1-2005

Assessment centers: Recent developments in practice and research

Filip LIEVENS
Singapore Management University, filiplievens@smu.edu.sg

George C. THORNTON III
DOI: https://doi.org/10.1002/9781405164221.ch11

Assessment Centers: Recent Developments in Practice and Research

Filip Lievens

Ghent University

George C. Thornton III

Colorado State University

Address of correspondence: Filip Lievens, Department of Personnel Management and Work and Organizational Psychology, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium (filip.lievens@ugent.be).

Lievens, F., & Thornton, G. C. III (2005). Assessment centers: recent developments in practice and research. In A. Evers, O. Smit-Voskuijl, & N. Anderson (Eds.) *Handbook of Selection* (pp. 243-264). Blackwell Publishing.

Bio of authors

Filip Lievens is associate professor at the Department of Personnel Management and Work and Organizational Psychology at Ghent University, Belgium. In 1999, he earned his Ph.D. from the same university. His current research interests focus on alternative selection procedures (e.g., assessment centers, situational judgment tests, web-based assessment) and organizational attractiveness. He is the author of over 30 articles and has published among others in the Journal of Applied Psychology, Personnel Psychology, Journal of Organizational Behavior, Journal of Occupational and Organizational Psychology, Applied Psychology: An international Review, and International Journal of Selection and Assessment.

George Thornton is professor of industrial and organizational psychology in the Department of Psychology at Colorado State University. He earned his Ph.D. from Purdue University in 1966. His current research interests include the effectiveness of developmental assessment centers for managers and students, the role of industrial psychology in employment discrimination litigation, and the cross-cultural study of achievement motivation to work. He is the author of three books, several chapters, and numerous articles and conference presentations on assessment center methods. He has helped both public and private organizations develop assessment centers. He has lectured on the assessment center method in numerous countries in Europe and Asia.

Assessment Centers: Recent Developments in Practice and Research

In an assessment center, candidates who participate in various simulation exercises are evaluated by a multiple trained assessors on job-related dimensions. Examples of commonly used simulation exercises are role-plays, presentations, in-baskets, or group discussions. For nearly fifty years, assessment centers have remained a popular approach for managerial selection and development (Spychalski, Quinones, Gaugler, & Pohley, 1997), and they have been shown to have substantial validity (Thornton & Rupp, 2004). Assessment centers are also very much an international affair as they are used around the globe (Byham, 2001; Kudisch, Avis, Fallon, Thibodeaux, Roberts, Rollier, & Rotolo, 2001; Sarges, 2001). Over the last years, several innovative trends have emerged in assessment center practice. At the same time, various scholars (e.g., Arthur, Woehr, & Maldegen, 2000; Arthur, Day, McNelly & Edens, 2003; Haaland & Christiansen, 2002; Kolk, Born, & Van der Flier, 2002; Lance, Newbolt, Gatewood, Foster, French, & Smith, 2000; Lievens & Conway, 2001; Lievens & Klimoski, 2001) have given a new impetus to assessment center research. The aim of this chapter is to inform both practitioners and researchers of these recent intriguing developments. In particular, we focus on developments in assessment center practice and research that occurred between the last five years (1998-2003).

Recent Developments in Assessment Center Practice

In this section, we give an overview of recent developments in assessment center practice. To identify these recent developments we examined assessment center operations described in research studies, surveys of practitioners around the world, presentations at the International Congress on Assessment Center Methods, and innovations we learned about from colleagues around the world in the last years.

The following developments in assessment center practice are described: assessment centers for non-managerial jobs, assessment centers in cross-cultural settings, new methods of

analyzing job requirements, assessment of new dimensions, types of exercises, use of technology and virtual assessment centers, integrating assessment centers with organizational strategy, using assessment centers for developmental purposes, and assessment centers as criterion measures. If relevant, we begin our discussion of each trend by outlining the changes in business practices and organizations that have triggered the trend.

Assessing Non-managerial Jobs

Historically, assessment centers have been applied most frequently to managerial jobs ranging from supervisor to executive. More recently, they have been used to assess a wider range of non-managerial jobs. For many years, Diamond Star Motors has used an assessment center process to select manufacturing employees (Henry, 1988). This practice has spread to other manufacturing organizations such as Cessna (Hiatt, 2000) and BASF (Howard, L. & McNelly, 2000). The State of Connecticut has used assessment center methods to certify the competence of teachers (Jacobson, 2000). Other organizations have used assessment centers to select entry-level police officers (Dayan, Kasten, & Fox, 2002) and airline pilots (Damitz, Manzey, Kleinmann, & Severin, 2003) and to assess and certify consultants (Howard, A. & Metzger, 2002; Rupp & Thornton, 2003) and lawyers (Sackett, 1998).

These examples demonstrate the applicability of assessment center principles to a wide range of jobs. As noted below, the good news is that recent research has found evidence for the validity of some of these assessment centers in non-managerial populations.

Applying Assessment Centers in Multi-national and Cross-cultural Settings

The emergence of global businesses has increased the need to design assessment centers that have cross-cultural and cross-national applicability. Assessment centers have been implemented in an ever-increasing variety of countries around the world. Soon after their inception in England and the US in the 1950s, assessment centers spread to selected organizations in Canada and Japan. Next came extension to Germany, Switzerland, Israel,

South Africa, and Indonesia in the 1970s. But, it was not until the last several years that assessment centers cropped up in virtually every industrialized country in the world.

The internationalization of assessment centers is also revealed by examination of the lists of persons attending the International Congress on Assessment Center Methods over the past 31 years. There has been a steady increase in the number and percentage of attendees coming from countries outside North America. In 1974 at the second Congress, only 5 of 76 attendees were from countries other than the US and Canada. The percentage of attendees from outside North America grew steadily: 1974 to 1983 - approximately 5%; 1984 to 1993 - approximately 15%; 1994 to 2003 - approximately 25%. In 2003, 27 of the 104 attendees were from such diverse countries as Kuwait, Indonesia, Korea, and the Philippines.

Many challenging issues about the design and implementation of assessment centers arise when they are used in cross-cultural situations. Two approaches to these issues can be considered: etic and emic (Chawla & Cronshaw, 2002). The etic approach assumes that a) there are universal individual attributes that are relevant to organizational effectiveness, b) pre-existing assessment techniques can be adapted in different countries, c) standardization and validity extension require that a fixed set of dimensions and procedures must be used, d) and the adoption of uniform selection procedures across cultures contributes to a homogeneous organizational culture. The emic approach assumes that a) generic assessment methods will be invalid (i.e., they under-specify unique aspects of criterion performance), b) each culture must be studied to identify its unique features, c) the acceptance of various assessment techniques varies across cultures, and d) assessor training must include an appreciation of contextual information. An unresolved issue in this discussion is the relative gains and losses that come from modifying assessment center elements. For example, modification of the exercises may accommodate unique local demands, but render comparisons of assessments across locations problematic.

The spread of assessment centers around the world, the cross cultural applications of assessment centers, the globalization of businesses, the need for global executives (McCall, 2002), and the establishment of consultancies offering assessment center services in many other countries have raised questions about the application of assessment practices in diverse countries. Are assessment centers useful in selecting persons from a home country to serve in another country? Along these lines, Briscoe (1997) suggested that careful attention should be paid to the design of other exercises, the use of different dimensions, the use of assessors from both the home and host country, the evaluation of behaviors, and the provision of feedback. Briscoe (1997) and Howard (1997) also provided case studies that illustrated some of the challenges in using assessment centers to select international personnel. Kozloff (2003) discussed some of the complex issues of selecting leaders, along with their spouses and families, to live and work in diverse settings around the world. In the only empirical, predictive validity study on this topic of which we are aware, Lievens, Harris, Van Keer, and Bisqueret (2003) found that ratings from selected assessment exercises contributed to predictive accuracy over cognitive ability and personality tests in predicting success in a training program for European managers in Japan.

