13.64/.99	19/37.46/.96	Φ	8/15.56/.98
Sagie & Magnezy (1997)					
1	$\Lambda, \Phi, \Theta_{\delta}$	72/122.63/.97	87/169.85/.96	Φ	60/89.06/.99
2	Φ	72/155.88/.93	87/269.18/.90	50/77.39/.98	60/134.78/.96
Schleicher et al. (1999)					
1	$\Lambda, \Phi, \Theta_{\delta}$	15/6.91/1.00	24/19.43/.98	15/7.91/1.00	18/16.93/1.00
2	Λ, Θ_{δ}	Λ, Θ_{δ}	24/36.05/.96	15/21.40/.98	18/33.59/94
Schneider & Schmit (1992)	Φ	Λ, Θ_{δ}	48/65.92/.97	39/40.85/1.00	42/57.02/.98
Sweeney (1976)					
1	$-df$	Λ, Θ_{δ}	8/8.27/1.00	0/.34/1.00	3/2.21/1.00
2	$-df$	Λ, Θ_{δ}	8/40.40/.95	Φ	3/0.98/1.00
Van der Velde et al. (1994)	Φ	Λ, Θ_{δ}	101/198.98/.93	$\Lambda, \Phi, \Theta_{\delta}$	69/91.82/.98
Veldman (1994)					
1	26/41.95/.98	41/123.74/.90	53/238.90/.78	9/17.96/.99	24/91.49/.92
2	$\Lambda, \Phi, \Theta_{\delta}$	41/95.41/.87	53/191.99/.67	9/9.14/1.00	24/69.02/.89

Note. CDCE = correlated Dimension-correlated Exercise model; 1DCE = one-dimension-correlated Exercise model; 0DCE = zero-dimension-correlated exercise model; CDCU = correlated Dimension-correlated uniqueness model; 1DCU = one-dimension-correlated uniqueness model; Λ = the LISREL matrix of factor loadings; Φ = the LISREL matrix of factor correlations; Θ_{δ} = the LISREL matrix of unique variances (and covariances). Presentation of these matrices in the table identifies the matrices in which improper estimates were located (λ_{ij} , ϕ_{ij} , and Θ_{δ} elements $> |1.00|$, $\text{diag}(\Theta_{\delta}) \leq 0.0$) in the event that a solution was inadmissible. Tabled numbers refer to the model *dfs*, the χ^2 statistic, and comparative fit index (CFI) values (i.e., $df/\chi^2/CFI$) for admissible solutions. Numbers in boldface indicate models that were selected for further comparative analysis of Dimension and Exercise variance components. Numbers associated with citations throughout the table indicate that multiple data sets were obtained for reanalysis from some cited articles. The numbers indicate Dataset 1, Dataset 2, and so forth.

exercise variance components because the 0DCE model specifies no Dimension factors. Those data sets for which the 1DCE model returned an admissible solution were retained along with the corresponding 1DCU models because these models do specify a general Dimension factor along with multiple Exercise factors. Models selected for further analysis are shown in boldface in Table 1.

Estimation of Variance Components

Dimension variance components were calculated, for both the CE and CU models and each data set separately, by squaring Dimension factor loadings (so as to index the absolute effect size and to avoid the problem of having some negative factor loadings

Table 2
Summary of Variance Components Analysis for Studies That Returned Admissible CDCE and CDCU Model Solutions

Study	Mean squared Dimension factor loadings		Mean squared Dimension factor correlations		Mean squared Exercise factor correlations	Mean squared Exercise factor loadings	
	CDCE	CDCU	CDCE	CDCU	CDCE	CDCE	CDCU
Arthur et al. (2000)	.51	.62	.48	.35	.24	.21	.08
Veldman-1 (1994)	.43	.37	.29	.35	.01	.28	.26
<i>M</i>	.48	.50	.38	.35	.16	.25	.17

