Factor Structure of Two Different Developmental Assessment Center Rating Formats

Boris Yanovsky
Veterans Health Administration National Center for Organization Development

Katerine Osatuke
Veterans Health Administration National Center for Organization Development

Robin Graff-Reed
Veterans Health Administration National Center for Organization Development

Sue R. Dyrenforth
Veterans Health Administration National Center for Organization Development

Construct validity of assessment center (AC) dimensions has long been scrutinized. Although ACs remain some of the best predictors of job performance, their construct validity repeatedly falls short. We examined construct validity of two rating formats within a developmental AC (DAC). Within-exercise ratings loaded onto their exercises, while the within-dimension rating factors represented multiple dimensions. Our findings suggest that dimensions within ACs exist, but AC designers must be aware of how to structure the rating process to enable the raters to accurately observe these dimensions. This is particularly important for DACs, where dimensions serve an important role.

INTRODUCTION

The first use of assessment centers (ACs) in a civilian setting took place nearly 50 years ago when AT&T used this method to study adult development (Thornton & Krause, 2009). Since then, ACs have become popular tools adopted by various types of organizations worldwide. Although first used in the military by the Office of Strategic Services (OSS) for the selection of intelligence agents during World War II (Bray, 1995; Kello, 2007; Office of Strategic Services, 1948), ACs quickly made the transition to the private sector and are well received and widely used today. The popularity of ACs is not surprising as they have been shown to be good predictors of job performance (Arthur, Day, McNelly, & Edens, 2003) and they're preferred to other predictors by applicants (Macan, Avedon, Paese, & Smith, 1994). In fact, in a recent meta-analysis, Meriac, Hoffman, Woehr, and Fleischer (2008) found that in addition to sharing 18% of variance with measures of job performance, ACs also explain incremental validity above personality and cognitive ability tests. Taken together, the findings across the literature suggest that ACs are one of the best predictors of job performance.

In addition to being valid selection tools, ACs are also sought after because of their flexibility. The design and format of ACs over the last 50 years has been as diverse as the organizations adopting, creating, and implementing them. Although ACs can differ on a case by case basis, in order to be considered an AC, the method must involve multiple raters observing multiple behaviors. A set of guidelines for ACs does exist, but it still gives AC designers a great deal of flexibility to tailor ACs to suit their needs (International Task Force on Assessment Center Guidelines [ITFACG], 2009). Additionally, ACs can be used for purposes other than just selection. Other popular uses of ACs are for the purposes of promotion, diagnosis of training needs, and employee development. When used for decision making purposes, ACs best serve as predictors of future performance and behavior. In contrast, when used for identifying areas for development, ACs need to be effective diagnostic tools of current performance and behavior. Additionally, ACs must balance learning and development with diagnosis in order to be useful for developmental purposes (ITFACG, 2009).

Until recently, the popular use of ACs has been for administrative and decision making purposes in selection and promotion (Kudisch, et al., 2001; Spychalski, Quinones, Gaugler, & Pohley, 1997). This trend may be declining, however, as a survey showed that slightly more organizations were using ACs for developmental purposes (including diagnosis of training needs) than for selection or promotion (Thornton and Krause, 2009). Another survey showed ACs being used for developmental purposes more frequently than for selection purposes, but the most common purpose was the use of ACs for promotion within an organization (Eurich, Krause, Cigularov, Thornton, 2009). It is clear that ACs are not strictly used for selection purposes and are designed with different goals in mind; therefore it's important to separately evaluate ACs with respect to their intended goals.

Much attention has been given to ACs in the Industrial-Organizational (I-O) psychology literature, with a focus mostly on ACs for selection purposes. Although the findings have been mostly favorable (e.g. criterion-related validity), a concern over construct validity exists. Researchers are finding that the dimensions in ACs continue to routinely demonstrate exercise effects instead of dimension effects. That is, different AC dimensions are usually highly intercorrelated within exercises, while correlations for the same dimensions across exercises are low (Howard, 1997; Lance, Lambert, Gewin, Lievens, & Conway, 2004). This can pose a major setback to the AC method, especially considering the multiple purposes and intensions of the AC and the reliance some organizations place upon them.