Another challenging question is whether the *Guidelines* apply universally around the world. Over the years there has been a trend to consider international issues. The task forces who wrote the 1975 and 1979 editions of the *Guidelines and Ethical Considerations for Assessment Center Operations (International Task Force on Assessment Center Guidelines, 2000)* contained only North American practitioners. The 1989 and 2000 task forces each included one Dutch representative, and there was increased input sought from practitioners from outside North America. In 2001, a group of practitioners in Europe met to consider whether the *Guidelines* needed revision for application in those countries (Seegers & Huck,

2001). In 2000, a task force was set up in Indonesia to write a code of conduct for assessment center operations for that country (Pendit & Thornton, 2001).

We predict that assessment centers will be used more frequently in international settings. This will occur in three different ways. Home-country organizations will use assessment centers to assess persons going to host-countries. Home-country organizations will use their assessment methods to assess host-country persons in those other countries. Organizations in countries not currently using assessment centers will adopt the method. Each of these applications of the assessment center method presents unique challenges. Assessment center proponents and adopters will have to make choices about what elements and specific practices of the method can be kept the same from their points of origin to the new location, and what adaptations need to be made to accommodate the unique aspects of the new location.

Job Analysis Methods

Recently, the breakdown of rigid divisions of labor among jobs has led to the search for broader competencies to serve as the dimensions for assessment. Therefore, traditional methods of analyzing tasks involved in job accomplishment and the knowledge, skills, and other attributes needed for performance on specific current jobs have been supplemented by future-oriented methods such as strategic job analysis (Schneider & Konz, 1989) and by competency modeling to analyze more general competencies that organizations expect employees to possess in order to achieve broader organizational objectives (Schippman, Ash, Battista, Carr, Eyde, Hesketh, Kehoe, Pearlman, & Sanchez, 2000; see also the chapter on job analysis in this Handbook). These broader objectives may be translated into the roles that employees are expected to play (Mahoney-Philips, 2002; Sackett & Tuzinski, 2000).

In our opinion, the current techniques of competency modeling have several advantages of aligning performance in specific jobs to broader organizational objectives,

defining requirements of broader sets of jobs rather than isolated positions, and gaining acceptance from higher level managers and executives. Conversely, competencies are often defined so broadly as to defy reliable and valid assessment. Thus, there is often a clear need to develop techniques to translate competencies into performance dimensions that can be assessed with reasonable accuracy.

Different Dimensions

Traditionally, assessment centers have been designed to measure relatively specific sets of behaviors known as "dimensions" (Thornton & Byham, 1982). Recently assessment center architects began to assess broader competencies (e.g., customer service orientation, team work) each of which is often a complex combination of traditional dimensions. In our opinion, broad organizational competencies do not provide specific, objective attributes appropriate for assessment in assessment centers. For example, "customer service" and "continuous quality improvement" are worthy organizational goals, but they need operationalization into behavioral dimensions. The former can be specified as behaviors classified into active listening, information seeking, and oral communication, and the latter as behaviors such as problem analysis, creativity, and decision analysis.

Other trends include increased emphasis on the assessment of interpersonal dimensions such as team work, cooperation, and informal leadership. These sorts of broader dimensions are especially relevant to success in international settings. Kozloff (2003) describes the need to consider broader sets of personality factors (e.g., tolerance for ambiguity and emotional balance) and family relations when selecting global leaders. Some organizations are also devising ways to assess sets of values, using systematic techniques such as Systematic Multiple Level of Observation of Groups (SYMLOG) (Wilson & Pilgram, 2000). Other programs have assessed roles that employees are expected to play (Mahoney-Phillips, 2002). On a grander scale, the U.S. Office of Personnel Management has developed a

national framework of dimensions defining all jobs in the U.S. economy (Gowing, 1999). This taxonomy of jobs listed in the Standard Occupational Classification system provides a common language for the attributes needed for all jobs. Taking a quite different approach, some assessment center architects argue that no dimensions at all should be assessed (Thoreson, 2002), but rather that the performance of behavior in the exercise as a whole should be assessed.

In our estimation, virtually any performance attribute is amenable to assessment if two conditions are met: (1) the dimension is clearly defined in terms of behaviors on the job and behaviors observable in simulation exercises, and (2) the exercises are constructed carefully so as to elicit the relevant behaviors. The second condition implies that exercises can and should be constructed in different ways to assess different attributes. Techniques for constructing various types of exercises for various purposes and for eliciting behaviors relevant to different dimensions are described in *Developing Organizational Simulations* (Thornton & Mueller-Hanson, 2004).

Assessment Exercises

One might think that new types of exercises would have been invented to assess new dimensions for new jobs in new settings, but this does not seem to be the case. Most of the old standby types of exercises seem to persist, including in-baskets, case studies, and interaction simulations. There seems to be a trend toward lesser use of the group discussion technique. There may be three explanations. First, in police and fire departments where the assessments are used as one method in promotional examinations, there is a strong pressure for strict standardization that does not exist in the highly variable group dynamics that typically unfold in a leaderless group discussion. Second, there are practical, logistical problems of gathering all candidates at one specific location at one specific time. Designers often wish to have a process that does not require that a group of participants are at the same location at the same

time. Third, the complex interactions among 5 or 6 persons in a group discussion are often difficult to observe and thus defy the systematic observation and evaluation by novice assessors.

The elimination of group discussion exercises is understandable when the assessment center is being used for selection or promotion, and legal challenges to standardization are highly likely. However, in light of organizations' interest in assessing the fit of individuals to teams and organizations, the group exercise is one of the more content valid assessment exercises.

Use of Technology and Virtual Assessment Centers

The availability of computers and electronic media has provided the opportunity to increase the use of technology in assessment centers. Initially, computers were used to compile and analyze ratings from a team of assessors. Recently, more sophisticated applications have emerged, primarily in the methods to present stimuli. Exercise stimuli have been presented via video monitors and on computer-based simulations (Bobrow & Schulz, 2002). At Sprint, a virtual office has been simulated on the company's intranet for the administration of exercises (Hale, Jaffee, & Chapman, 1999). Reynolds (2003) described the movement toward web based delivery of exercises for the assessment of executives and leaders. Other applications of technology involve capturing behavior on audio and video recordings, sometimes from remote locations. These recordings can then be analyzed in a traditional manner by trained observers, or by using sophisticated software programs. Other assessment programs have used the web to capture electronic records of various achievements including text, audio, and video media (Richards, 2002). Automated analysis of written responses can evaluate the content and quality of writing samples (Ford, 2001). In addition, software can analyze voice tone (Bobrow & Schulz, 2002). Special software has been developed to automate the process of writing reports (Lovler & Goldsmith, 2002).

Furthermore, the web can be used to facilitate all stages of an assessment process including administration, exercise delivery, scoring, data tracking, report writing, and feedback (Smith & Reynolds, 2002). Reynolds (2003) traced the progression of technology applications toward a web services model of assessment using the manager's on-line desk to complete work in a simulated "day in the life" set of assessment activities.