Note. CDCE = correlated Dimension-correlated Exercise model; CDCU = correlated Dimension-correlated uniqueness model. The hyphen and number associated with Veldman indicates Dataset 1.

cancel out other positive loadings), converting the squared factor loadings to z scores (because averaging raw correlational effect sizes yields biased estimates of the means, but averaging z score equivalents yields unbiased estimates; see James, Demaree, & Mulaik, 1986), averaging the z scores, and back-transforming the average z to the estimated mean squared factor loading. Exercise variance components were computed similarly for the CE models from the Exercise factor loadings. Mean squared Dimension and Exercise factor correlations were also computed similarly by squaring the r s, converting the r^2 s to z s, averaging the z s and converting the mean z back to an r^2 . Exercise variance components for the CU models were calculated using Scullen's (1999) method in which (a) the Θ_{δ} matrix is saved from the CTCU or IDCU model, (b) separate unidimensional CFAs are conducted on covariances among uniquenesses for PEDRs measured in the same Exercise, and (c) variance components are calculated as the squared unstandardized factor loadings.

Results

Table 2 shows results for the analysis of variance components from the CDCE models. Results for the Arthur et al. (2000) data set show the expected upward bias in estimated dimension variance components (mean dimension $\lambda^2 = .62$ and $.51$ for the CDCU model and the CDCE model, respectively). The amount of bias here is not trivial, but it is also not substantial (the mean dimension variance component estimate was 21.6% higher under the CDCU model compared with the CDCE model) due to the fact that (a) exercise variance components are not large (mean Exercise $\lambda^2 = .21$) and (b) the mean Exercise factor correlation is moderate (mean Exercise factor $\phi^2 = .21$, so that the mean estimated $\phi = .46$). Bias in dimension variance components is negligible for the Veldman (1994) data set, owing to the fact that the mean Exercise factor correlation was $\phi = .09$. As such, results in Table 2 are inconclusive because there were only two data sets for which the CDCE model (and its corresponding CDCU model) was selected for analysis of variance components. Nevertheless, and consistent with Lance et al.'s (2002) algebraic developments and Conway et al.'s (2002) Monte Carlo findings, upward bias in trait (i.e., dimension) factor loadings incurred only when Exercise factor loadings and Exercise factor correlations were nonzero (the Arthur (2000) data set).

Table 3 shows results for the analysis of variance components from the IDCU models, models that have been supported empir-

ically in AC literature. Findings here are clear. Bias in dimension variance components for the IDCU model (mean dimension $\lambda^2 = .27$) is evident as compared with the IDCCE model (mean Dimension $\lambda^2 = .14$)—the IDCU model indicates that nearly twice as much variance in PEDRs is attributable to dimensions, as compared with estimates from the IDCCE model, and this difference is statistically significant (paired samples $t[19] = 4.176, p < .001$). This bias arises from the omission of significant exercise variance components (mean Exercise $\lambda^2 = .52$) and from nontrivial Exercise factor correlations (mean Exercise factor $\phi^2 = .18$; mean $\phi = .42$) in the estimation of Dimension factor loadings under the IDCU model. Interestingly, the IDCU model appeared to substantially underestimate exercise variance components (mean Exercise $\lambda^2 = .36$, which is comparable with Lievens & Conway's [2001] estimated value of $.34$) as compared with the IDCCE model (mean Exercise $\lambda^2 = .52$, paired samples $t[19] = 4.348, p < .001$), although we had not predicted this result. As such, results in Table 3 further verify algebraic developments presented by Lance et al. (2002) and Monte Carlo findings presented by Conway et al. (2002) indicating that the CU model yields upward-biased estimates of Trait (in the present case, Dimension) factor variance components. Compared with the IDCCE model, the IDCU model overestimated dimension variance components by 93% and underestimated exercise variance components by 31% on the average. These biases are substantial.¹

Discussion

Results from Lievens and Conway's (2001) quantitative review of MTMM studies on the construct validity of PEDRs seemed

