

Guidelines and Ethical Considerations for Assessment Center Operations

Prepared and Endorsed by:

The International Taskforce on Assessment Center Guidelines

Deborah E. Rupp, Purdue University, USA, Chair
Brian J. Hoffman, University of Georgia, USA, Co-Chair
David Bischof, Deloitte, South Africa, Co-Chair

Taskforce Members:

William Byham, Development Dimensions International, USA
Lynn Collins, Fenestra, USA
Alyssa Gibbons, Colorado State University, USA
Shinichi Hirose, International University of Japan, Japan
Martin Kleinmann, University of Zurich, Switzerland
Jeffrey D. Kudisch, University of Maryland, USA
Martin Lanik, Pinsight, USA
Duncan J. R. Jackson, Birkbeck, the University of London, UK
Myungjoon Kim, Assesta, South Korea
Filip Lievens, Ghent University, Belgium
Deon Meiring, University of Pretoria, South Africa
Klaus G. Melchers, Universität Ulm, Germany
Vina G. Pendit, Daya Dimensi, Indonesia
Dan J. Putka, Human Resources Research Organization, USA
Nigel Povah, Assessment and Development Consultants, UK
Doug Reynolds, Development Dimensions International, USA
Sandra Schlebusch, LEMASA, South Africa
John Scott, APTMetrics, USA
Svetlana Simonenko, Detech, Russia
George Thornton, Colorado State University, USA

Table of Contents

- I. Purpose**
- II. History of Guidelines**
- III. Assessment Center Defined**
- IV. Non-Assessment Center Activities**
- V. Assessment Centers for Different Purposes**
- VI. Assessment Center Policy Document**
- VII. Assessor Training**
- VIII. Training and Qualifications of Other Assessment Center Staff**
- IX. Validation Issues**
- X. Technology**
- XI. Ethics, Legal Compliance, and Social Responsibility**
- XII. Conducting Assessment Centers Across Cultural Contexts**
- XIII. National Assessment Center Guidelines**

Appendix A: Past Taskforce Members

Appendix B: Glossary of Relevant Terms

Appendix C: Relevant Professional Guidelines

Appendix D: Key Sources/Recommended Readings

I. Purpose

This document's intended purpose is to provide professional guidelines and ethical considerations for users of the assessment center method. These guidelines are designed to cover both existing and future applications. The title "assessment center" is restricted to those methods that follow these guidelines.

These guidelines will provide: (1) guidance to industrial/organizational/work psychologists, organizational consultants, human resource management specialists and generalists, and others who design and conduct assessment centers; (2) information to managers deciding whether or not to institute assessment center methods; (3) instruction to assessors serving on the staff of an assessment center; and (4) guidance on the use of technology and navigating multicultural contexts; (5) information for relevant legal bodies on what are considered standard professional practices in this area.

II. History of Guidelines

The growth in the use of the assessment center method over the last several decades has resulted in a proliferation of applications in a variety of organizations. Assessment centers currently are being used in industrial, educational, military, government, law enforcement, and other organizational settings, and being used all over the world. Background on each Edition of the *Guidelines* is provided below. Appendix A provides Taskforce members for each Edition.

1st Edition. From the beginning of its use in modern organizational settings, practitioners raised concerns that reflected a need for standards or guidelines for users of the assessment center method. This resulted in the first International Taskforce on Assessment Center Guidelines to be formed, chaired by Joseph L. Moses. The 3rd International Congress on the Assessment Center Method, which met in Quebec in May of 1975, endorsed the first set of Guidelines, which were based on the observations and experience of a group of professionals representing many of the largest users of the method.

2nd Edition. Developments in the period of 1975 to 1979 concerning federal guidelines related to testing, as well as professional experience with the original guidelines, suggested that the guidelines should be evaluated and revised. Therefore, the 1979 Guidelines included essential items from the original Guidelines, but also addressed the recognized need for: (1) further definitions; (2) clarification of impact on organizations and participants; (3) expanded guidelines on training; and (4) additional information on validation. The taskforce for the Second Edition was chaired by Joel Moses, and endorsed by the 7th International Congress on the Assessment Center Method, which met in New Orleans, Louisiana in June of 1979.

3rd Edition. Following the publication of the Second Edition, a wider variety of organizations were adopting the assessment center method and using it to assess individuals for a more diverse array of jobs. Stakeholders had begun to demand more streamlined procedures that were less time-consuming and expensive. In addition, new theoretical arguments and evidence from empirical research had been interpreted to mean that the assessment center method does not work

exactly as its proponents originally had believed, suggesting that the method should be modified. Finally, many procedures purporting to be assessment centers had not complied with previous Guidelines—which was thought to be due to the previous Guidelines being too ambiguous. The 1989 revision of these Guidelines was begun at the 15th International Congress on the Assessment Center Method in Boston (April 1987), led by Douglas Bray. Subsequently, the third Taskforce was formed, chaired by Douglas Bray and George Thornton, who solicited feedback from the 16th International Congress held in May of 1988 in Tampa. The final version of the Third Edition was endorsed by a majority of the Taskforce and by the 17th International Congress held in May of 1989 in Pittsburgh. Revisions/additions involved: (1) specification of the role of job analysis; (2) clarification of the types of attributes/dimensions to be assessed and whether or not attributes/dimensions must be used; (3) delineation of the processes of observing, recording, evaluating, and aggregating behavioral information; and (4) further specification regarding assessor training.

4th Edition. The 2000 revision of these Guidelines was initiated at the 27th International Congress on Assessment Center Methods in Orlando (June 1999). The Taskforce for the 4th Edition, chaired by David MacDonald, conducted discussions with a number of assessment center experts in attendance and also solicited input at a general session regarding aspects of the Guidelines that needed to be (re)addressed. A primary factor driving this revision was the passage of a full decade since the 3rd Edition. Other factors included a desire to integrate technology into assessment center methods and recognition of the need for more specific definitions of several concepts and terms. Input was synthesized into a final draft that was presented and endorsed at the 28th International Congress held in May of 2000 in San Francisco, which was attended by 150 delegates representing Australia, Belgium, Brazil, Canada, Columbia, Germany, India, Indonesia, Italy, Japan, Mexico, the Netherlands, the Philippines, Singapore, Sweden, Switzerland, Taiwan, the United Arab Emirates, the United Kingdom, and the United States of America.