ACs that incorporate feedback into the exercises and assessments to help facilitate participants' professional development are known as developmental assessment centers (DACs) (Rupp, Snyder, Gibbons, & Thornton, 2006; Thornton & Rupp, 2003, 2006). In order to maximize the utility of DACs, the dimensions used within them must allow for the potential of development (Rupp et al., 2006). Furthermore, specific feedback is provided on behaviors that represent these dimensions. In fact, Thornton and Rupp (2006) recommend that feedback is detailed, extensive, provided constantly, and should be used for action planning in a DAC. The use of feedback in DACs is markedly different from traditional ACs, where feedback is seldom used, and only presented at the end of the assessment with a pass/fail. Thus, the role that dimensions play in DACs is fundamentally different from traditional ACs. Because feedback should be specific, the importance of distinct dimensions and clear behaviors within them is underscored in DACs. This feature of the DAC exemplifies how poor construct validity can undermine the utility of this method.

The construct validity of traditional AC dimensions has been questioned by many AC scholars (Howard, 1997; Lance et al., 2004). Although evidence for the poor convergent and discriminant validities exists in the literature, some recent studies have nevertheless found signs of hope for AC dimensions. For example, Dilchert and Ones (2009) examined the incremental validity of AC dimension scores and overall AC scores over personality and cognitive ability tests. Results showed that separate dimension scores demonstrated incremental validity above personality and cognitive ability tests, while overall ratings did not (Dilchert & Ones, 2009). Also, in a meta-analysis described earlier, Meriac et al. (2008) found that six of the seven AC dimensions examined uniquely accounted for significant variance in job performance. These findings suggest that although previous studies failed to find differences between AC dimensions, these dimensions still uniquely contribute to the criterion.

One other artifact of AC research that encumbers our understanding of the roles of dimensions is the overwhelming focus on traditional ACs. Studies investigating DACs are scarce compared to their selection counterparts (Rupp et al., 2006; Thornton and Krause, 2009), and findings from traditional ACs should not be generalized to ACs used for other purposes (Carrick & Williams, 1999). This is further exemplified by the differences between various AC methods and purposes discussed in an earlier section of the paper. This inquiry into understanding the roles of dimensions within ACs may be even more critical for DACs due to the reliance on the dimensions to deliver feedback and foster development. Therefore, the purpose of this paper is to contribute to DAC research by examining the dimensions within a DAC.

Although DACs do not get as much attention in the literature, the studies that have focused on DACs appear to obtain similar findings as those obtained in traditional AC literature. For example, Fleenor (1996) found that exercise factors dominated DAC ratings rather than dimension factors. This problem, one that seems to be common to most ACs, may not have as much to do with the dimensions as with the rating process. This was examined in a study by Robie, Osburn, Morris, and Etchegaray (2000), in which the researchers manipulated how the raters observe and evaluate dimensions. In this experiment, raters were assigned to two groups: Some raters were instructed to only provide ratings on a single dimension in each exercise (referred to as a *Within-Dimension* format), and others were instructed to provide ratings of all dimensions within a single exercise, the latter of which takes the approach of the common *Within-Exercise* rating method. Factor analysis results showed the presence of dimension factors in the *Within-Dimension* approach, and a presence of exercise factors in the *Within-Exercise* approach. Additionally, the authors found that, compared to the *Within-Exercise* format, monotrait-heteromethod correlations were higher and heterotrait-monomethod correlations were lower for the *Within-Dimension* format. In another study, Silverman, Dalessio, Woods, and Johnson (1986) found similar results when examining the *Within-Exercise* rating process versus the *Within-Dimension* ratings process, which led them to conclude that "different assessment center evaluation methods forced the raters to process and organize the assessment center data in different ways" (p. 573). Raters in the *Within-Exercise* approach, which is widely used in ACs, were forced to think of their ratings in the context of the performance on a particular exercise. On the other hand, the raters in the *Within-Dimension* approach were forced to focus on one dimension in particular, allowing for differences in the dimensions to surface.

As suggested by some of the research findings in the literature, the *Within-Dimension* approach to ratings uncovers the dimensions and bolsters the construct validity of ACs. However, this rating format is not without its own limitations. One limitation of the *Within-Dimension* approach is that the evidence of discriminant validity may be caused by differences between raters perceptions. That is, the raters in these conditions observed only one dimension and were not able to provide ratings on other dimensions to ensure agreement. Additionally, this rating format would be difficult and impractical to incorporate in an applied setting. To properly implement this method in an AC, the AC designer must assign a separate rater for each dimension being assessed. The number of raters would then have to double if one also wanted to examine rater agreement. This type of rating format appears to be costly in terms of resources invested, and may not be a feasible choice in organizational settings.