A number of these technological developments increase the fidelity of exercise in terms of the stimuli presented to the participant (e.g., managers nowadays typically receive information via electronic media and respond online). Thus, in our estimation, high technology in an exercise may increase the realism of the exercise. Other aspects of high tech assessment exercises may in fact decrease the fidelity of the assessment, especially response fidelity. For example, some computerized in-baskets call for the participant to respond by choosing among a number of pre-established alternatives. In real-life, managers do not typically have the alternatives presented. In fact, they must generate alternatives and then overtly write a response. In some exercises a video depicts a subordinate's comments and the participant selects among a set of pre-established responses. This sort of assessment method does not have fidelity with dynamic interpersonal interactions. Computerized in-baskets and video-based assessment techniques may have predictive validity, but they are qualitatively different from the overt behaviors required in the typical interpersonal and decision making simulations that have been the hallmark of the assessment center method.

Integrating Assessment Centers with HR Management and Organizational Strategy

There has been increasing recognition in recent years that assessment centers must be integrated carefully with other human resource management practices and with the overall organizational strategy. Although this is not a new idea (Thornton, 1992), pressures to make all HRM practices more efficient have placed more emphasis on making assessment center practices more compatible with broader organizational strategies. Thus, we see a recent trend

to more systematically build the assessment center into a larger system of recruitment, selection, promotion, development, and succession planning for management talent (Byham, 2002; Byham, Smith, & Pease, 2001; Roth & Smith, 2000). This trend is also manifest in organizations operating global HR practices (Eckhardt, 2001).

The integration of assessment centers into broader organizational strategic planning and the use of assessment centers to foster organizational change is also apparent in recent applications. For example, assessment centers have been used to help achieve redeployment of existing staff (Adler, 1995), downsizing (Gebelein, Warrenfeltz, & Guinn, 1995), executive team development (Fleisch & Cohen, 1995), restructuring from functional to product focus (Fleisch, 1995), and climate change (Dailey, Cohen, & Lockwood, 1999) in such diverse organizations as manufacturing, telecommunication, trucking, customer service, high-tech, and security. The integration of assessment centers into organizational change efforts requires involvement of high-level executives in the program (Dowell & Elder, 2002).

Assessment Centers for Developmental Purposes

The most pronounced trend in assessment centers activities in recent years is the shift in their predominant purpose from selection/promotion to development. This increased interest in using assessment centers to develop the talent of managers remaining in their current positions results among other things from the flattening and downsizing of organizations and the fewer promotional opportunities available. The original purpose of assessment centers (i.e., identification of managerial talent and decision-making for promotion, Bray & Grant, 1966; Thornton & Byham, 1982) is still predominant in public safety organizations. Conversely, in most business organizations in recent years, the most frequent application is for developmental purposes (Kudisch, et al., 2001; Spychalski et al., 1997). A more skeptical view of this application was presented by Tillema (1998) who found

only minimal use of development centers in a survey of Dutch organizations because of lack of familiarity with and difficulties in implementation of this application.

There are several variants of developmental assessment centers. In some, the emphasis is on the diagnosis of training needs of individuals. The design of these centers, including the dimensions and exercises, is very similar to promotional centers. Another variant is a true *development* center in which the objective is to foster skill development (Ballantyne & Povah, 1995). To turn the program into a learning experience, steps are taken to provide immediate feedback, practice, reinforcement of learning, transfer of training, and follow-up developmental support in the organization. A third variant of developmental assessment centers are programs designed to promote development of organizational units. The use of simulation technology for development purposes typically involves the assessment of intact work groups participating in complex organization games (Thornton & Cleveland, 1990). The use of one assessment center program for dual purposes of selection and development is problematic (Arnold, 2002) and requires careful attention to factors beyond psychometric precision (e.g., motivated participants, clear feedback, supportive context) (Kudisch, Lundquist, & Smith, 2002).

Developmental assessment centers have become quite popular, but have met with numerous challenges. One of the primary challenges is the necessity to demonstrate adequate psychometric evidence of construct validity. As discussed in a later section of this paper, there is mixed evidence regarding the ability of assessors' ratings to demonstrate evidence of construct validity. The second challenge of development assessment centers is to provide evidence that the program has some impact on participants. Impact may take the form of a) intentions to take action to develop, b) engagement in some form of developmental experience, c) change of understanding of the performance dimensions, d) improvement in skills, e) change of behavior on the job, or f) improvement in organizational effectiveness.

Jones and Whitmore (1995) found that career advancement of assessed and non-assessed managers did not differ, except when the assessed managers engaged in developmental activities. Unfortunately, most managers do not follow up assessment center diagnoses with developmental activities (T. Byham, 2003). Only recently has research evidence begun to emerge that demonstrates the conditions under which developmental assessment centers are effective (Mauer, Eidson, Atchley, Kudisch, Poteet, Byham, & Wilkerson, 2003). Positive effects do not automatically ensue and are likely to occur only if there are a number of other support systems in place in the organization to help the assessee after the assessment center experience (Bernthal, Cook, & Smith, 2001).

Assessment Centers as Criterion Measures

Similar to work samples, assessment centers are increasingly used as criterion measures in studying various aspects of managerial and student performance. For example, Thomas, Dickson, and Bliese (2001) used an assessment center of leader effectiveness in a study of the role of values, motives, and personality among cadets. Barlay and York (2002) and Riggio, Mayes, and Schleicher (2003) used assessment centers to measure undergraduate student achievement. Recently, Atkins and Wood (2002) validated a 360-degree feedback program on the basis of assessment center ratings of the candidates.

The underlying rationale of the use of assessment centers as criterion measures is that they correspond closely to the job and therefore can be considered as miniaturized settings for observing job performance. Although this rationale makes sense, it is important to note that the criterion data obtained with assessment centers are also inherently different from more traditional job performance data (i.e., ratings). Assessment center performance reflects maximal performance, whereas job performance ratings represent typical performance.

Disturbing Trends

Two disturbing trends have been noted in the implementation of assessment centers in recent years. First, in response to the economic downturn in recent years (2000-2003), organizations have sought ways to streamline the process. Unfortunately, in many cases, modifications of essential steps in the development and implementation of programs have led to short cuts that may affect accuracy and effectiveness. Caldwell, Thornton, and Gruys (2003) summarize ten errors that diminish assessment center validity (e.g., inadequate job analysis, ill-defined dimensions, inadequate assessor training).

A second disturbing trend is that the term "assessment center" has been used to refer to many methods that do not conform to the essential elements of the assessment center method. Examples that in our opinion do not qualify as assessment centers are methods that include only paper and pencil tests, methods that involve only one assessor, and methods that do not involve the observation of overt behavior. Thus, even though they are valid, the following methods do not constitute an assessment center: computerized in-baskets that call for the participant to pick among a set of pre-defined alternative behaviors; situational interviews that ask the respondent to state what he or she would do when faced with hypothetical situations, written "low fidelity" simulations or situational judgment tests that call for choosing among alternative actions; clinical or individual psychological assessments that are carried out by one assessor.

While there is no legal restriction, patent registration, or copyright proprietary claim for the words "assessment center," there are strong reasons for wishing to restrict the use of the term. First, for over 50 years the term has been used in the personnel assessment profession to refer to a common set of practices. Second, extensive research has been conducted on the method, and while there are certainly different instantiations of many elements of the method, there are enough commonalities to the claim that a coherent body of

research exists. Summaries of that research have led to several meta-analyses of validity findings and comparisons with alternative assessment techniques. Such comparative studies are not meaningful if the alternative techniques cannot be clearly defined and classified. Third, for over 25 years, the International Congress on Assessment Center Methods has attracted hundreds of participants who have a common interest in the design, implementation, and evaluation of this commonly understood method. *Guidelines and Ethical Considerations for Assessment Center Operations* (International Task Force on Assessment Center Guidelines, 2000) clearly defines what is and what is not an assessment center. It provides a standard for students, practitioners, and researchers to follow.