¹ Of course, this assumes that findings from the IDCCE model are unbiased, that is, that the IDCCE is the correct model and the IDCU model is not. However, because the data that we reanalyzed were from operational ACs, we could not know which model, the IDCCE, IDCU, or one of many other conceivable models, was the true population model. Consequently, although there is evidence that the CU model tends to provide biased model parameter estimates (Conway et al., 2002), we have no way of assessing the relative accuracy of the IDCCE and IDCU models' variance components estimates here. Nevertheless, our interpretations of results reported here are consistent with (a) Lance et al.'s (2002) algebraic developments, (b) Conway et al.'s (2002) Monte Carlo findings that the CU model does return biased estimates for trait factor loadings, and (c) previous findings in the literature on construct validity of ACs (e.g., Sackett & Tuzinski, 2001).

Table 3
Summary of Variance Components Analysis for Studies That Returned Admissible 1DCE and 1DCU Model Solutions

Study	Mean squared Dimension factor loadings		Mean squared Exercise factor correlations	Mean squared Exercise factor loadings	
	1DCE	1DCU	1DCE	1DCE	1DCU
Becker-2 (1990)	.11	.21	.14	.43	.33
Bobrow & Leonards (1997)	.14	.18	.08	.57	.45
Bycio et al. (1987)	.03	.33	.21	.74	.40
Chorvat (1994)	.17	.18	.17	.09	.11
Donahue et al. (1997)	.06	.35	.23	.55	.33
Fredericks (1989)	.08	.07	.16	.63	.63
Harris et al. (1993)	.14	.22	.11	.35	.26
Kleinman-4 (1997)	.26	.43	.18	.52	.30
Kudisch et al. (1997)	.09	.20	.17	.45	.33
Lance, Newbolt, et al. (2000)					
1	.24	.22	.06	.40	.35
2	.08	.12	.05	.33	.27
3	.07	.11	.05	.71	.64
Lance et al. (2004)					
4	.35	.38	.10	.54	.55
5	.08	.47	.10	.75	.35
Lievens & Van Keer (1999)	.32	.30	.08	.39	.33
Robie (2000)	.04	.37	.37	.65	.27
Sagie & Magnezy (1997)					
1	.11	.22	.13	.41	.30
2	.11	.21	.13	.41	.30
Schleicher-1 et al. (1999)	.05	.49	.38	.76	.29
Veldman-2 (1994)	.18	.31	.52	.43	.28
<i>M</i>	.14	.27	.18	.52	.36
<i>SD</i>	.09	.12	.12	.17	.16

Note. 1DCE = one-dimension-correlated Exercise model; 1DCU = one-dimension-correlated uniqueness model. Numbers associated with citations throughout table indicate that multiple data sets were obtained for reanalysis from some cited articles. The numbers indicate Dataset 1, Dataset 2, and so forth.

counter to conventional wisdom as represented in previous reviews of the AC literature in suggesting that PEDRs reflect equal proportions of dimension and exercise variance. Lievens and Conway's study was cleverly conceived and meticulously executed, but some of their major findings were based on an analytic approach to MTMM data, the CU model, that is now known to provide biased estimates of Trait factor variance, and in some cases this bias can be substantial (Conway et al., 2002). Lievens and Conway acknowledged this as a potential limitation to their study and suggested that "future research should . . . examine more thoroughly the possible upward bias of the correlated uniqueness model" (p. 1212). This was the primary aim of this study. We reanalyzed Lievens and Conway's database using a CFA model that (a) has been supported empirically in the AC literature and (b) may be less prone than the CU model to biased parameter estimates (i.e., the 1DCE model). Results of these reanalyses indicate that, consistent with previous narrative reviews of the AC literature, exercise variance components (mean Exercise $\lambda^2 = .52$) dominated over dimension variance components (mean Dimension $\lambda^2 = .14$). This is good news and bad news. On the one hand, the present study's results help explain the discrepancy between conclusions from narrative reviews of AC construct validity and those offered by Lievens and Conway—it seems that biased estimates of Trait variance components obtained from the CU model provided a more optimistic picture of the proportion of dimension variance

in PEDRs than was actually warranted. This reconciliation is the good news.