5th Edition. The 5th Edition of these Guidelines was initiated at the 32nd International Congress on Assessment Center Methods, which was held in Las Vegas in October of 2004. A roundtable discussion addressed contemporary assessment center issues on which there had been little previous guidance. Subsequently, this Congress decided that additions and revisions were needed in two areas: First, because of the proliferation of multinational organizations using assessment centers across geographic regions, more guidance was needed on global assessment center practices. The 32nd Congress established a sub-taskforce to examine this issue. A report from this taskforce served as the foundation for a new section of the Guidelines. Second, given recent research on the effectiveness of various assessor training components, the Congress suggested an expansion of the Guidelines in this area as well. A second round of discussions on these issues was held in 2006 at the 33rd International Congress in London. These discussions suggested additional guidance in two areas: (1) the use of technology in assessment center practices; and (2) recognition of methodological differences among assessment centers used for different purposes. The resulting revision, led by Deborah Rupp and Doug Reynolds, was unanimously endorsed by the 34th International Congress (2008, Washington, DC), which was attended by delegates representing Austria, Belgium, Canada, China, Germany, India, Indonesia, Mexico, the Netherlands, Romania, Russia, Singapore, South Africa, South Korea, Spain, Sweden, the United Arab Emirates, the United Kingdom, and the United States of America.

6th Edition. The current, 6th Edition, presented herein, was initiated due to three recent developments since 2009. First, new and compelling research has amassed, generally regarding the construct validity of assessment center ratings. This evidence has important implications for the focal constructs assessed by assessment centers, the development of simulation exercises, assessor training, and the use of assessment center ratings. Second, continued delineation was seen as needed between assessment center programs serving different HR functions, and supporting different talent management objectives. Finally, multicultural and technological challenges were seen as continuing to pervade assessment center applications. A Taskforce for the 6th Edition was formed, chaired by Deborah Rupp, Brian Hoffman, and David Bischof. A revision was prepared by the Taskforce, which included the following additions and revisions:

- a) The use of the broader term “behavioral constructs” to refer to what is assessed via the assessment center method (to include dimensions, competencies, tasks, KSAs (knowledge, skills, and abilities), and other constructs, so long as they are defined behaviorally and comply with the criteria outlined herein)
- b) Recognition of the state of the research literature supporting the construct validity of, and thus the use of, these various types of behavioral constructs
- c) Acknowledgment of the state of the research literature supporting the use of various types of behavioral constructs
- d) More comprehensive coverage of assessment centers used for different purposes and used to serve different talent management (and strategic management) functions
- e) New sections on:
 - i. The training/certification of other assessment center staff (beyond assessors)
 - ii. The incorporation of technology into assessment center operations
 - iii. Ethical, legal, and social responsibilities
- f) Additional information on:
 - i. Translations of AC materials and the simultaneous use of multi-language versions
 - ii. Data security and (international) data transfer
 - iii. The complementary role played by the International Guidelines alongside various countries’ national assessment center guidelines
- g) Other additions and expansions reflective of the current state of science and practice

The Taskforce is currently seeking the endorsement of the 38th International Congress on Assessment Center Methods, convening in October 2014 in Alexandria, Virginia, USA.

III. Assessment Center Defined

An *assessment center* consists of a standardized evaluation of behavior based on multiple inputs. Any single assessment center consists of multiple components, which include behavioral simulation exercises¹, within which multiple trained assessors observe and record behaviors, classify them according to the behavioral constructs of interest, and (either individually or collectively) rate (either individual or pooled) behaviors. Using either a consensus meeting among assessors or statistical aggregation, assessment scores are derived that represent an assessee's standing on the behavioral constructs and/or an aggregated overall assessment rating (OAR).

Assessment centers can be used for multiple purposes. Most commonly, these purposes include *prediction* (e.g., for personnel selection, promotion, or succession planning), *diagnosis* (i.e., to identify strengths and areas for training/development for the purpose of development planning), and *development* (i.e., as a training intervention in and of itself, or as part of a larger initiative). Assessment centers must be developed, implemented, and validated/evaluated in ways specific to the intended purpose of the program, and according to the talent management goals of the hosting organization (see Section V).

All assessment center programs must contain ten essential elements:

1. Systematic Analysis to Determine Job-Relevant Behavioral Constructs—The focal constructs assessed in an assessment center have traditionally been called “behavioral dimensions” or simply “dimensions” within assessment center science and practice, and are defined as a constellation or group of behaviors that are specific, observable, and verifiable; that can be reliably and logically classified together; and that relate to job success. The term dimension is sometimes used synonymously with competency or KSA (knowledge, skills, or ability). Other assessment center applications have classified relevant behaviors according to tasks or job roles. Regardless of the label for the focal constructs to be assessed, they must be defined behaviorally, and as such are referred to hereafter as “behavioral constructs.”² Behaviors in any definition of a behavioral construct may be either broad or specific in relation to a particular context or job.

Further, these behavioral constructs must be derived via a rigorous and systematic process (e.g., job analysis, competency modeling) that considers how the construct manifests in the actual

¹ and which may also include other tests and forms of assessment

² The expansion of this definition has led some assessment center researchers and practitioners to use overall performance in each simulation exercise as the behavioral constructs, while other applications have begun to use dimension performance linked to specific simulated situations as a meaningful unit of behavioral information. The research evidence to date provides support for the use of traditional dimensions, and new research is amassing that supports the incorporation of situation-dependent behaviors into the interpretation of dimension-level performance. A smaller number of studies have presented evidence for exercise-based interpretations of assessment center performance. These statements are supported by studies demonstrating the reliability of assessor ratings, the multi-faceted structure of ratings within assessment centers, and relationships of assessment center ratings with comparable measures outside the assessment center including criteria of job performance and tests of cognitive ability and personality.

job/organizational context, and documents the job relevance of the final behavioral constructs incorporated into the assessment context. The type and extent of analysis will depend on the purpose of the assessment; the complexity of the job; the adequacy and appropriateness of prior information about the job; and the similarity of the job to jobs that have been studied previously. If past research/analyses are used to select behavioral constructs and exercises, evidence of the comparability or generalizability of the jobs must be provided. When the job does not currently exist, analyses can be done of actual or projected tasks or roles that will comprise the new job, position, job level, or job family. Analysis of the organization's vision, values, strategies, or key objectives may also inform identification of appropriate behavioral constructs. However, if the assessment center is designed to inform selection decisions, then in certain countries (e.g., the U.S.), basing the choice of behavioral constructs largely on analysis of the organization's vision, values, strategies, or key objectives with little consideration of behavioral requirements of the target job, would be inconsistent with legal and professional guidelines for the development of selection measures.