Recent Developments in Assessment Center Research

Whereas the first part of this chapter focused on recent developments in assessment center practice, this part delves into recent assessment center research. Inspection of assessment center research published between 1998 and 2003 showed that the vast majority of studies could be grouped under the following four broad themes: criterion-related validity research, incremental validity research, construct validity research, and process-related research. While these are recurring themes in the assessment center literature, recent research has often given a new twist to them.

Assessment Centers and Criterion-Related Validity

Over the last five years, further support for the criterion-related validity of assessment centers has been gathered. One set of studies extends validity evidence of the overall assessment rating, and another set of studies extends validity evidence of dimension ratings. Specifically, recent studies have provided evidence that the criterion-related validity of assessment centers holds across jobs, time, and contexts. With respect to jobs, two studies were most noteworthy. Damitz et al. (2003) broadened existing selection procedures for selecting airline pilots by including various assessment center exercises to assess both

interpersonal and performance-related dimensions. The overall assessment rating was a valid predictor of peer criterion ratings. Similarly, Dayan et al. (2002) argued that assessment centers can be a vital tool to capture the interpersonally-oriented dimensions of police work. Their assertion was supported among Israeli police force candidates using both supervisory and peer ratings as criteria. Other studies have also confirmed the relevance of assessment centers for student selection (Bartels, Bommer, & Rubin, 2000; Riggio et al. 2003).

With respect to the validity of assessment centers in the long run, Jansen and Stoop (2001) validated an assessment center over a 7-year period with average salary growth as the criterion. The corrected validity of the overall assessment rating was .39. An interesting contribution of Jansen and Stoop was that they also examined how the validity of assessment center dimensions changed over time. They found that the firmness dimension was predictive over the whole period, whereas the interpersonal dimension became valid only after some years. The latter finding is consistent with research showing that noncognitive predictors become more important when the criterion data are gathered later on (Goldstein, Zedeck, & Goldstein, 2002).

In recent years, there has also been some evidence that assessment centers can be used in contexts other than domestic selection. Stahl (2000) developed an assessment center for selecting German expatriates. Although the criterion-related validity was not examined, Stahl found that candidates scoring high on different criteria of intercultural competence were also appraised by their peers as being more adaptable to a foreign environment. Lievens et al. (2003) developed and validated an assessment center for selecting European managers for a cross-cultural training program in Japan. Besides assessment center exercises, the procedure included cognitive ability and personality tests and a behavior description interview. The dimensions of adaptability, team work, and communication as measured by a group discussion exercise emerged as valid predictors, beyond cognitive ability and personality

tests. Dimensions measured in a presentation did not emerge as significant predictors, showing that exercise design is an important issue in assessment centers for international applications (see above).

Finally, Arthur et al. (2003) conducted a meta-analysis of the criterion-related validity of assessment center dimensions. They distinguished six meta dimensions: (1) consideration/awareness of others, (2) communication, (3) drive, (4) influencing others, (5) organizing and planning, and (6) problem solving. True criterion-related validities varied from .25 to .39. Moreover, a regression-based composite consisting of four out of the six dimensions accounted for the criterion-related validity of assessment center ratings and explained somewhat more variance in performance than the prior meta-analysis of Gaugler, Rosenthal, Thornton, and Bentson (1987). The assessment center dimensions yielded a multiple correlation of .45 (R^2 = .20). Thus, a focus on assessment center constructs (dimensions) instead of on the overall assessment rating seems to increase the predictiveness of assessment centers.

In summary, these last five years, recent studies have found evidence that assessment center validities hold across a wide range of jobs, over longer time periods, and in international contexts. In addition, a recent meta-analysis has further supported the criterion-related validity of assessment centers. An important new finding was that assessment centers have higher predictive validity when they are not seen as a monolithic entity (cf. the overall assessment rating) but as a measure to provide information on various constructs (cf., assessment center dimensions). Despite this positive news, an intriguing finding is that the validity of assessment centers is not higher than the validity of less expensive predictors such as highly-structured interviews or situational judgment tests. Two methodological issues might explain this. First, prior assessment center meta-analyses used values of .77 (Gaugler et al., 1987, p .496) and .76 (Arthur et al., 2002, p. 153) respectively to correct for criterion

unreliability. Hence, these values are much higher than the .52 inter-rater reliability value of job performance ratings that has typically been used in recent meta-analyses of selection procedures (e.g., structured interviews) (Viswesvaran, Ones, & Schmidt, 1996). Since prior meta-analyses used such conservative estimates for correcting criterion unreliability, their corrected values underestimate the "true" validity of assessment centers. For example, if we correct the validity coefficient of Gaugler et al. (1987) with the usual .52 value instead of the more conservative values, the corrected validity of assessment centers rises to .45 instead of to .37. In a similar vein, the corrected validity coefficient of .45 of Arthur et al. (2002) would be higher when corrected with the usual .52 value. Another key methodological issue when interpreting assessment center validities relates to range restriction in the KSAOs measured. Typically, assessment centers are used in final selection stages so that assessment center candidates have been screened in prior selection stages on the basis of both cognitive ability and personality. Consequently, the variance in terms of both cognitively-oriented and interpersonally-oriented competencies among assessment center candidates is more limited, leading to a possible decrease in predictive validity (Hardison & Sackett, 2004). Future research should put all of this to the test

Assessment Centers and Incremental Validity

Despite the widespread agreement that assessment centers have strong predictive validity, there is more debate as to whether assessment centers have incremental validity over and above traditional selection procedures such as cognitive ability and personality tests. A meta-analysis of Collins, Schmidt, Sanchez-Ku, Thomas, McDaniel, and Le (2003) found that the multiple correlation of personality and cognitive ability tests with overall assessment center ratings was .84. In addition, Schmidt and Hunter (1998) reported that assessment centers had a small incremental validity (2%) when combined with cognitive ability tests. A recent study (Dayan et al., 2002), however, found the opposite results as assessment centers

had significant unique validities beyond cognitive ability tests. In addition, O'Connell, Hattrup, Doverspike, and Cober (2002) found that role play simulations added incremental validity over biodata in predicting retail sales performance.

How can these conflicting findings be reconciled? First, it should be noted that the assessment centers included in the two aforementioned large-scale reviews (Collins et al., 2003; Schmidt & Hunter, 1998) often incorporated cognitive ability and personality tests. Thus, the overall assessment ratings were partially based on information from cognitive ability and personality tests. Given this contamination, it is less surprising that assessment centers did not explain much additional variance over cognitive ability and personality tests. Second, both large-scale studies focused on the overall assessment rating. Although the overall assessment rating is of great practical importance (hiring decisions are contingent upon it), it is a summary rating of evaluations on a variety of dimensions in a diverse set of exercises (Howard, 1997). The fact that the overall assessment rating is such an amalgam of various ratings may reduce its conceptual value. Arthur et al. (2003) cogently argued that assessment centers are best conceptualized as a method for measuring a variety of constructs. Therefore, it makes little sense to state that assessment centers *per se* measure cognitive ability. Instead, depending on the job-related constructs measured, assessment centers might (or might not) have strong correlations with cognitive ability. For instance, if assessment center exercises (in-baskets, case-analyses) primarily measure cognitively-oriented dimensions, strong correlations with cognitive ability tests are to be expected. If this is not the case, correlations with cognitive ability tests will be lower. In support of this, Goldstein, Yusko, Braverman, Smith, and Chung (1998) reported that the relationship between assessment centers and cognitive ability tests varied as a function of the cognitive "loading" of assessment center exercises. When exercises (e.g., in-basket exercise) tapped cognitively-

oriented dimensions (e.g., problem analysis), there were stronger relationships between the exercise and the cognitive ability test (see also Goldstein, Yusko, & Nicolopoulos, 2001).