The bad news, at least from a traditional perspective on AC construct validity, is that the present reanalyses indicate that AC construct validity cannot be salvaged from reanalysis of previously reported data. That is, Lievens and Conway's (2001) reassuring findings may have resulted as an artifact of adopting the CU model as the analytic model of choice. However, an emerging literature suggests that the ways researchers have asked questions about the construct validity of ACs may also not have been appropriate to begin with (e.g., Kolk, Born, & van der Flier, 2001; Lance, Newbolt, et al., 2000; Lance et al., 2004; Lievens, 2001). For example, initial expectations that candidate performance should typically be cross-situationally consistent and differentiated across dimensions may be inconsistent with actual candidate behaviors that are typically cross-situationally (i.e., cross-exercise) specific and undifferentiated (i.e., generally good or poor) across dimensions (Lance, Newbolt, et al., 2000). This interpretation is consistent with recent experimental findings indicating that when candidate performance is manipulated to be consistent with AC researchers' initial expectations (i.e., cross-situationally consistent and differentiated across dimensions), assessors rate candidate performance accordingly, but when candidate performance is manipulated to be cross-situationally specific and not differentiated across dimensions (the pattern that is consistent with findings

reported here), assessors also rate this pattern accordingly (Lievens, 2002). In other words, assessors' judgments in experimental settings tend to be quite veridical. We speculate that assessors' judgments in operational ACs are also quite veridical, owing to their experience with positions to which AC candidates aspire and the extensive training that is usually afforded assessors of candidate performance that is, in fact, cross-situationally specific (varies across Exercises) and is not well differentiated across dimensions.

Along with other recent research (Lance, Newbolt, et al., 2000, Lance, et al., 2004), the present findings are troubling to traditional AC theory and practice. The original design of the AC architecture was based on the theory that the dimensions assessed in ACs represent cross-situationally stable traits that can be assessed in various exercises (International Task Force on Assessment Center Guidelines, 2002). That is, exercises were implicitly viewed as representing different opportunities for candidates to display behaviors reflecting cross-situationally (i.e., cross-exercise) stable traits. However, empirical findings accumulated over the last 20 years do not support this theoretical architecture—correlations between different dimensions within the same exercise tend to be much larger than the correlations between dimension across different exercises (Woehr & Arthur, 1999). In other words, exercise effects dominate over dimension effects on PEDRs and, as has been documented by several reviews (e.g., Lievens, 1998; Lievens & Conway, 2001; Sackett & Tuzinski, 2001; Woehr & Arthur, 1999), various AC design modifications have had little or no effect on this basic pattern of findings. Rather, and contrary to the original design of the AC architecture, candidate performance seems to be relatively undifferentiated across the various dimensions defined for each exercise and cross-situationally specific as exercises define different performance situations.

This raises the question of what constructs are reflected by AC PEDRs. One way of interpreting the last 20 years of research on AC construct validity is that ACs lack construct validity, a position that seems to have driven various attempts at modifying AC technology toward making ACs more construct valid. Sackett and Tuzinski (2001) took an alternative view, stating that "assessment centers do not 'lack construct validity,' but rather lack clear consensus as to the constructs they do assess . . . at issue is whether the traditional explanation of what constructs assessment centers measure is the correct one." (p. 118). We speculate that AC performance is best characterized as consisting of two broad components: (a) general, but situation-specific, performance factors corresponding to overall performance in each exercise qua work sample task (Lance, Johnson, Douthitt, Bennett, & Harville, 2000; Smith, 1991; Teachout & Pllum, 1991), and (b) a cross-situationally stable overall performance factor that is driven by stable traits such as cognitive ability, conscientiousness, and experience.