Rigor in this regard is defined as the involvement of subject matter experts who are knowledgeable about job requirements; the collection and quantitative evaluation of essential job elements; and the production of evidence that assessment center scores are reliable. Any job analysis, competency-modeling, or related undertaking must result in clearly-specified categories of behavior that can be observed over the course of the assessment procedures. The behavioral constructs must be defined precisely and expressed in terms of behaviors observable on the job (or within the job family) and in the simulation exercises used within the assessment center. Behavioral constructs must also be shown to be related to success in the target job, position, or job family.

2. Behavioral Classification—The behaviors captured within the assessment context (e.g., trained assessors' behavioral observations of assessees participating in simulation exercises), must be classified according to the behavioral constructs. Further classification might also take place, such as into broader performance categories or an overall assessment rating (OAR).

3. Multiple Assessment Center Components—Any assessment center must contain multiple assessment components, some of which consist of behavioral simulation exercises. As such, assessment centers may be entirely comprised of multiple behavioral simulation exercises, or some combination of simulations and other measures, such as tests (referred to in some countries as "psychometric tests"), structured interviews, situational judgment tests, questionnaires, and the like. The assessment center components are developed or chosen to elicit a variety of behaviors and information relevant to the behavioral constructs. Self-assessment and multisource assessment data may also be gathered as assessment information. Each assessment component must be pretested to ensure that it provides reliable, objective, and relevant behavioral information for the organization in question. Pretesting might entail trial administration with participants similar to the intended assessees, thorough review by subject matter experts as to the accuracy and representativeness of behavioral sampling, and/or evidence from the use of these techniques for similar jobs in similar organizations.

4. Linkages Between Behavioral Constructs and Assessment Center Components—A matrix mapping what behavioral constructs are assessed in each assessment center component must be

constructed. This is most commonly referred to as a dimension-by-exercise matrix. Evidence must be established supporting the inferences made as the assessment center developer moves from job analysis (or competency modeling) information to the choice of behavioral constructs, and then to the choice of assessment components to measure each construct in multiple ways.

5. *Simulation Exercises*—An assessment center must contain multiple opportunities to observe behaviors relevant to the behavioral constructs to be assessed. At least some job-related simulation exercises must be included.

A simulation exercise is an assessment technique designed to elicit behaviors representative of the targeted behavioral constructs and within a context consistent with the focal job. They require assessees to respond behaviorally to situational stimuli. Examples of simulations include, but are not limited to, in-box exercises, leaderless group discussions, case study analyses/presentations, role plays, and fact-finding exercises. Stimuli can be presented via a variety of formats, including face-to-face interaction, paper, video, audio, computers, telephones, or the Internet. The format used to present stimuli should,³ as far as possible, be consistent in nature to how such information would be delivered in the actual job environment. For simple jobs, one or two job-related simulations may be used if the job analysis clearly indicates that one or two simulations alone sufficiently simulate a substantial portion of the job being evaluated. If a single comprehensive assessment technique is used (e.g., a computer-delivered simulation that simulates a number of tasks and situations), then it must include distinct, job-related segments.

Simulation exercises must be carefully designed and constructed such that a large number of behavioral construct-related behaviors can be reliably elicited and detected by assessors. Behavioral cues (i.e., prompts provided by role players or via other stimuli provided within the context of a simulation exercise, incorporated for the purpose of creating opportunities for displaying behavior relevant to the behavioral constructs) should be determined and documented prior to or during exercise development, and incorporated into both assessor training and scoring protocol. The stimuli contained in a simulation must parallel or resemble stimuli in the work situation, although they may be in different settings. The desirable degree of fidelity is a function of the assessment center's purpose. Fidelity may be relatively low for early identification and selection programs for non-managerial personnel and may be relatively high for programs designed to diagnose the training needs of experienced managers, executives, and other professionals. Assessment center designers must take steps to ensure that the exercise content does not unfairly favor certain assessees (e.g., those in certain racial, ethnic, age, or sex groups).

To qualify as a behavioral simulation for an assessment center as defined herein, the assessment method must require the assessee to overtly display certain behaviors. The assessee must be required to demonstrate a constructed response (i.e., as opposed to choosing among pre-determined behavioral options). Assessment procedures that only require the assessee to select among provided alternative responses (e.g., multiple-choice tests, situational judgment tests, and some computerized in-baskets and 3D virtual games) do not conform to this requirement. Similarly, a situational interview that calls for only an expression of behavioral intentions would

³ Note that the use of the term “should” throughout these Guidelines refers to strongly recommended/desirable practices. Whereas these refer to expected practices, the Taskforce does recognize that there may be some instances when they are not feasible or applicable.

not be seen as conforming to this criterion. Whereas such techniques may yield highly reliable and valid assessment ratings, they would not be classified as a behavioral simulation exercise.

6. Assessors—Multiple assessors must be used to observe and evaluate each assessee. When selecting assessors, where appropriate, the assessment center program must strive to have diverse assessors, both in terms of demographics (e.g., race, ethnicity, age, sex) and experience (e.g., organizational level, functional work area, managers, psychologists, etc.). The maximum ratio of assessees to assessors is a function of several variables, including the type of exercises used, the behavioral constructs to be evaluated, the roles of the assessors, the type of data integration carried out, the amount of assessor training conducted, the experience of the assessors, and the purpose of the assessment center. The ratio of assessees to assessors should be minimized where practicable in the interests of reducing cognitive load (and for group simulation exercises, the number of assessees an assessor must assess simultaneously should be kept to a minimum). To minimize potential bias, an assessee’s current supervisor should not be involved in the assessment of a direct subordinate when the resulting data will be used for selection or promotional purposes.

7. Assessor Training—Assessors must receive thorough training and demonstrate performance that meets pre-specified criteria. Training must include instruction on the purpose and goals of the assessment center; behavioral constructs to be assessed and associated behaviors; the assessment center components to be utilized; the materials and rubrics with which to document, classify, and evaluate behaviors, as well as the rights and responsibilities of assessees, assessors, and the host organization and affiliated consulting bodies. It must also include instruction on making ratings and calibrating scoring levels associated with specific behaviors and behavioral constructs (often referred to as “frame of reference training”). Assessors must only be allowed to assess actual assessees after demonstrating their competence and reliability, both individually and as a group. If assessors also serve as feedback providers, then training should also address strategies for enhancing feedback acceptance and behavior change. More information on assessor training is provided in Section VII below.