In a similar vein, the relationship between an overall assessment rating and personality tests will differ according to the job-related constructs measured in assessment center exercises. Various recent studies (Craik, Ware, Kamp, O'Reilly, Staw, & Zedeck, 2002; Lievens, De Fruyt, & Van Dam, 2001; Spector, Schneider, Vance, & Hezlett, 2000) support this reasoning. For instance, Spector et al. (2000) discovered that "interpersonal" exercises correlated with personality constructs such as Emotional stability, Extraversion, and Openness and that "problem-solving" exercises correlated with cognitive ability and Conscientiousness. In another study, Craik et al. (2002) reported that in-basket performance was related to Conscientiousness, Openness, and strategic dimensions such as decision making. Conversely, group discussion performance was best described by interpersonal dimensions and personality constructs such as Agreeableness, Extraversion, and Openness. Finally, Lievens et al. (2001) linked the personality and assessment center domains by scrutinizing the notes of assessors for personality-descriptive adjectives and by classifying them according to the Big Five. Again, results revealed that the distribution of the Big Five categories varied across exercises. For example, the in-basket elicited most frequently Conscientiousness descriptors, whereas the group discussion was characterized by many Extraversion descriptors.

In recent years, assessment centers have been challenged not only by personality and cognitive ability tests, but also by other assessment methods. In particular, situational judgment tests, situational interviews, and behavior description interviews have gained in popularity because they are easy to administer, good predictors of job performance, and not very expensive. Therefore, an important question is whether assessment centers have incremental validity over them. So far, research seems to support the continued use of assessment centers. In fact, Lievens et al. (2003) showed that dimensions measured by an

assessment center had incremental validity over dimensions assessed in a behavior description interview for predicting cross-cultural training performance. Further, Harel, Arditi, and Janz (2003) reported that the validity of a behavior description interview was .53, whereas the assessment center's validity was .62.

In summary, recent studies have scrutinized the incremental validity of assessment centers over traditional selection procedures (personality and cognitive ability tests) and emerging ones (behavioral description interviews). Unfortunately, only a few studies have been conducted so far. A drawback of most incremental validity studies is that they confound methods (e.g., assessment centers, interviews, tests) with constructs (e.g., Conscientiousness, sociability). For example, the validity of two constructs (cognitive ability and personality) was typically compared to the validity of a method (assessment center). As already noted, these comparisons are not meaningful unless one either holds the constructs constant and varies the method, or holds the method constant and varies the content (Arthur et al., 2003). For instance, future studies should examine whether sociability as measured by an assessment center exercise has incremental validity over sociability as measured by a personality inventory or situational interview.

Assessment Centers and Constructs Measured

Generally, two analytical methods have been used for examining assessment center construct validity. First, final dimension ratings have been placed in a nomological network to investigate their relationships with similar constructs measured by other methods such as tests, interviews, etc. As described above, assessment center ratings have been found to correlate with the same or similar dimensions assessed by other methods. As a second analytical strategy, dimensional ratings made per exercise (i.e., within-exercise dimension ratings) have been cast as a multitrait-multimethod matrix in which dimensions serve as traits and exercises as methods. The general conclusion from the latter strategy has been that ratings

on the same dimensions across exercises correlate lowly (i.e., low convergent validity), whereas ratings on different dimensions in a single exercise correlate highly (i.e., low discriminant validity, or method bias). This has lead to the debate whether assessment centers actually measure the dimensions that they purport to measure. This is not to say that assessment centers do not have construct validity. Rather than questioning *if* there are constructs measured, the issue is *what* constructs are measured (Lievens & Klimoski, 2001; Sackett & Tuzinski, 2001).

Over the last five years, research on this theme has expanded (see Hoeft & Schuler, 2001; Lievens, 1998; Lievens & Conway, 2001; Sackett & Tuzinski, 2001; Woehr & Arthur, 2003, for reviews). Researchers have tried to unravel why the aforementioned construct validity results are found. Although the debate is still ongoing, current thinking seems to be that at least three factors are responsible.

First, poorly designed assessment centers seem to show less construct validity evidence. To examine the effects of assessment center design, Lievens and Conway (2001) reanalyzed a large number of studies. They reported significantly more evidence of construct validity when fewer dimensions were used and when assessors were psychologists. Use of behavioral checklists, a lower dimension-exercise ratio, and similar exercises also increased dimension variance. Recently, Woehr and Arthur (2003) confirmed the influence of many of these design considerations. These two large-scale studies demonstrate that assessment center design is important and matters. Therefore, we are generally enthusiastic regarding this body of research. Yet, a caveat is in order. It is important to consider which design recommendations are artificial and which are not. For instance, asking assessors to integrate behavior observations and ratings for each dimension across all exercises prior to evaluating the subsequent dimensions (see Arthur et al., 2000) might be stretching design changes too

far. When assessors are first required to look at consistency of candidates across exercises, one might artificially inflate the correlations of the dimensions across exercises.

As a second factor affecting construct validity evidence, there should be high inter-rater reliability among assessors. If inter-reliability is at best moderate, variance due to assessors will be necessarily confounded with variance due to exercises because assessors typically rotate through the various exercises (they do not rate candidates in all exercises). Due to this confounding, part of the large exercise variance observed in construct validity studies of operational centers might be assessor variance (Howard, 1997). To examine this, two recent studies (Kolk et al., 2002; Robie, Adams, Osburn, Morris, & Etchegaray, 2000) compared construct validity evidence when assessors rated all dimensions in a single exercise (as is often the case in practice) to construct validity evidence when an assessor rated only a single dimension across exercises. Construct validity evidence increased with the latter method. Although having one assessor per dimension may not be practically feasible, these studies do indicate that the large exercise variance typically found may at least partly be due to rating variability across assessors.

Recent studies have further revealed that the aforementioned factors (i.e., careful design and assessor reliability) might be necessary but insufficient conditions for establishing construct validity. Specifically, two studies (Lance et al., 2000; Lievens, 2002) identified the nature of candidate performances as a third key factor. Lance et al. examined whether exercise variance represented bias or true cross-situational performance differences. They correlated latent exercise factors with external correlates such as cognitive ability measures and concluded that exercise factors captured true variance instead of bias. Apparently, assessors provide relatively accurate assessments of candidates. These candidates, however, do not show performance consistency across exercises. Lievens (2002) reached similar conclusions, showing that convergent and discriminant validity evidence could be established

only for candidates whose performances varied across dimensions and were relatively consistent across exercises. This suggests that assessors are capable of detecting performance differences on dimensions, when these differences truly exist.

Now that we know that candidate performances affect construct validity evidence, the next question becomes what makes candidates perform differently across exercises. To answer this question, recent studies have built on interactionist models in social psychology. In particular, Tett and Guterman (2000) used the principle of trait activation (Tett & Burnett, 2003) to emphasize how the behavioral expression of a trait requires arousal by trait-relevant situational cues (i.e., exercise demands). On the basis of this interactionist approach, Tett and Guterman (2000) and Haaland and Christiansen (2002) showed that cross-exercise consistency in assessor ratings is found only when exercises share trait-expressive opportunities.

In sum, in recent years substantial advancements have been made to unravel the puzzle of assessment center construct validity. We have better insight in the factors contributing to the typically low construct validity of operational assessment centers when an internal validation strategy is used. These findings seem to result from a combination of poor assessment center design, moderate inter-rater reliability, and inconsistent and undifferentiated performance levels of candidates. To shed further light on this issue, future research might especially benefit from using interactionist models in social psychology (e.g., trait activation). We also believe that trait activation theory might serve to not only understand what is happening in assessment centers, but also be useful as a prescriptive framework to modify assessment center design (e.g., design of exercise-dimension matrix, role-player instructions).