We see at least two implications of the present findings for AC practice. On the one hand, findings of a lack of AC construct validity (in a traditional sense) do not threaten, and are not inconsistent with, accumulated findings that ACs demonstrate predictive validity (Sackett & Tuzinski, 2001; Schmidt & Hunter, 1998). ACs are valid predictors of performance criteria, just not for the reasons originally thought. On the other hand, the present findings suggest that the AC practice of providing performance feedback to candidates according to AC dimensions may not be justified because AC PEDRs substantially reflect exercise effects and not dimension

effects. Instead, developmental feedback might be more profitably focused on specific aspects of exercise performance that are closely tied to job performance requirements.

So what are we to do? We think that investigating additional AC design modifications intended to make PEDRs' covariance structure conform to expectations based on the original design of ACs is pointless because efforts so far have made little difference. Rather, researchers might focus on job analysis techniques that yield information on critical behaviors required on the job that may consist of important tasks and dimensions. Even another approach could be based on the ideas that (a) candidate AC performance is generally not differentiated across dimensions within exercises and is largely cross-situationally specific, (b) exercises should more properly be viewed as work samples rather than mere opportunities for candidates to display dimension-relevant behavior, (c) exercises qua work samples can be carefully designed to capture important aspects of jobs for which assessment is conducted, and (d) developmental feedback can be provided with reference to key work sample elements or task steps.

Lowry (1995, 1997) provided examples of this approach. Lowry (1997) distinguished between traditional "dimension-specific" (p. 53) ACs that are designed to assess traits, such as organizing and planning, leadership, and communication, and "task-specific" ACs, which are designed to evaluate "how well the subject performs important tasks encountered on the job" (p. 54). Both types of ACs are developed on the basis of job analysis, but whereas dimension-specific ACs are founded on identification of important job dimensions, task-specific ACs are founded on identifying critical job tasks, based on analyses of "the importance, frequency, and need to perform on entry to the position" (p. 55). The goal in the task-specific AC is to construct work samples such as tactical simulations, role plays, and in-baskets that simulate these critical job tasks. Lowry (1997) also presented an example of how checklists of actions required to successfully accomplish tasks for each exercise can be developed. These can be used to structure assessor observation, scale the effectiveness of possible actions toward task accomplishment, facilitate overall assessment of task accomplishment, and structure detailed feedback to candidates as to actions taken that were particularly effective and ineffective. Although Lowry (1995) presented some preliminary and favorable psychometric evidence for task-specific ACs, there is otherwise little or no additional evidence on their reliability and validity. We see this as one key need for future research.

Conclusion

Two decades of research on the construct validity of AC PEDRs suggests that they substantially reflect exercise effects and not the dimensions that they were designed to measure, and our reanalyses of Lievens and Conway's (2001) database confirms this. Clearly, ACs exhibit criterion-related validity but do not seem to reflect the constructs that were intended. We think the time has come to recognize this and redirect research on ACs toward a more thorough understanding of what constructs are being tapped by assessor ratings and how the effectiveness of these ratings might be enhanced.

References

References marked with an asterisk indicate studies included in the meta-analysis.