8. Recording and Scoring of Behaviors—A systematic procedure must be used by assessors to record (and if appropriate, rate) specific behavioral observations accurately at the time of observation. This procedure might include the use of note-taking, behavioral observation scales (BOS), behavioral checklists, or behaviorally anchored rating scales (BARS). Observations may also occur *post hoc* by accessing audio and/or video recordings taken as assessees complete simulation exercises. Assessors must prepare a record/report of the observations made during each exercise before the integration discussion or before statistical integration takes place. Behavioral categorization, scoring, and reporting must always be according to the pre-determined/validated set of behavioral constructs that form the foundation of the assessment center.

9. Data Integration—The integration of observations and/or ratings of each assessee’s behaviors must be based on a discussion of pooled observations and ratings from various assessors and/or a statistical integration of assessors’ ratings. The process used must be carried out in accordance with professionally accepted standards. Depending on the purpose of the assessment center,

integration may result in exercise-specific “dimension”⁴ scores; exercise scores; across-exercise dimension scores; and/or an overall assessment rating (OAR). If an integration discussion amongst assessors (also known as a “consensus discussion”) is used, assessors must consider the behavioral construct-relevant information collected from the assessment components, and not consider information obtained outside the documented processes of the AC. Regardless of method of integration, the scores yielded by the integration process must be reliable. In both computing and interpreting assessment center scores, consideration of how assesseees perform across diverse situations should be considered. Depending on the purpose and design of the assessment center, this might include weighting behaviors based on the extent to which they manifest themselves on the job (e.g., the number of critical job tasks that a particular behavioral dimension is linked to based on a job analysis); providing feedback on exercise-specific dimension performance; considering “split” ratings (when performance on a given behavioral construct is high in one situation but low in another) as potentially meaningful information; or providing exercise-specific feedback.

10. Standardization—The procedures for administering all aspects of an assessment center must be standardized so that all assesseees have the same opportunities to demonstrate behaviors relevant to the behavioral constructs. Standardization is especially important for high-stakes assessment centers, where the outcomes are used to make decisions about the employment status of individuals (e.g., assessment centers that inform selection and promotion decisions). Standardization may be compromised in many aspects of AC administration, including the instructions given, time allowed for completion of exercises, materials available, the room and other facilities, the composition of groups in group interaction exercises, the behavior of role players, follow-up questions asked by assessors after a presentation, differing sequences of assessment components, etc. Other considerations for standardization are discussed in Sections X and XII. Exceptions to strict adherence to standardized procedures may be allowed in response to legitimate, documented requests for accommodation for a disability (e.g., more time for a person with a reading disability). Similarly, the requirement for strict standardization does not apply to individually customized assessments used in developmental settings (although even in such settings, when assesseees participate in the same assessment components, these components should be carried out in a standardized way).

IV. Non-Assessment Center Activities

There is a difference between an assessment center and the application of assessment center methodology more generally. Various features of the assessment center methodology are used in procedures that do not meet all the Guidelines set forth herein, such as when a psychologist or human resource professional, acting alone, uses a simulation as part of an individual’s evaluation. Such personnel assessment procedures are not covered by these Guidelines; each should be judged on its own merits. Procedures that do not conform to all the Guidelines herein should not be represented as assessment centers or imply that they are assessment centers by using the term “assessment center” as part of the title.

The following kinds of activities *do not* constitute an assessment center:

⁴ Or alternative behavioral constructs.

1. Assessment procedures that do not require the assessee to demonstrate overt behavioral responses are not behavioral simulations; thus, any assessment program that consists solely of such procedures is not an assessment center as defined herein. Examples of these are computerized in-baskets and situational judgment tests marketed as “simulations” calling only for closed-ended responses (e.g., rating the effectiveness of behavioral response options, ranking potential behavioral responses, and multiple choice responses), situational interviews calling only for behavioral intentions, and written competency tests. Note that procedures not requiring an assessee to demonstrate overt behavioral responses may be used within an assessment center, but must be coupled with some simulation exercises requiring the overt display of behaviors.⁵
2. Panel interviews or a series of sequential interviews as the sole technique.
3. Reliance on a single assessment component (regardless of whether it is a simulation) as the sole basis for evaluation. This restriction does not preclude a comprehensive assessment that includes distinct job-related segments (e.g., large, complex simulations or virtual assessment centers with several definable sub-components and with multiple opportunities for observation in different situations).
4. A test battery (lacking any behavioral simulation exercises), regardless of whether the scores on the individual tests are combined via a statistical or judgmental pooling of scores.
5. Single-assessor evaluation (i.e., measurement by one individual using a variety of techniques, such as paper-and-pencil tests, interviews, personality measures, or simulations). Even if multiple assessors are used to assess multiple assessees, if each individual assessee is not evaluated by multiple assessors over the course of the assessment, the program cannot be referred to as an assessment center.
6. The use of several simulation exercises, which incorporate multiple assessors, but that does not pool the assessment data in any way (i.e., across assessors, exercises, dimensions, and/or alternative constructs).
7. A physical location labeled as an “assessment center” that does not conform to the methodological requirements noted above.
8. A website or catalog that warehouses various tests, measures, and assessments.
9. Fully automated, computerized assessments that either do not elicit overt behavior on the part of the assessee or do not require assessor observation and evaluation of overt behavior.

V. Assessment Centers for Different Purposes

Assessment centers are generally used for three major purposes: (1) to predict future behavior for decision making, (2) to diagnose development needs, and (3) to develop assessees on

⁵ Note that overt behaviors may be displayed as a production of a written, constructed response.

behavioral constructs of interest. However, additional purposes for the assessment center method currently exist and will continue to evolve with further use.

The design and operation of an assessment center must vary, depending on its intended purpose. For example, assessment centers designed purely to support personnel decisions (e.g., promotion), must place emphasis on reliable and valid overall assessment ratings. Alternatively, diagnostic assessment centers may require the generation of reliable and valid behavioral construct scores only. Here, illustrative behaviors to support strengths and development needs are particularly important.

Developmental assessment centers (DACs) seek to both assess and develop assessees on behavioral constructs.⁶ Here, it is essential that the behavioral constructs chosen for the program can actually be improved upon within the scope of the program (which may extend beyond the assessment events themselves). DACs involve multiple points of feedback and repeated practice, and may repeat exercises of the same type(s) as a way to track improvement on the behavioral constructs over time. As a result, such programs may be longer than assessment centers for prediction and diagnosis. Feedback is an essential component of a DAC program, and in order to foster learning, this feedback needs to be immediate. Often, the role of assessors in DACs is not only to observe and evaluate behavior, but also to facilitate learning and development by delivering feedback during the assessment process as well as follow-up coaching.