A *Within-Dimension* rating approach does offer AC researchers a better understanding of the rating process from the raters point of view, and the promising signs of convergent and discriminant validity are encouraging. Therefore, this study aims to examine AC dimensions in a *Within-Exercise* rating approach and a modified *Within-Dimension* rating approach which allows raters to provide ratings on all dimensions. This study also focuses on a DAC, where the dimensions may have even greater implications.

The limitation of having each rater rate only one dimension can be a deterrent that prevents the *Within-Dimension* rating format from being accepted in the field. In this paper, we propose a rating format that modifies the *Within-Dimension* format and allows for a rater to rate all of the dimensions. This rating method, along with some benefits it could introduce, is detailed below.

Several versions of rating processes have been presented in the AC literature, with the *Within-Exercise* and the *Within-Dimension* approaches emerging as the two primary methods (Meriac et al.,

2008; Robie et al., 2000; Sackett & Dreher, 1982; Woehr & Arthur, 2003). In a *Within-Exercise* approach, raters rate all dimensions after each exercise has been completed, whereas in the *Within-Dimension* approach, only one dimension is rated by a rater.

Both of these common AC rating methods have their respective strengths and weaknesses. The *Within-Exercise* rating approach, while providing the raters with the opportunity to evaluate all behaviors immediately after observation, suffers from a presence of exercise factors rather than dimension factors (Fleenor, 1996; Sackett & Dreher 1982). On the other hand, the *Within-Dimension* rating approach provides evidence for the existence of dimensions (Silverman et al., 1986). However, as explained above, the *Within-Dimension* approach is not free from flaws as it may be seen as unpractical by AC administrators. In addition, the *Within-Dimension* approach could potentially suffer from measurement error through a lack of agreement if the number of raters is not, at the very least, doubled.

The method presented in this paper is a slight variation on both of these approaches, with a focus on preserving the beneficial features of each. This method is examined in comparison to a typical *Within-Exercise* method and factor structures of dimensions are presented.

METHOD

Participants

Mid-level managers (*N*=101) participated in the DAC in January, 2010 and January, 2011 as part of a national leadership development program. This DAC is used within the Veterans Health Administration's (VHA) Executive Career Field (ECF) to prepare high-potential leaders selected from the top end of mid-management ranks.

Materials

The ECF Feedback & Critical Skills Assessment Center is a four-day program that uses simulations and other experiential activities to assess relevant skills and assist the participants in focusing their developmental efforts. Each participant is accompanied by his/her assigned mentor, as well as a preceptor from their home facility. A mentor is someone who currently holds a position in the participant's chosen career track, while a preceptor is a supervisor of the participant from their facility. Both mentors and preceptors will share the similar role of providing feedback to the participant and both will have the same opportunities to observe and assess their assigned participant throughout the DAC.

Participants in this DAC were rotated through exercises designed to simulate actual roles that would be assumed once they step into their executive positions. These exercises were constructed to assess eight specific competencies relevant to all VHA training and development programs. The dimensions within this DAC are the competencies themselves; where performance, ratings, and feedback are based on the behaviors grouped within the competencies.

Some exercises were modified from the 2010 administration of the DAC to the 2011 administration. Only the five exercises that were preserved from year to year were used for this study. In addition to the five exercises, the *Within-Dimension* ratings were also available in both DAC administrations. The matrix in Table 1 shows which exercises were used and the competencies that were represented. Additionally, the list of competencies and their definitions can be found in Table 2.

TABLE 1
EXERCISES AND DIMENSIONS

	Executive Team Simulation 1	Executive Team Simulation 2	In-Box	Meeting Management	Human Resource Skills
Interpersonal Effectiveness	**X**	**X**	**X**	**X**	**X**
Customer Service	**X**	**X**	**X**		
Systems Thinking	**X**	**X**	**X**	**X**	
Flexibility/Adaptability	**X**	**X**			
Creative Thinking	**X**	**X**			
Organizational Stewardship	**X**	**X**	**X**		
Personal Mastery	**X**	**X**			
Technical Knowledge & Skills	**X**	**X**	**X**	**X**	**X**

*Exercises are listed in top row and dimensions are listed in first column.