- *Arthur, W. A., Jr., Woehr, D. J., & Maldegen, R. (2000). Convergent and discriminant validity of assessment center dimensions: A conceptual and empirical re-examination of the assessment center construct-related validity paradox. *Journal of Management* 26, 813–835.
- *Becker, A. S. (1990). *The effects of a reduction in assessor roles on the convergent and discriminant validity of assessment center ratings*. Unpublished doctoral dissertation, University of Missouri, St. Louis.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238–246.
- *Bobrow, W., & Leonards, J. S. (1997). Development and validation of an assessment center during organizational change. *Journal of Social Behavior and Personality*, 12, 217–236.
- Brannick, M. T., & Spector, P. E. (1990). Estimation problems in the block-diagonal model of the multitrait–multimethod matrix. *Applied Psychological Measurement*, 14, 325–339.
- Browne, M. W. (1984). The decomposition of multitrait–multimethod matrices by generalized least squares. *British Journal of Mathematical and Statistical Psychology*, 37, 1–21.
- *Bycio, P., Alvares, K. M., & Hahn, J. (1987). Situational specificity in assessment center ratings: A confirmatory factor analysis. *Journal of Applied Psychology*, 72, 463–474.
- Cheung, G. W., & Rensvold, R. B. (2001). The effects of model parsimony and sampling error on the fit of structural equation models. *Organizational Research Methods*, 4, 236–264.
- *Chorvat, V. P. (1994). *Toward the construct validity of assessment center leadership dimensions: A multitrait-multimethod investigation using confirmatory factor analysis*. Unpublished doctoral dissertation, University of South Florida, Tampa.
- Conway, J. M., Lievens, F., Scullen, S. E., & Lance, C. E. (2002, April). *Bias in the correlated uniqueness model for MTMM data: A simulation study*. Paper presented at the meeting of the Society for Industrial and Organizational Psychology, Toronto, Ontario, Canada.
- *Donahue, L. M., Truxillo, D. M., Cornwell, J. M., & Gerrity, M. J. (1997). Assessment center construct validity and behavioral checklists: Some additional findings. *Journal of Social Behavior and Personality*, 12, 85–108.
- *Fleenor, J. W. (1996). Constructs and developmental assessment centers: Further troubling empirical findings. *Journal of Business and Psychology*, 10, 319–333.
- *Fredricks, A. J. (1989). *Assessment center ratings: Models and processes*. Unpublished doctoral dissertation, University of Nebraska, Lincoln.
- Gaugler, B. B., Rosenthal, D. B., Thornton, G. C., III, & Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology*, 72, 493–511.
- *Harris, M. M., Becker, A. S., & Smith, D. E. (1993). Does the assessment center scoring method affect the cross-situational consistency of ratings? *Journal of Applied Psychology*, 78, 675–678.
- Howard, A. (1997). A reassessment of assessment centers: Challenges for the 21st century. *Journal of Social Behavior and Personality*, 12(5), 13–52.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis. *Structural Equation Modeling*, 6, 1–55.
- International Task Force on Assessment Center Guidelines (2000). Guidelines and ethical considerations for assessment center operations. *Public Personnel Management*, 28, 315–331.
- James, L. R. (1980). The unmeasured variables problem in path analysis. *Journal of Applied Psychology*, 65, 415–421.
- James, L. R., Demaree, R. G., & Mulaik, S. A. (1986). A note on validity generalization procedures. *Journal of Applied Psychology*, 71, 440–450.
- *Joyce, L. W., Thayer, P. W., & Pond, S. B. (1994). Managerial functions: An alternative to traditional assessment center dimensions? *Personnel Psychology*, 47, 109–121.
- Kenny, D. A., & Kashy, D. A. (1992). Analysis of the multitrait–multimethod matrix by confirmatory factor analysis. *Psychological Bulletin*, 112, 165–172.