Assessment Centers and Process-related Research

In the last five years, researchers have also shown a renewed interest in the assessment center process. A first group of studies has examined potentially biasing factors in this process. In particular, researchers have explored whether assessors' judgments are prone to effects related to repeated assessee participation (Kelbetz & Schuler, 2002), exercise order (Bycio & Zoogah, 2002), assessee impression management (Kuptsch, Kleinmann, & Köller, 1998; McFarland, Ryan, & Kriska, 2003), and assessor-assessee acquaintance (Moser, Schuler, & Funke, 1999). Many of these potentially biasing factors exerted relatively minor effects. For example, Bycio and Zoogah (2002) found that the order wherein candidates participated in exercises explained only about 1% of the rating variance. Kelbetz and Schuler (2003) reported that prior assessment center experience explained no more than 3% of the variance of the overall assessment rating. Generally, repeated participation in an assessment center provided candidates with a gain equivalent to an effect size of .40. McFarland et al. (2003) found less use of candidate impression management tactics in an assessment center exercise (a role-play) than in a situational interview. Apparently, candidates are already so busy acting out their designated role-play character that they have little cognitive resources left to engage in impression management. Whereas the aforementioned studies found only minor effects, Moser et al. (1999) found a large effect of assessor-assessee acquaintance. When acquaintance between the candidate and the assessor was less than or equal to two years, the criterion-related validity was .09. This value increased dramatically to .50 when assessor-assessee acquaintance was greater than two years. Although there might be drawbacks in terms of fairness, we believe that assessor-assessee acquaintance is not always bad. It might be beneficial in assessment centers for developmental purposes. To facilitate follow-up developmental actions, the best "assessor" might well be the participant's boss. In

fact, this process is followed by a branch of Suisse Credit Bank in Italy (D. Hippendorf, personal communication, October 7, 1999).

Another group of process-related studies has confirmed the importance of assessor type (psychologist versus manager). Specifically, Lievens (2001a, 2001b) found that managers had more difficulty in discriminating among dimensions than psychology student assessors. However, managerial assessors also rated candidates with higher accuracy. Other studies found that psychologists outperformed nonpsychologists only when the criterion-related validity of the interpersonal ratings made was examined (r = .24 versus r = .09) (Damitz et al., 2003) and that experienced assessors yielded significantly higher accuracy than inexperienced assessors (Kolk, Born, Van der Flier, & Olman, 2002). As a whole, these studies have shown that both types of assessors have their strengths and weaknesses. Therefore, it seems recommendable to continue the common practice of including a mix of experienced line managers and psychologists in the assessor team.

Third, recent studies have examined how assessors' observation and evaluation task can be facilitated. An obvious intervention consists of providing assessors with better training. There seems to be some evidence that especially schema-driven training might be beneficial in terms of increasing inter-rater reliability, dimension differentiation, differential accuracy, and even criterion-related validity (Goodstone & Lopez, 2001; Lievens, 2001a; Schleicher, Day, Mayes, & Riggio, 2002). Schema-driven training (frame-of-reference training) teaches raters to use a specific performance theory as a mental scheme to 'scan' the behavioral stream for relevant incidents and to place these incidents –as they are observed- in performance categories. Such a training seems to be a useful complement to the traditional data-driven training that teaches assessors to strictly distinguish various rating phases (observation, classification, and evaluation) and to proceed to another phase, only when the previous one is finished.

Other researchers have explored whether modifications to the existing observation and evaluation procedures might yield beneficial effects. Hennessy, Mabey, and Warr (1998) compared three observation procedures: note taking, behavioral checklists, and behavioral coding. The methods yielded similar outcomes in terms of accuracy, halo, and attitude toward the method, with a preference for behavioral coding. Kolk et al. (2002) found no positive effects of asking assessors to postpone note-taking until immediately after the exercise on accuracy, inter-rater reliability, or halo.

In summary, in the last five years, research on the assessment center process has revealed valuable findings. Specifically, the importance of the type of assessor has been corroborated. Furthermore, frame-of-reference training has emerged as one of the best assessor training strategies. Results on different observation formats has not yielded beneficial effects. Although the studies reviewed have advanced our understanding of the assessment center process, they also constitute only the proverbial tip of the iceberg. Few studies have actually profited from current thinking in person perception, social information processing, interpersonal judgments, and decision making. More specifically, examples of interesting research avenues might involve the roles of social judgment accuracy, assessor expectancies, cognitive structures, motivated cognition, and accountability in assessor judgments (Lievens & Klimoski, 2001).

Epilogue

The assessment center method continues to be used in a variety of organizational settings and to generate numerous research studies. In recent years, assessment centers have been used for a variety of purposes with an increasingly diverse set of jobs in countries around the world. Developments in assessment center practice in the past few years include new dimensions being assessed with innovations in assessment methods employing computer and web-based technology. Although these are often innovative applications, it is unfortunate

that systematic research about their validity and utility in comparison with established practices is typically lacking.

Developments in research include innovative studies regarding the criterion-related validity of assessment centers and regarding the unique contribution of assessment centers over alternative assessment procedures. Recent studies have also increased our understanding of the construct validity issue. Specifically, research identified that poor assessment center design, assessor unreliability, and lack of performance variability all contribute to poor measurement of constructs in assessment centers. Finally, process-related studies on assessment centers have emphasized the criticality of type of assessor and type of assessor training.

Additional research is needed to demonstrate the conditions under which developmental assessment centers have impact. Evidence is sorely lacking to demonstrate that participants take some follow up action in response to developmental feedback, show changes in behavior on the job, to contribute to increasing levels of individual and organizational effectiveness. Initial research has demonstrated some of the individual characteristics and organizational support mechanisms that contribute to the positive impact of developmental assessment centers, but more studies in these areas are needed.

References

Adler, S. (1995, May). *Using assessment for strategic organization change.* Paper presented at the 23rd International Congress on the Assessment Center Method, Kansas City, MO.

Arnold, J. (2002). Tensions between assessment, grading and development in development centres: A Case Study. *International Journal of Human Resource Management, 13,* 975-991.

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

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,* 125-154.

Atkins, P.W.B., & Wood, R.E. (2002). Self- versus others' ratings as predictors of assessment center ratings: Validation evidence for 360-degree feedback programs. *Personnel Psychology, 55,* 871-904.

Ballantyne, I & Povah, N. (1995). *Assessment and development centres.* Aldershot: Gower.

Barlay, L.A., & York, K.M. (2002, October). *Assessment centers for program evaluation: Assessing student academic achievement.* Paper presented at the 30th International Congress on Assessment Center Methods. Pittsburgh, PA.

Bartels, L.K., Bommer, W.H., & Rubin, R.S. (2000). Student performance: Assessment centers versus traditional classroom evaluation techniques. *Journal of Education for Business, 75,* 198-201.

Bernthal, P., Cook, K., & Smith, A. (2001). Needs and outcomes in an executive development program. *Journal of Applied Behavioral Science, 37,* 488-512.

Bobrow, W. & Schulz, M. (2002, October). *Applying technical advances in assessment centers.* Paper presented at the 30th International Congress on Assessment Center Methods. Pittsburgh, PA.

Bray, D.W., & Grant, D.L. (1966). The assessment center in the measurement of potential for business management. *Psychological Monographs, 80*, (17, Whole No. 625, 1 – 27.

Briscoe, D.R. (1997). Assessment centers: Cross-cultural and cross-national issues. *Journal of Social Behavior and Personality, 12,* 261-270.

Bycio, P., & Zoogah, B. (2002). Exercise order and assessment centre performance. *Journal of Occupational and Organizational Psychology, 75,* 109-114.