TABLE 2
COMPETENCY DEFINITIONS

Competency Name	Competency Definition
Interpersonal Effectiveness	The ability to build and sustain relationships, to resolve conflict, to handle negotiation effectively, and to develop collaborative working relationships. A high scorer on this competency displays empathy, empowers others, and possesses written and oral communication skills.
Customer Service	The ability to integrate customer service (including patient satisfaction and stakeholder support) into a management plan. A high scorer on this competency enhances internal and external customer satisfaction; models customer service by handling complaints effectively and promptly and by ensuring a customer-centered focus in direction and daily work; uses customer feedback in planning and providing products and services; and encourages subordinates to meet or exceed customer needs and expectations.
Systems Thinking	The ability to understand the pieces as a whole and appreciate the consequences of actions on other parts of the system. A high scorer on this competency thinks in context; knows how to link actions with others in the organization; demonstrates awareness of process, procedures, and outcomes; and possesses a big (whole) picture view of the world.
Flexibility & Adaptability	The ability to quickly adapt to change, handle multiple inputs and tasks simultaneously, and accommodate new situations and realities. A high scorer on this competency works well with all levels and types of people, welcomes divergent ideas, and maximizes limited resources.
Creative Thinking	The ability to think and act innovatively, to look beyond current reality to forecast future direction, to take risks, to challenge traditional assumptions, and to solve problems creatively. A high scorer on this competency takes advantage of difficult or unusual situations to develop unique, original approaches and useful solutions.
Organizational Stewardship	A high scorer on the competency is sensitive to the needs of individuals and of the organization and provides services to both; assumes accountability for self, others, and the organization; demonstrates commitment to people and trusts others.
Personal Mastery	The ability to recognize personal strengths and weaknesses and to engage in continuous learning and self-development. A high scorer on this competency demonstrates a willingness to take actions to change and takes charge of own career.
Technical Knowledge & Skills	The knowledge and skills to perform and evaluate the work of the organization based upon a clear understanding of the processes, standards, methods, and technologies of the organization.

Procedure

Two types of rating formats were adopted for this DAC. In the *Within-Exercise* approach, the mentor and preceptor provided ratings on dimensions at the conclusion of each exercise. Raters were able to provide multiple ratings in each exercise since all exercises used for the purpose of this study assessed multiple dimensions.

Additionally, in the modified *Within-Dimension* approach, the mentor and preceptor of each participant was asked to provide ratings on each dimension at the completion of the entire DAC. For each dimension, the raters were asked to focus on examples of behaviors that represented the specific dimension across all exercises during the DAC. In this modified format, instead of observing and rating only one dimension after each exercise, the raters were instructed to think of the dimension in question in isolation. The raters were allowed to provide separate ratings for all eight dimensions assessed in the DAC. Each dimension was comprised of three items to allow for construct validity of the dimensions to be examined, resulting in 24 items. The latter approach is somewhat a hybrid approach of overall AC ratings and the *Within-Dimension* approach.

RESULTS

The two types of rating formats were examined using principal components factor analysis. The criterion of eigenvalues greater than 1 was used to determine the number of factors to extract. All factors were rotated using a Varimax rotation procedure.

Ratings by mentors and preceptors were combined into one rating for each candidate because these two ratings were not different from each other. Interrater agreement scores were calculated between mentors and preceptors for each participant using both moderate and slightly skewed distribution estimates of expected variance, as recommended by LeBreton & Senter (2007). The moderate and slightly skewed null distribution r_{WG} indices indicated strong agreement between mentor and preceptor ratings (moderate skew r_{WG}=.76; slight skew r_{WG}=.82), exceeding the traditional cut-off point of .70 (Lance, Butts, & Michels, 2006; LeBreton, Burgess, Kaiser, Atchley, & James, 2003). Based on the interrater agreement indices, mentor and preceptor scores were aggregated, resulting in 101 possible ratings for each item.

For the *Within-Exercise* rating format, ratings on 26 items from the five DAC exercises were used. Five factors were initially extracted, explaining a total of 76% of the variance. All of the 26 items loaded respectively on the exercises in which they were observed, resulting in a total of five interpretable factors (See Table 3).