- *Kleinmann, M. (1997). Transparenz der Anforderungsdimensionen: Ein Moderator der Konstrukt- und Kriteriumsvalidität des Assessment-Centers [Transparency of the required dimensions: A moderator of assessment centers' construct and criterion validity]. *Zeitschrift für Arbeits- und Organisationspsychologie*, 41, 171–181.
- *Kleinmann, M., Andres, J., Fedtke, C., Godbersen, F., & Köller, O. (1994). Der Einfluss unterschiedlicher auswertungsmethoden auf die Konstruktvalidität von Assessment-Centern [The influence of different rating procedures on the construct validity of assessment center methods]. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 41, 184–210.
- *Kleinmann, M., Kuptsch, C., & Köller, O. (1996). Transparency: A necessary requirement for the construct validity of assessment centers. *Applied Psychology: An International Review*, 45, 67–84.
- *Kolk, N. J., Born, M. P., & Van der Flier, H. (2000, April). *The transparent assessment center: The effect of revealing dimensions to applicants*. Paper presented at the meeting of the Society for Industrial and Organizational Psychology, New Orleans, LA.
- Kolk, N. J., Born, M. P., & van der Flier, H. (2001, April). *Common rater variance as explanation for the lack of construct validity of assessment center dimensions*. Paper presented at the meeting of the Society for Industrial and Organizational Psychology, San Diego, CA.
- *Kudisch, J. D., Ladd, R. T., & Dobbins, G. H. (1997). New evidence on the construct validity of diagnostic assessment centers: The findings may not be so troubling after all. *Journal of Social Behavior and Personality*, 12, 129–144.
- *Lance, C. E., Foster, M. R., Gentry, W. A., & Thorensen, J. D. (2004). Assessor cognitive processes in an operational assessment center. *Journal of Applied Psychology*, 89, 22–35.
- Lance, C. E., Johnson, C. D., Douthitt, S. S., Bennett, W., Jr., & Harville, D. L. (2000). Good news: Work sample administrators' global performance judgments are (about) as valid as we've suspected. *Human Performance*, 13, 253–277.
- *Lance, C. E., Newbolt, W. H., Gatewood, R. D., Foster, M. R., French, N., & Smith, D. E. (2000). Assessment center exercise factors represent cross-situational specificity, not method bias. *Human Performance*, 13, 323–353.
- Lance, C. E., Noble, C. L., & Scullen, S. E. (2002). A critique of the correlated trait-correlated method and correlated uniqueness models for multitrait–multimethod data. *Psychological Methods*, 7, 228–244.
- Lievens, F. (1998). Factors which improve the construct validity of assessment centers: A review. *International Journal of Selection and Assessment*, 6, 141–152.
- Lievens, F. (2001). Assessors and use of assessment center dimensions: A fresh look at a troubling issue. *Journal of Organizational Behavior*, 22, 203–221.
- Lievens, F. (2002). Trying to understand the different pieces of the construct validity puzzle of assessment centers: An examination of assessor and assessee effects. *Journal of Applied Psychology*, 87, 675–686.
- Lievens, F., & Conway, J. M. (2001). Dimension and exercise variance in assessment center scores: A large-scale evaluation of multitrait–multimethod studies. *Journal of Applied Psychology*, 86, 1202–1222.
- Lievens, F., & Klimoski, R. J. (2001). Understanding the assessment centre process: Where are we now? *International Review of Industrial and Organizational Psychology*, 16, 246–286.
- *Lievens, F., & Van Keer, E. (1999, May). *Modeling method effects in assessment centers: An application of the correlated uniqueness approach*. Paper presented at the meeting of the Society for Industrial and Organizational Psychology, Atlanta, GA.
- Lowry, P. E. (1995). The assessment center process: Assessing leadership in the public sector. *Public Personnel Management*, 22, 443–450.
- Lowry, P. E. (1997). The assessment center process: New directions. *Journal of Social Behavior and Personality*, 12, 53–62.
- Marsh, H. W. (1989). Confirmatory factor analysis of multitrait–