Byham, T. (2003). Ac follow up. Paper presented at the meeting of the Society for Industrial and Organizational Psychology, Orlando, Florida.

Byham, W.C. (2001, October). *What's happening in assessment centers around the world*. Paper presented at the 29th International Congress on Assessment Center Methods. Frankfurt, Germany.

Byham, W.C. (2002, October). *Growing leaders: We must do better.* Paper presented at the 30th International Congress on Assessment Center Methods. Pittsburgh, PA.

Byham, W.C., Smith, A., & Paese, M.J. (2001). *Grow your own leaders.* Pittsburgh, PA: DDI Press.

Caldwell, C., Thornton, G.C.III, & Gruys, M.L. (2003). Ten classic assessment center errors: Challenges to selection validity. *Public Personnel Management, 32,* 73-88.

Chawla, A. & Cronshaw, S. (2002, October). *Top-down vs bottom-up leadership assessment: Practical implications for validation in assessment centers.* Paper presented at the 30th International Congress on Assessment Center Methods. Pittsburgh, PA.

Collins, J.M., Schmidt, F.L., Sanchez-Ku, M., Thomas, M.A., McDaniel, M.A., & Le, H. (2003). Can basic individual differences shed light on the construct meaning of assessment center evaluations. *International Journal of Selection and Assessment, 11,* 17-29.

Craik, K.H., Ware, A.P., Kamp, J., O'Reilly, C., Staw, B., & Zedeck, S. (2002). Explorations of construct validity in a combined managerial and personality assessment programme. *Journal of Occupational and Organizational Psychology, 75,* 171-193.

Dailey, L., Cohen, B., & Lockwood, W. (1999, June). *Using assessment centers as a change strategy in a global company.* Paper presented at the 27th International Congress on Assessment Center Methods. Orlando, Florida.

Damitz, M., Manzey, D., Kleinmann, M., & Severin, K. (2003). Assessment center for pilot selection, Construct and criterion validity and the impact of assessor type. *Applied Psychology, An International Review, 52*, 193-212.

Dayan, K., Kasten, R., & Fox, S. (2002). Entry-level police candidate assessment center, An efficient tool or a hammer to kill a fly? *Personnel Psychology, 55,* 827-849.

Dowell, B. E., & Elder, E.D. (2002, October). *Accelerating the development of tomorrow's leaders.* Paper presented at the 30th International Congress on Assessment Center Methods. Pittsburgh, PA.

Eckardt, T. (2001, October). *Implementing integrated human resource development globally*. Paper presented at the 29th International Congress on Assessment Center Methods. Frankfurt, Germany.

Fleisch, J.M. (1995, May). *The human side of reengineering.* Paper presented at the 23rd International Congress on the Assessment Center Method, Kansas City, MO.

Fleisch, J. M., & Cohen, B. M. (1995, May). *Organizational change: You want it? You got it!* Paper presented at the 23rd International Congress on the Assessment Center Method, Kansas City, MO.

Ford, J. (2001, October). *Automating the collection and scoring of written (non-multiple choice) assessment center data*. Paper presented at the 29th International Congress on Assessment Center Methods. Frankfurt, Germany.

Gaugler, B.B., Rosenthal, D.B., Thornton, G.C. III, & Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology, 40,* 243-259.

Gebelein, S., Warrenfeltz, W., & Guinn, S. (1995, May). *Change in organizations – Assessment tools as solutions.* Paper presented at the 23rd International Congress on the Assessment Center Method, Kansas City, MO.

Goldstein, H.W., Zedeck, S., & Goldstein, I.L. (2002). *g*: Is this your final answer. *Human Performance, 15,* 123-142.

Goldstein, H.A., Yusko, K.P., Braverman, E.P., Smith, D.B. and Chung, B. (1998). The role of cognitive ability in the subgroup differences and incremental validity of assessment center exercises. *Personnel Psychology, 51*, 357-374.

Goldstein, H.W. Yusko, K.P., & Nicolopoulos, V. (2001). Exploring Black-White subgroup differences of managerial competencies. *Personnel-Psychology, 54*, 783-807.

Goodstone, M.S., & Lopez, F.E. (2001). The frame of reference approach as a solution to an assessment center dilemma. *Consulting Psychology Journal: Practice and Research, 53*, 96-107.

Gowing, M. (1999, June). *Defining assessment center dimensions: A national framework for global use*. Paper presented at the 27th International Congress on Assessment Center Methods. Orlando, Florida.

Haaland, S., & Christiansen, N.D. (2002). Implications of trait-activation theory for evaluating the construct validity of assessment center ratings. *Personnel Psychology, 55*, 137-163.

Hale, B., Jaffee, C., & Chapman, J. (1999, June). *How technology has changed assessment centers.* Paper presented at the 27th International Congress on Assessment Center Methods. Orlando, Florida.

Hardison, C.M., & Sackett, P.R. (2004, April). *Assessment center criterion related validity: A meta-analytic update.* Paper presented at the 18th Annual Conference of the Society for Industrial and Organizational Psychology, Chicago, IL.

Harel, G.H, Arditi, V.A., & Janz, T. (2003). Comparing the validity and utility of behavior description interview versus assessment center ratings. *Journal of Managerial Psychology, 18*, 94-104.

Hennessy, J., Mabey, B., & Warr, P. (1998). Assessment centre observation procedures: An experimental comparison of traditional, checklist and coding methods, *International Journal of Selection and Assessment, 6*, 222-231.

Henry, S.E. (1988). *Nontraditional applications of assessment centers: Assessment in staffing plant start ups.* Paper presented at the meeting of the American Psychological Association, Atlanta, GA.

Hiatt, J. (2000, May). *Selection for positions in a manufacturing startup.* Paper presented at the 28th International Congress on Assessment Center Methods. San Francisco, California.

Hoeft, S., & Schuler, H. (2001). The conceptual basis of assessment centre ratings. *International Journal of Selection and Assessment, 9*, 114-123.

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

Howard, A., & Metzger, J. (2002, October). *Assessment of complex, consultative sales performance*. Paper presented at the 30th International Congress on Assessment Center Methods. Pittsburgh, PA.

Howard, L. & McNelly, T. (2000). Assessment center for team member level and supervisory development. Paper presented at the 28th International Congress on Assessment Center Methods. San Francisco, CA.

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

Jacobson, L. (2000, May). *Portfolio assessment: Off the drawing board into the f*ire. Paper presented at the 28th International Congress on Assessment Center Methods. San Francisco, California.

Jansen, P.G.W., & Stoop, B.A.M.(2001). The dynamics of assessment center validity, Results of a 7-year study. *Journal of Applied Psychology, 86,* 741-753.

Jones, R.G. & Whitmore, M.D. (1995). Evaluating developmental assessment centers as interventions. *Personnel Psychology, 48,* 377-388.

Kelbetz, G., & Schuler, H. (2002). Does practice improve assessment center performance? [In German] *Zeitschrift für Personalpsychologie, 1,* 4-18.

Kohne, W. (2001, October). *Executive view of assessment centers.* Paper presented at the 29th International Congress on Assessment Center Methods. Frankfurt, Germany.

Kolk, N. J., Born, M. P., & Van der Flier, H. (2002). Impact of common rater variance on construct validity of assessment center dimension judgments. *Human-Performance, 15*, 325-338.

Kolk, N.J., Born, M.Ph., Van Der Flier, H., & Olman, J.M. (2002). Assessment center procedures, Cognitive load during the observation phase. *International Journal of Selection and Assessment, 10,* 271-278.