If the focus is purely on learning, DACs may be customized to meet assessees' developmental needs. When validating or otherwise evaluating DACs, the appropriate criterion is positive change in assessees' understanding, behavior, and proficiency on behavioral constructs.

Due to the differences outlined above, assessment centers must be designed, implemented, and validated appropriately for their intended purpose.

VI. Assessment Center Policy Document

Assessment centers operate more effectively as part of an integrated human resource/talent management system, which should be further integrated into the organization's strategic management portfolio. Prior to the introduction of an assessment center into an organization, a policy document must be prepared and approved by the organization. The following lists the items to be included in the policy document. The procedures described in the policy document must be carefully carried out. The policy document will specify what has been done and what will be done to develop, implement, and evaluate the AC.

1. Objective—The purpose of the assessment center program. An assessment center may be used for a variety of purposes. Falling into the broad categories of selection vs. diagnosis vs. development, such purposes might include prescreening, hiring, early identification and

⁶ Some countries have adopted the term “development center (DC)” as a broad term that would encompass *both* assessment centers solely designed to diagnose development needs, and for assessment centers designed to catalyze development within the course of the assessment center program. The term “developmental assessment center” is used herein to refer only to the latter.

evaluation of potential, succession planning, and professional development. The objective must be included in the assessment center policy document. It must also be stated (or at the very least ensured as general practice) that:

- a) Assesseees will be told, prior to the assessment, what decision(s) will be made with the assessment center data.
- b) If the organization desires to make decisions with the data other than those previously communicated to the assessee, the decision(s) will be clearly described to the assessee and consent obtained. The policy document should also clearly specify who will have access to assessment center data as noted in point 4 below.
- c) The design, development, implementation, and validation of the program will all be carried out in ways appropriate to the purpose of the assessment center.
- d) Decisions about the choice of behavioral constructs, content of simulations, selection and training of assessors, scoring, feedback, and evaluation will all be made with the objective in mind.

2. Assesseees—The population to be assessed; the method for selecting assesseees from this population; the procedure for notification; and the activities that the assesseees will carry out.

3. Assessors and other program staff—The assessor population (including sex, age, race, ethnicity, and relevant background/expertise/qualifications); the method for selecting assessors from this population; the diversity goals for the final assessor pool (including both demographics and experience/qualification); how the final assessor credentials will be reported in final documentation; the plan, format, length, and general content of the assessor training program, including information on how assessors will be evaluated and/or certified; other details pertinent to the selection and training of other program staff (e.g., role players, coordinators, see Sections VII and VIII).

4. Use of Data—The process flow of assessment records within the organization; specification of who will receive access to reports (e.g., supervisors, upper level management, HR); restrictions on access to information; procedures and controls for research and program evaluation purposes; feedback procedures to management and employees; the length of time for which data will be maintained in files. Particularly for a selection application, it is recommended that the data be used within two years of the date of administration because of the likelihood of changes in both assesseees and the organizational context. This section will also describe the collection, storage, and use of data electronically and/or over the Internet, as well as planned compliance with any relevant electronic data security laws or standards (see Sections X and XI).

5. Qualifications of Consultant(s) or Assessment Center Developer(s)—The internal or external consultant(s) responsible for the development of the assessment center, individual assessment center components, assessor training, feedback, and evaluation/validation, along with his/her/their professional qualifications, experience, and related training.

6. Validation—The validation model to be used, and the evidence supporting the use of the assessment center for its intended purpose. If a content-oriented validation strategy is used, this will include documentation of the relationship of the job/job family content to the behavioral

constructs and exercises, along with evidence of the reliability and/or agreement⁷ of the observations and ratings of behavior. If evidence is being taken from prior validation research, which may have been summarized in meta-analyses, this will include documentation that the current job/job family and assessment center are comparable and generalizable to the jobs and assessment centers studied elsewhere (often referred to as a transportability study). If a local, criterion-related validation strategy is used, this will include full documentation of the study. If the assessment center is being used for developmental purposes, this will include training evaluation results documenting learning and improvement on the behavioral constructs. If validation studies are under way, a schedule indicating when a validation report will be available should be provided. Information should also be provided pertaining to ongoing evaluation and periodic review of program validity over time.

Although these Guidelines do not prescribe use of a specific type of score, as this will vary across assessment centers, what is paramount is that the validation evidence supporting the way in which the scores are ultimately used (in terms of their validity and reliability for the purpose at hand) is provided by the AC developer/user. Whether these scores are exercise-specific dimension scores, across-exercise dimension scores, or some other type of aggregate score is not critical—what matters here is that the developer defends the validity of those scores in reference to how they are being used.

For all assessment center programs, the policy document will additionally disclose both potential benefits and risks to stakeholder groups impacted by the assessment center program. Risks include potentially unintended negative consequences on vulnerable and legally protected groups.

7. Legal Context—The particular laws and policies that are relevant for the assessment center program and how legal compliance will be ensured. Laws existing both in an organization's/agency's home state, province, or nation—as well as in the state, province, or nation where the assessment center program is being carried out—may have implications for program design, validation, implementation, and documentation. Most nations have disadvantaged and protected groups (such as native/aboriginal people, racial groups, religious groups, and those protected on the basis of age, gender, disability, sexual orientation, etc.) with accompanying regulations providing various legal protections. Laws and standards also exist governing the delivery of assessment content over the Internet (and across international borders), as well as electronic data security and management. Such regulations must also be considered (see #8 below, and Sections X and XI).

8. Use of Technology—A list of technical requirements for administering the assessment center program. This includes requirements for conducting assessor (and other staff) training, scheduling and communicating with assessees and program staff, administering the assessment

⁷ If the assessment center's scores are used to make relative comparisons among candidates, then using an appropriate index of reliability is sufficient for estimating the amount of measurement error present in scores. If, however, assessment center scores are used to make pass/fail, hire/no hire, or some other decision that involves comparing candidates' scores to some fixed cut-off or standard, then providing an appropriate index of absolute agreement among observations or ratings of behavior is recommended as best practice. In the latter case, indices of reliability can underestimate the amount of measurement error present in the assessment center scores. For guidance on this issue, see the Haertel chapter listed in Appendix D.

components, scoring and integration, report generation, feedback delivery, and data/report storage, as well as details surrounding system maintenance and the overall security protocol (see Sections X and XI).