- multimethod data: Many problems and a few solutions. *Applied Psychological Measurement*, 13, 335–261.
- Marsh, H. W. (1994). Confirmatory factor analyses of factorial invariance: A multifaceted approach. *Structural Equation Modeling*, 1, 5–34.
- Marsh, H. W., & Bailey, M. (1991). Confirmatory factor analysis of multitrait–multimethod data: A comparison of alternative models. *Applied Psychological Measurement*, 15, 47–70.
- *Parker, M. W. (1992). *A construct validation of the Florida Principal Competencies Assessment Center using confirmatory factor analysis*. Unpublished doctoral dissertation, University of South Florida, Tampa.
- Robertson, I., Gratton, L., & Sharpley, D. (1987). The psychometric properties and design of managerial assessment centres: Dimensions into exercises won't go. *Journal of Occupational Psychology*, 60, 187–195.
- *Robie, C., Adams, K. A., Osburn, H. G., Morris, M. A., & Etchegaray, J. M. (2000). Effects of the rating process on the construct validity of assessment center dimension evaluations. *Human Performance*, 13, 355–370.
- Sackett, P. R., & Dreher, G. F. (1982). Constructs and assessment center dimensions: Some troubling empirical findings. *Journal of Applied Psychology*, 67, 401–410.
- Sackett, P. R., & Tuzinski, K. (2001). The role of dimensions and exercises in assessment center judgments. In M. London (Ed.), *How people evaluate others in organizations* (pp. 111–129). Mahwah, NJ: Erlbaum.
- *Sagie, A., & Magnezy, R. (1997). Assessor type, number of distinguishable dimension categories, and assessment centre construct validity. *Journal of Occupational and Organizational Psychology*, 70, 103–108. (Sagie – 1 & Sagie – 2).
- *Schleicher, D. J., Day, D. V., Mayes, B. T., & Riggio, R. E. (1999, May). *A new frame for frame-of-reference training: Enhancing the construct validity of assessment centers*. Paper presented at the meeting of the Society for Industrial and Organizational Psychology, Atlanta, GA.
- Schleicher, D. J., Day, C. V., Mayes, B. T., & Riggio, R. E. (2002). A new frame for frame-of-reference training: Enhancing the construct validity of assessment centers. *Journal of Applied Psychology*, 87, 735–746.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262–274.
- *Schneider, J. R., & Schmitt, N. (1992). An exercise design approach to understanding assessment center dimension and exercise constructs. *Journal of Applied Psychology*, 77, 32–41.
- Scullen, S. E. (1999). Using confirmatory factor analysis of correlated uniquenesses to estimate method variance in multitrait–multimethod matrices. *Organizational Research Methods*, 2, 275–292.
- Smith, F. D. (1991). Work samples as measures of performance. In A. K. Wigdor & B. F. Green, Jr. (Eds.), *Performance assessment for the workplace* (Vol. 2, pp. 27–52). Washington, DC: National Academy Press.
- *Sweeney, D. C. (1976). *The development and analysis of rating scales for the Chicago Police Recruit Assessment Center*. Unpublished manuscript, Bowling Green State University, OH.
- Teachout, M. S., & Pellum, M. W. (1991). *Air force research to link standards for enlistment to on-the-job performance* (AFHRL-TR-90–90). Brooks AFB, TX: Air Force Human Resources Laboratory, Training Systems Division.
- *Van der Velde, E. G., Born, M. P., & Hofkes, K. (1994). Begripsvalidering van een assessment center met behulp van confirmatorische factoranalyse [Construct validity of an assessment center using confirmatory factor analysis]. *Gedrag en Organisatie*, 7, 18–26.
- *Veldman, W. M. (1994). *Assessment centers and candidates' personal qualities: A study on the correlations between assessment center ratings*. Unpublished doctoral dissertation, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands.
- Woehr, D. J., & Arthur, W., Jr. (1999, April). *The construct-related validity of assessment center ratings: A review and meta-analysis of the role of methodological factors*. Paper presented at the meeting of the Society for Industrial and Organizational Psychology, Atlanta, GA.

Received January 30, 2003
 Revision received April 8, 2003
 Accepted May 6, 2003 ■

Wanted: Your Old Issues!

As APA continues its efforts to digitize journal issues for the PsycARTICLES database, we are finding that older issues are increasingly unavailable in our inventory. We are turning to our long-time subscribers for assistance. If you would like to donate any back issues toward this effort (preceding 1982), please get in touch with us at journals@apa.org and specify the journal titles, volumes, and issue numbers that you would like us to take off your hands.