TABLE 3
FACTOR LOADINGS OF WITHIN-EXERCISE RATINGS

	Factor Loadings				
Ratings	1	2	3	4	5
Executive Team Simulation Day 1					
Flexibility/Adaptability	**.74**	.42	.20	.11	.01
Interpersonal Effectiveness	**.84**	.17	.22	.10	-.01
Systems Thinking	**.80**	.30	.17	.09	.04
Personal Mastery	**.83**	.25	.14	.09	.01
Technical Knowledge & Skills	**.74**	.31	.31	.06	.01
Creative Thinking	**.75**	.39	.26	.05	-.01
Organizational Stewardship	**.74**	.38	.20	.01	-.01
Customer Service	**.76**	.37	.08	-.04	.05

Executive Team Simulation Day 2					
Personal Mastery	.36	**.68**	.21	.13	.04
Interpersonal Effectiveness	.44	**.68**	.20	.07	.18
Creative Thinking	.41	**.68**	.24	.05	.02
Flexibility/Adaptability	.27	**.81**	.17	.03	.06
Technical Knowledge & Skills	.43	**.66**	.31	.02	-.04
Systems Thinking	.24	**.82**	.19	.01	.01
Customer Service	.35	**.72**	.19	-.05	-.09
Organizational Stewardship	.23	**.78**	.16	-.05	.02
In-Box					
Customer Service	.26	.25	**.79**	.08	-.11
Interpersonal Effectiveness	.24	.12	**.82**	.08	.03
Technical Knowledge & Skills	.13	.24	**.81**	.04	-.09
Systems Thinking	.22	.27	**.82**	.04	.10
Organizational Stewardship	.20	.20	**.85**	.00	.13
Meeting Management					
Interpersonal Effectiveness	.12	.01	.04	**.89**	.06
Technical Knowledge & Skills	.08	.09	.12	**.89**	.14
Systems Thinking	.03	-.03	.01	**.88**	.07
Human Resource Skills					
Interpersonal Effectiveness	.04	.06	.09	.22	**.87**
Technical Knowledge & Skills	.00	.02	-.05	.04	**.93**

For the *Within-Dimension* rating format, 24 items were used (three ratings per dimension) from the final ratings obtained at the end of the DAC. Three factors were extracted, which explained 70% of the variance. The dimensions in this rating process were reduced from eight to three, and a majority of the ratings grouped together under their intended dimension (See Table 4).

TABLE 4
FACTOR LOADINGS OF WITHIN-DIMENSION RATINGS

	Factor Loadings		
	1	2	3
First Dimension			
Creative Thinking item 1	**.81**	.38	.20
Creative Thinking item 2	**.79**	.39	.23
Creative Thinking item 3	**.68**	.45	.23
Flexibility/Adaptability item 1	**.68**	.23	.36
Flexibility/Adaptability item 2	**.67**	.36	.39
Flexibility/Adaptability item 3	**.58**	.39	.47

Interpersonal Effectiveness item 1	**.77**	.21	.32
Interpersonal Effectiveness item 2	**.59**	.38	.54
Customer Service item 1	**.77**	.25	.28
Organizational Stewardship item 1	**.69**	.42	.24
Second Dimension			
Systems Thinking item 1	.16	**.81**	.20
Systems Thinking item 2	.35	**.81**	.21
Systems Thinking item 3	.35	**.73**	.20
Technical Skills item 1	.31	**.76**	.26
Technical Skills item 2	.38	**.74**	.24
Customer Service item 2	.24	**.75**	.15
Organizational Stewardship item 2	.19	**.50**	.43
Third Dimension			
Personal Mastery item 1	.15	.14	**.79**
Personal Mastery item 2	.34	.30	**.68**
Personal Mastery item 3	.51	.02	**.56**
Technical Skills item 3	.32	.47	**.64**
Interpersonal Effectiveness item 3	.45	.28	**.59**
Customer Service item 3	.43	.46	**.54**
Organizational Stewardship item 3	.44	.42	**.52**

DISCUSSION

This study aimed to further investigate the roles of dimensions within ACs, specifically focusing on DACs. Developmental Assessment Centers are increasingly becoming popular training and professional development tools across various organizations (Ballantyne & Povah, 2004; Kudisch et al., 2001; Spychalski et al., 1997). Whereas traditional ACs have been used in the private sector for decades, in contrast, DACs are still in their stages of infancy. It is widely reported that ACs possess exemplary criterion related validity, but lack construct related validity (Arthur et al., 2003; Howard, 1997; Lance et al., 2004; Meriac et al., 2008). According to Arthur, Woehr, and Maldegen (2000), however, these findings are paradoxical and AC construct related validity does exist. Furthermore, studies have demonstrated construct related validity of ACs by manipulating the way ratings are given (Robie et al., 2000; Silverman et al., 1986). Lack of construct related validity would be especially problematic to DACs due to their central purpose of using dimensions for feedback and development.