VII. Assessor Training

Assessor training is an integral part of the assessment center program. Assessor training must have clearly stated training objectives, performance guidelines, and quality standards. The following issues related to training must be considered:

1. Training Content—Whatever the approach to assessor training, the objective is to obtain reliable and accurate assessor judgments. A variety of training approaches may be used (e.g., lectures, discussion, observation of practice assesseees, video demonstrations, observation of other assessors) as long as it can be shown that reliable, accurate assessor judgments are obtained. At a general level, all assessor training programs must include training on:

- a) The behavioral constructs to be assessed, including their behavioral definitions
- b) The observation, recording, classification, and evaluation of behaviors relevant to the behavioral constructs to be assessed
- c) The content of the simulation exercises as well as which behavioral constructs are targeted in which simulation exercises; including examples of effective and ineffective performance in each simulation exercise
- d) Awareness of the nature of common observational and rating errors (including how to distinguish behaviors from inferences)
- e) Security/confidentiality, standards of professionalism, and issues of fairness and non-discrimination

Depending on the purpose of the assessment center, the training might include additional components such as knowledge of the organization, knowledge of the target job, the ability to give accurate oral or written feedback, and consistency in role playing.

The following minimum training goals are required:

- a) Knowledge of the organization and job/job family or normative group being assessed to provide an effective context for assessor judgments where appropriate
- b) Thorough knowledge and understanding of the behavioral constructs, their definitions, their relationship to job performance, and examples of effective and ineffective performance
- c) Thorough knowledge and understanding of the assessment techniques, exercise content, relevant behavioral constructs to be observed in each assessment center component, expected or typical behavior, and examples or samples of actual behaviors
- d) Demonstrated ability to observe, record, and classify behavior (or lack of behavior) into behavioral constructs, including knowledge of the protocol for documenting behavior
- e) Thorough knowledge and understanding of evaluation and rating procedures, including how data are integrated

- f) Demonstrated evidence of inter-rater reliability, inter-rater agreement, and/or agreement with standard/expert ratings⁸
- g) Thorough knowledge and understanding of assessment policies and practices of the organization, including restrictions on how assessment data are to be used
- h) Thorough knowledge and understanding of feedback procedures and strategies, where appropriate, to maximize assessee's acceptance of feedback and behavior change
- i) Demonstrated ability to give accurate oral and written behavioral feedback, when feedback is given by the assessors, and to do so in a manner that maintains or enhances assessee self-esteem
- j) Demonstrated knowledge and ability to play objectively and consistently the role called for in interactive exercises (e.g., one-on-one simulations or fact-finding exercises), when role playing is required of assessors. Non-assessor role players also may be used if their training results in their ability to play the role objectively and consistently (see Section VIII).

2. Training Length—The length of assessor training may vary due to a variety of considerations that can be categorized into three major areas:

- a) Trainer and instructional design considerations:
 - i. The instructional mode(s) utilized
 - ii. The qualifications and expertise of the trainer
 - iii. The training and instructional sequence
- b) Assessor considerations:
 - i. Previous knowledge and experience with similar assessment techniques
 - ii. Type of assessors used (e.g., professional psychologists vs. managers)
 - iii. Experience and familiarity with the organization and the target position(s)/job(s)/job families/target level
 - iv. The frequency of assessor participation
 - v. Other related qualifications and expertise (e.g., testing and assessment, executive coaching)
- c) Assessment program considerations:
 - i. The target position's level of difficulty
 - ii. The number of behavioral constructs to be assessed
 - iii. The anticipated use of the assessment information (e.g., immediate selection, broad placement considerations, diagnosis, development)
 - iv. The number of exercises and their complexity
 - v. The division of roles and responsibilities between assessors and others on the assessment staff (e.g., administrators, role players, support staff)
 - vi. The degree of support provided to assessors in the form of observation and evaluation guides

It should be noted that length and quality of training are not synonymous. Precise guidelines for the minimum number of hours or days required for assessor training are difficult to specify. One

⁸ See footnote 5.

day of training may be sufficient for a well-structured assessment center using a small number of exercises, a qualified trainer, and carefully selected assessors. However, for the initial training of assessors who have no prior experience, considerably more training may be needed (e.g., two days of assessor training for each day of assessment center exercises). Assessors who have experience with similar assessment techniques in other programs may require less training. More complex assessment centers with varied formats of simulation exercises may require additional training; simple assessment centers may require less. In any event, assessor training is an essential aspect of an assessment program. The true test of training quality is assessor competence as described below.

3. Performance Guidelines and Certification—Each assessment center must have clearly stated performance guidelines for assessors contingent on the purpose of the assessment center and the various assessor roles. These performance guidelines must include, at a minimum, the ability to:

- a) Observe, record, and rate behavior in a standardized fashion
- b) Classify behaviors according to behavioral constructs
- c) Provide ratings that are calibrated in scale to the assessor team or an expert standard
- d) If applicable, report behavioral construct-relevant behaviors to the administrator or assessor team
- e) If assessors also serve as exercise administrators, administer exercises
- f) If assessors also serve as role players, objectively and consistently perform the role called for in interactive exercises
- g) If assessors are to provide feedback to assessees, deliver positive and negative behavioral feedback with supporting evidence in a manner that conveys concern/empathy and maintains or enhances assessees' self-esteem
- h) If assessors serve in a coaching role, establish clear expectations at the outset of the program (i.e., what behaviors can be expected from the assessor; what behaviors are expected of the assessee), motivate assessees, provide constructive and challenging feedback, and engage in coaching, developmental action planning, and goal setting
- i) If assessors are to provide feedback to line management, deliver clear, unambiguous and well-constructed feedback on assessees' strengths and developmental needs
- j) If assessors are to write reports for organizational decision-making or assessee feedback purposes, deliver reports that are clear, well-written, comprehensive, well-integrated, and proofread

Some measurement is needed to indicate that the individual being trained is capable of functioning as an assessor. This measurement may vary and could include data in terms of (1) accuracy and reliability of rating performance (defined with regard to either an "expert" standard or convergence with other assessors), (2) critiques of assessor reports, and (3) observation or shadowing of assessors in training by the assessment center staff. It is important that, prior to carrying out their actual duties, assessors' performance is evaluated to ensure that they are sufficiently trained to function as assessors and that such performance is periodically monitored to ensure that the skills learned in training are applied.

Each organization must be able to demonstrate that its assessors can meet minimum performance standards. This may require the development of additional training or other prescribed actions for assessors not meeting these standards.

The trainer of assessors must be competent to enable individuals to develop the assessor skills stated above and to evaluate the acquisition of these skills.