Kozloff, B. (2003, October). *Expatriate selection.* Paper presented at the 31st International Congress on Assessment Center Methods, Atlanta, GA

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

Kudisch, J.D., Lundquist, C., & Smith, A.F.R. (2002, October). *Reactions to "dual purpose" assessment center feedback.* Paper presented at the 30th International Congress on Assessment Center Methods. Pittsburgh, PA.

Kuptsch, C., Kleinmann, M., & Köller, O. (1998). The chameleon effect in assessment centers: The influence of cross-situational behavioral consistency on the convergent validity of assessment centers, *Journal of Social Behavior and Personality, 13,* 102-116.

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.

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. (2001a). Assessor training strategies and their effects on accuracy, inter-rater reliability, and discriminant validity. *Journal of Applied Psychology, 86,* 255-264.

Lievens, F. (2001b). Assessors and use of assessment center dimensions: A fresh look at a

troubling issue. *Journal of Organizational Behavior, 65*, 1-19.

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., & Harris, M.M., Van Keer, E., & Bisqueret, C. (2003). Predicting cross-cultural training performance: The validity of personality, cognitive ability, and dimensions measured by an assessment center and a behavior description interview. *Journal of Applied Psychology, 88*, 476-489.

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., De Fruyt, F., & Van Dam K. (2001). Assessors' use of personality traits in descriptions of assessment centre candidates: A five-factor model perspective. *Journal of Occupational and Organizational Psychology, 74*, 623-636.

Lovler, R. & Goldsmith, R.F. (2002, October). *Cutting edge developments in assessment center technology*. Paper presented at the 30th International Congress on Assessment Center Methods. Pittsburgh, PA.

Mahoney-Phillips, J. (2002, October). *Role profiling and assessment as an organizational management tool.* Paper presented at the 30th International Congress on Assessment Center Methods. Pittsburgh, PA.

Maurer, T., Eidson, C., Atchley, K., Kudisch, J., Poteet, M. Byham, T., Wilkerson, B. (2003). *Where do we go from here? Accepting and applying assessment center feedback.*

Paper presented at the 31st International Congress on Assessment Center Methods.

McCall, M.W. Jr. (2002, October). *Good news and bad news about developing global executives.* Paper presented at the 30th International Congress on Assessment Center Methods. Pittsburgh, PA.

McFarland, L.A., Ryan, A.M., & Kriska, S.D. (2003). Impression management use and effectiveness across assessment methods. *Journal of Management, 29*, 641-661.

Moser, K., Schuler, H., & Funke, U. (1999). The moderating effect of raters' opportunities to observe ratees' job performance on the validity of an assessment centre. *International Journal of Selection and Assessment, 7,* 133-141.

O'Connell, M.S., Hattrup, K, Doverspike, D., & Cober, A. (2002). The validity of "mini" simulations for Mexican retail salespeople. *Journal of Business and Psychology, 16,* 593-599.

Pendit, V. & Thornton, G.C. III. (2001, October). *Development of a code of conduct for personnel assessment in Indonesia*. Paper presented at the 29th International Congress on Assessment Center Methods. Frankfurt, *Germany.*

Reynolds, D. (2003, October). *Assessing executives and leaders through a technology-based assessment center.* Paper presented at the 31st International Congress on Assessment Center Methods, Atlanta, GA.

Richards, W. (2002,October). *A digital portfolio to support learning and development.* Paper presented at the 30th International Congress on Assessment Center Methods. Pittsburgh, PA.

Riggio, R.E., Mayes, B.T., Schleicher, D.J. (2003). Using assessment center methods for measuring undergraduate business outcomes. *Journal of Management Inquiry, 12,* 68-78.

Robie, C., Osburn, H. G., Morris, M. A., Etchegaray, J. M., & Adams, K. A. (2000). Effects

of the rating process on the construct validity of assessment center dimension evaluations. *Human Performance, 13,* 355-370.

Roth, E. & Smith, A. (2000, May). *The United States Postal Service: Reinventing executive succession planning.* Paper presented at the 28th International Congress on Assessment Center Methods. San Francisco, California.

Rupp, D.E. & Thornton, G.C. III. (2003). *Development of simulations for certification of competence of IT consultants.* Paper presented at the 18th Annual Conference of the Society for Industrial and Organizational Psychology, Orlando, FL.

Sackett, P.R. (1998). Performance assessment in education and professional certification: Lessons for personnel selection? In M.D. Hakel (Ed.), *Beyond multiple choice tests* (pp. 113-129). Mahwah: Lawrence Erlbaum.

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.

Schippman, J.S., Ash, R.A., Battista, M., Carr, L., Eyde, L.D., Hesketh, B., Kehoe, J. Pearlman, K, & Sanchez, J.I. (2000). The practice of competency modeling. *Personnel Psychology, 53,* 703-740.

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, B., & Konz, A.M. (1989). Strategic job analysis. *Human Resource Management, 28,* 51-63.

Seegers, J. & Huck, J. (2001, October). *European review of new assessment center guidelines.* Paper presented at the 29th International Congress on Assessment Center Methods. Frankfurt, Germany.

Smith, A., & Reynolds, D. (2002, October). *Automating the assessment experience: The latest chapter.* Paper presented at the 30th International Congress on Assessment Center Methods. Pittsburgh, PA.

Spector, P.E., Schneider, J.R., Vance, C.A., & Hezlett, S.A. (2000). The relation of cognitive ability and personality traits to assessment center performance. *Journal of Social Psychology, 30,* 1474-1491.

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

Stahl, G.K. (2000). Using assessment centers as tools for international management development: An exploratory study. In T. Kuehlmann, M. Mendenhall, & G. Stahl (Eds.), *Developing global leaders: Policies, processes, and innovations.* Quorum Books.

Tett, R.P., & Burnett, D.D. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology, 88,* 500-517.

Tett, R.P., & Guterman, H.A. (2000). Situation trait relevance, trait expression, and cross-situational consistency: Testing a principle of trait activation. *Journal of Research in Personality, 34,* 397-423.

Thomas, J.L., Dickson, M.W., & Bliese, P.D. (2001). Values predicting leader performance in the U.S. Army Reserve Officer Training Corps assessment center: Evidence for a personality-mediated model. *Leadership Quarterly, 12,* 181-196.

Thoreson, J. (2002, October). *Do we need dimensions? Dimensions limited unlimited*. Paper presented at the 30th International Congress on Assessment Center Methods. Pittsburgh, PA.

Thornton, G.C. III. (1992). *Assessment centers and human resource management.* Reading, MA: Addison-Wesley.

Thornton, G.C.III, & Byham, W.C. (1982). *Assessment centers and managerial performance*. New York: Academic Press.

Thornton, G.C.III, & Cleveland, J.N. (1990). Developing managerial talent through simulation. *American Psychologist, 45*, 190-199.

Thornton, G.C.III, & Mueller-Hanson, R.A. (2004). *Developing organizational simulations: A guide for practitioners and students.* Mahwah, NJ: Erlbaum.

Thornton, G.C.III, & Rupp, D.E. (2004). Simulations and assessment centers. In M. Hersen (Ed.). *Comprehensive handbook of psychological assessment: Vol. 4* J.C. Thomas (Ed.), *Industrial and organizational assessment.* (pp. 319-344). Hoboken, N.J.: Wiley.

Tillema, H. H. (1998). Assessment of potential, from assessment centers to development centers. . *International Journal of Selection and Assessment, 6*, 185-191.

Viswesvaran, C., Ones, D.S., & Schmidt, F.L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology, 81*, 557-574.

Wilson, L. & Pilgram, M. (2000, May). *Diagnose…then prescribe.* Paper presented at the 28th International Congress on Assessment Center Methods. San Francisco, California.

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.