Previous studies have examined different rating processes of traditional ACs, with some studies concluding that it is not the dimensions that are flawed, but the way ratings are given that prevent accurate observations from being captured (Silverman et al., 1986). The current study examined whether the rating process affected dimensions by adopting a rating format that allowed for raters to provide scores on each dimension at the end of the AC. To allow for construct related validity analysis, we expanded each dimension to three ratings. This method is a variation of overall AC ratings; the latter is a common rating technique used in traditional ACs that may not always allow for internal construct related validity analysis (Meriac et al., 2008).

Findings from the factor analyses suggest that when ratings are provided at the end of the exercise, raters focus on the overall performance of that exercise and do not discriminate between dimensions. However, when ratings are provided at the end of the AC, some distinct dimensions seem to appear. This is evidenced by the presence of five factors in the *Within-Exercise* method that loaded perfectly on their respective five exercises. The competencies used in all of the exercises were defined consistently and where therefore expected to be present throughout the DAC. Contrary to the expectation and the design of the DAC, the competencies were not related to each other across exercises, but were more similar to the unique exercise in which they were observed. On the other hand, in the *Within-Dimension* method, some dimensions appeared to be distinct from each other and three factors emerged. That is, the ratings within exercises were similar to each other, whereas the ratings given at the end of the AC measured at least some of the distinct dimensions that were the focus of the AC.

These findings demonstrate that not all rating approaches are appropriate for a DAC, suggesting that further research into this topic is warranted. Within a DAC, development occurs through feedback presented to the participant on specific dimensions (Thornton and Rupp, 2006). If raters are not able to differentiate between dimensions within an exercise, then their feedback to the participants will not be as precise as was intended within a DAC. The *Within-Exercise* rating approach could undermine a crucial aspect of a DAC, participant development. By failing to identify and differentiate between varying performances across distinct competencies, this rating format represents an obstacle: it prevents accurate feedback from ever reaching candidates. An approach similar to a *Within-Dimension* process may be more appropriate within a DAC since it appears to allow the rater to focus less on the overall performance of an exercise and to think more about the role dimensions play across the DAC.

Limitations

An important consideration when conducting a factor analysis is sample size. Some disagreement exists in the literature regarding the appropriate sample size for a factor analysis procedure. For example, recommendations in the literature for a ratio of sample size, *N*, to items used, *p*, range anywhere from 3:1 to 10:1 (MacCallum, Widaman, Zhang, & Hong, 1999). This means that, depending on which literature is referenced, a researcher could end up with a recommended sample size of anywhere between 90 and 300 for a 30 item measure. In their paper addressing sample size requirements in factor analysis, MacCallum et al. (1999) offer some guidance. Specifically, the authors note that appropriate sample sizes will vary across studies.

In the interest of obtaining a larger sample size, data from two administrations of the DAC were combined, resulting in a sample size of 101. The number of items used in each method were 26 for the *Within-Exercise,* and 24 for the *Within-Dimension* rating method, resulting in an approximate *N*:*p* ratio of 4:1. Although the sample size of this study is adequate, a factor analysis on a larger and overall different sample would be of interest.

These findings are important for ACs and particularly DACs because they afford a better understanding of the mechanism underlying the rating process. Additionally, these findings are consistent with other literature on ACs in suggesting that dimensions within ACs do exist, but AC designers must recognize the value of structuring the rating process to enable raters to accurately observe these dimensions.

REFERENCES

Arthur, W., Jr., Day E. A., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology, 56*, 125-54.

Arthur, W., Jr., Woehr, D. J., & Maldegen, R. (2000). Convergent and discriminant validity of assessment center dimensions: A conceptual and empirical reexamination of the assessment center construct-related validity paradox. *Journal of Management, 26*, 813-835.

Ballantyne, I., & Povah, N. (2004). *Assessment and development centres* (2nd ed). Aldershot, England: Gower.

Bray, D. W. (1995). Centered on assessment. *Personnel Psychology, 48*, 468-471.