4. *Recency of Training and Experience*—The time between assessor training and initial service as an assessor must not exceed six months. If a longer period has elapsed, or even if experienced assessors do not have recent experience as an assessor (i.e., fewer than two assessment centers over two consecutive years), these (prospective) assessors should attend a refresher course or receive special coaching from a trained assessment center administrator. All assessors should be regularly checked for agreement and consistency in ratings, and provided refresher training as needed.

5. *Monitoring of Assessor Performance*—The performance of operational assessors should be periodically and systematically monitored, and action taken (via follow-up training, recertification, or de-certification and termination) when their ratings show a lack of reliability and/or accuracy; and/or when their behaviors show a lack of professionalism and adherence to performance standards set in training.

VIII. Training and Qualifications of Other Assessment Center Staff and Stakeholders

All staff members of any assessment center must be qualified and adequately trained to carry out their functions consistently, accurately, and effectively. “Other assessment center staff” refers to persons other than assessors who have contact with assessees in an assessment center, and who are responsible for aspects of assessment center operations (including communications, administration, training, validation, evaluation, and record keeping), including but not limited to the following roles:

1. *Assessment Center Administrator*—Referred to in some countries as the “assessment center manager,” this individual is the highest-level professional responsible for overseeing all assessment center operations. This individual may also be the assessment center developer/designer, may oversee the development and maintenance of the policy documents, and may be responsible for collecting ongoing validation/evaluation evidence. The assessment center administrator is responsible for the management of assessment center operations, logistics, assessor (and other staff) training, documentation, information sharing/confidentiality, risk management, and quality control.

2. *Assessment Center Coordinator*—Referred to in some countries as the “center administrator,” this individual plays an administrative support role, under the assessment center administrator/manager. This(These) individual(s) is(are) responsible for assessment center processes, scheduling, and logistics and may be responsible for administering simulation exercises and other assessment components, liaising with venue staff, collecting and managing documents and assessor ratings/reports, assembling scores for integration, preparing and proofreading feedback reports, and other duties as needed.

3. Role Player—A role player interacts with assessees in applicable behavioral simulation exercises in person, over the phone, or via other forms of communication technology. Role players must understand the overall assessment center context, as well as the simulation context in which they are playing a role. They must have a deep understanding of the demands of their role, as well as the importance of creating standardized responses toward various assessees. They must understand what behaviors are scripted and when they are permitted to act extemporaneously. They must also be well versed in program security/confidentiality, standards of professionalism, and issues of fairness and non-discrimination.

4. Organizational Decision Makers—To ensure the procedural fairness, integrity, and credibility of the assessment center program, assessment center administrators are strongly encouraged to provide training to those managers and/or organizational leaders who receive AC feedback reports/results to enhance the likelihood that data are appropriately interpreted and used. This is especially critical in those situations where ACs are used for administrative purposes (e.g., hiring, promotion, succession planning, etc.).

5. Other Roles—Other roles are carried out within an assessment center program, which may be carried out by individuals in the roles described above (as well as assessors), or may be carried out by separate individuals. These roles include:

- a) Those who communicate information about the assessment center program
- b) Persons who administer instructions to assessees
- c) Exercise facilitators
- d) Test administrators
- e) Persons who tabulate and report assessment center results
- f) Persons who write reports
- g) Coaches and other persons who are responsible for delivering feedback
- h) Other assistants and administrative support staff

Procedures for ensuring that all staff members can competently carry out their duties must be established. Reasonable steps should be taken to ensure all assessment center staff are appropriately trained and perform their roles effectively and consistently across participants and assessment sessions. Many of the recommendations provided in Section VII for assessors should be applied to the training and evaluation of such staff.

IX. Validation Issues

A major factor in the widespread acceptance and use of assessment centers is related directly to an emphasis on sound validation research. Numerous studies demonstrating the predictive validity of assessment center ratings have been conducted in a variety of organizational settings and reported in the professional literature. However, the historical record of this process's validity cannot be taken as a guarantee that a given assessment program (or new applications of existing programs) will or will not be valid.

Ascertaining the validity of assessment center ratings is a complicated technical process, and it is important that validation research meets both professional and legal standards. Research must be conducted by individuals knowledgeable in the technical and legal issues pertinent to validation procedures. In evaluating the validity of assessment center ratings, it is particularly important to document the process by which behavioral constructs are determined, their job relevance verified, and their linkages to the assessment components ensured.

Meta-analytic (also traditionally referred to as validity generalization) studies of assessment center research suggest that overall assessment ratings (OARs) show predictive validity across diverse settings. Such findings support the use of a new assessment center in a different setting if the job, exercises, assessors, and assessees in the new situation are similar to those studied in the validation research and if similar procedures are used to observe, report, and integrate the information. The meta-analytic studies substantiate the criterion-related validity of OARs, dimension ratings, and exercise ratings, but not necessarily the use of assessment center ratings for purposes other than the prediction of performance (or some other criterion, such as the diagnosis of training needs or to catalyze learning and development). The *Principles for the Validation and Use of Personnel Selection Procedures* and the *Standards for Educational and Psychological Testing* represent the definitive standards for validation. Assessment center practices must comply with these standards, as well as the professional testing/validation standards within the countries assessment centers are being carried out (e.g., the (UK) Council of the International Test Commission's International Guidelines for Test Use; the German DIN 33430 Requirements for Proficiency Assessment Procedures and Their Implementation; The Russian Standard of Psychodiagnostic Methods Requirements).

For assessment centers used for the sole purpose of training and development, in addition to the Guidelines provided herein, professional standards for training evaluation must be followed⁹. Evidence can be provided of improvements in such areas as cognitive (e.g., knowledge and concepts), skill-based (e.g., acquisition of new behaviors and abilities), and affective (e.g., attitude change, and motivational shifts) outcomes. Methods for compiling evidence must include sound evaluation procedures such as adequate samples of participants, research designs, measurement of relevant variables, controls, and statistical procedures.

X. Technology

It has become common practice to leverage information technology within assessment center practice in order to aid efficiency, lower costs, and provide a media-rich experience for assessees. Such techniques can aid in the elicitation, recording, rating, integrating, and feeding back of relevant behavioral information; they can influence assessees' perceptions of the program and organization; and they can assist organizations in aligning and connecting various talent management functions. Incorporation of technology must not result in the assessment program failing to comply with the essential elements of the assessment center method, *if* the new program is to continue to be referred to as an assessment center. As described earlier, for example, the new assessment program could no longer be referred to as an assessment center if

⁹ e.g., Kirkpatrick (1994); Quinnines & Tonidandel (2003); see Appendix D

the assessee no longer demonstrates overt behavior, or assessors no longer observe any overt behavior.