Carrick, P., & Williams, R. (1999). Development centres - A review of assumptions. *Human Resource Management Journal, 9*, 77-92.

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

Eurich, T. L., Krause, D. E., Cigularov, K., & Thornton, G. C. III (2009). Assessment centers: Current practices in the United States. *Journal of Business and Psychology, 24*, 387-407.

Fleenor, J. W. (1996). Constructs and development assessment centers: Further troubling empirical findings. *Journal of Business and Psychology, 10*, 319-335.

Howard, A. (1997). A reassessment of assessment centers: Challenges for the 21st century. *Journal of Social Behavior and Personality, 12*, 13-52.

International Task Force on Assessment Center Guidelines. (2000). Guidelines and ethical considerations for assessment center operations. *Public Personnel Management, 29*, 315-331.

Kello, J. (2007). Assessment center. In S. G. Rogelberg (Ed.), *The encyclopedia of industrial and organizational psychology*. Thousand Oaks, CA: Sage.

Kudisch, J. D., Avis, J. M., Fallon, J. D., Thibodeaux, H. F., Roberts, F. E., Rollier, T. J., & Rotolo, C. T. (2001, April). *A survey of assessment center practices in organizations worldwide: Maximizing innovation or business as usual?* Paper presented at the 16th annual conference for the Society of Industrial Organizational Psychology, San Diego, CA.

Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria: What did they really say? *Organizational Research Methods, 9*(2), 202-220.

Lance, C. E., Lambert, T. A., Gewin, A. G., Lievens, F., & Conway, J. M. (2004). Revised estimates of dimension and exercise variance components in assessment center postexercise dimension ratings. *Journal of Applied Psychology, 89*, 377-385.

LeBreton, J. M., Burgess, J.R.D., Kaiser, R. B., Atchley, E.K.P., & James, L. R. (2003). The restriction of variance hypothesis and interrater reliability and agreement: Are ratings from multiple sources really dissimilar? *Organizational Research Methods, 6*(1), 80-128.

LeBreton, J. M. & Senter, J. L. (2007). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods, 11*(4), 815-852.

Macan, T. H., Avedon, M. J., Paese, M., & Smith, D. E. (1994). The effects of applicants' reactions to cognitive ability tests and an assessment center. *Personnel Psychology, 47*, 715-738.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*, 84-99.

Meriac, J. P., Hoffman, B. J., Woehr, D. J., & Fleisher, M. S. (2008). Further evidence for the validity of the assessment center dimensions: A meta-analysis of the incremental criterion-related validity of dimension ratings. *Journal of Applied Psychology, 93*, 1042-1052.

Office of Strategic Services Assessment Staff. (1948). *Assessment of men: Selection of personnel for the Office of Strategic Services*. New York: Rinehart & Co.

Robie, C., Osburn, H. G., Morris, M. A. & Etchegaray, J. M. (2000). Effects of rating process on the construct validity of assessment center dimension evaluations. *Human Performance, 13,* 355-370.

Rupp, D. E., Snyder, L. A., Gibbons, A. M., & Thornton, G. C., III. (2006). What should developmental assessment centers be developing? *The Psychologist-Manager Journal, 9*, 75-98.

Sackett, P. R. & Dreher, G. F. (1982) Constructs and assessment center dimensions: Some troubling empirical findings. *Journal of Applied Psychology, 67*, 401-410.

Spychalski, A. C., Quinones, M. A., Gaugler, B. B., & Pohley, K. (1997). A survey of assessment center practices in organizations in the United States. *Personnel Psychology, 50*, 71-90.

Silverman, W. H., Dalessio, A., Woods, S. B., & Johnson, R. L., Jr. (1986). Influence of assessment center methods on assessors' ratings. *Personnel Psychology, 39*, 565-578.

Thornton, G. C., III, & Rupp, D. E. (2003). Simulations and assessment centers. In J. C. Thomas (Ed.) & M. Hersen (Series Ed.), *Comprehensive handbook of psychological assessment: Vol. 4. Industrial and organizational assessment* (pp. 319-344). Hoboken, NJ: Wiley.

Thornton, G. C., III, & Rupp, D. E. (2006). *Assessment centers in human resource management: Strategies for prediction, diagnosis, and development.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Thornton, G. C., III, & Krause, D. E. (2009). Selection versus development assessment centers: An international survey of design, execution, and evaluation. *The International Journal of Human Resource Management, 20*, 478-498.

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