1. Examples of ways in which technology has been leveraged within assessment center programs include:

- a) Technology to aid administrative tasks such as scheduling of assessees, assessors, and role players; carrying out assessor ratings, reporting, and integration (automatically or to aid discussion among assessors); and final reporting and feedback
- b) The use of video to aid delivery of instructions, administration of an exercise, assessment, feedback, and assessor training
- c) The use of multi-media tools to deliver simulation content over internal networks and the Internet

2. Incorporating such technologies into assessment center programs, especially when delivered, even in part, over the Internet, presents a number of legal and ethical challenges that must be addressed. A number of professional and legal guidelines must be consulted when carrying out such practices. The following lists some of these guidelines, as well as what their key recommendations suggest for assessment center operations:

- a) Guidelines provided by the American Psychological Association's *Taskforce on Psychological Testing on the Internet*
 - i. Institute a process for confirming the identification of assessees who may be assessed remotely.
 - ii. Use a multi-server configuration such that test/assessment content, data, scoring, and reporting information are stored on different servers, and that data (and back-ups) are stored on servers residing behind a secure firewall.
 - iii. Institute methods to discourage and disable (to the extent possible) the copying or printing of secure materials.
- b) International Test Commission's *Guidelines on the Security of Tests, Examinations, and Other Assessments* and *Guidelines on Computer-Based and Internet Delivered Testing*
 - i. Choose delivery methods (i.e., open access, controlled, supervised, managed) according to the level of control and security implied by the purpose of the assessment center. Assessment centers for selection and promotion require the most control and security.
- c) International laws and policies involving data privacy
 - i. If an assessment center uses remote assessment, and receives data from assessees in other countries, the program must comply with any data protection laws that might exist in those countries, as data have crossed international boundaries in this case. For example, see the European Union Directive on Data Protection and the U.S. Safe Harbor Privacy Principles.
- d) Assessment centers incorporating technology may also have to make special considerations involving accommodations for persons with disabilities, and for persons who may, for a variety of reasons, have lower than average computer literacy (when such

literacy is not an essential job requirement). The Web Accessibility Initiative (WAI) has been implemented by the World Wide Web Consortium to provide guidelines for ensuring Internet accessibility.

3. In addition, organizations and other applicable entities must also consider how the delivery of assessment center content over computer networks might threaten the standardization of the assessment context. Whereas this may be less of a threat in developmental contexts, if the purpose of the assessment center is personnel selection or promotion, standardization is paramount. Breakdowns in standardization can occur when users' experiences differ due to differences in:

- a) Operating systems
- b) Hardware
- c) Internet connection quality, speed, and bandwidth
- d) Browser compatibility and configuration
- e) Computer screen size and resolution
- f) Sound quality
- g) Keyboard type (e.g., onscreen touch keys may block visual display)
- h) Mouse capabilities (e.g., touch pads, differences in right click options)
- i) Working conditions
- j) The presence of other individuals nearby
- k) Access to network-delivered assessment center components (i.e., when "high"- and "low"-tech versions of assessment center components are offered simultaneously)

4. In addition, the security of assessment center content must be considered, as vulnerabilities may occur if delivering assessment content over the Internet. Assessment center developers must carefully consider these issues, identify potential vulnerabilities, and assess risks prior to making decisions about the incorporation of technology. Further, the assessment center policy document must include a section on the use of technology, data security, and all appropriate protocols (see Section VI). If different vendors are used to deliver assessment content, these considerations must also be made with regard to their deliveries.

5. Further, assessment center developers must consider whether the use of technology enhances or detracts from the fidelity of the assessment process. If the behaviors required to access and carry out the assessment are not essential to the focal job, then the use of technology may threaten the validity of the assessment center (e.g., by disadvantaging individuals who lack computer literacy or experience when these are not demonstrable job requirements). Simple training, tutorials, or help resources in lay language should be provided to assesseees when the technology is likely to be unfamiliar.

XI. Ethics, Legal Compliance, and Social Responsibility

The various stakeholder groups inherent to an assessment center program (e.g., assesseees, assessors, organizations, consultants) have various rights and responsibilities. Some have been articulated earlier in this document (e.g., qualifications of assessors, role players, and administrators; clearly stated purpose of program and use of data only for that purpose; professional standards; legal protections). Here we include additional ethical considerations.

1. Informed Participation—The organization is obligated to make an announcement prior to the assessment so that assessees will be fully informed about the program. This information should be made available in writing prior to assessment events. While the information provided will vary across organizations, the following basic information must be provided to all assessees:

- a) **Objective(s)**—The objective(s) of the program and the purpose of the assessment center. Depending on the purpose of the assessment center, the organization may choose to disclose the behavioral constructs measured and the general nature of the exercises prior to the assessment.
- b) **Selection**—How individuals are selected to participate in the assessment center
- c) **Choice(s)**—Any options the individual has regarding the choice of participating in the assessment center as a condition of employment, advancement, development, etc.
- d) **Staff**—General information on the assessment center staff and the role of the assessors, including composition, relevant experience, and assessor training
- e) **Materials**—What assessment center materials completed by the individual are collected and maintained by the organization
- f) **Results**—How the assessment center results will be used, what recommendations will be made, and how long the assessment results will be maintained on file
- g) **Feedback**—When, how (e.g., written, face-to-face, technology-aided), and what kind of feedback (e.g., by behavioral construct, by exercise, by a combination) will be given to the assessees
- h) **Development**—Mechanisms for follow-up support and monitoring, if any (e.g., coaching, training, mentoring; top management/supervisory support)
- i) **Alignment**—If applicable, how the assessment center results will be aligned with organizational strategy and culture, and how the results will be integrated with other human resource management functions
- j) **Reassessment**—The procedure for reassessment (if any)
- k) **Access**—Who will have access to the assessment center reports (and audio and/or video files, if applicable) and under what conditions
- l) **Contact**—The contact person responsible for the records and where the results will be stored or archived
- m) **Electronic Data Security**—Information on the security protocol for all electronic data files and compliance with relevant legal and professional standards pertaining to electronic data management and access to information

2. Assessee Rights—Assessment center activities typically generate a large volume of data on an individual who has gone through a center. These assessment data come in many forms and may include observer notes, reports on performance in the exercises, assessor ratings, peer ratings, paper-and-pencil or computerized tests, video files, and final assessment center reports. This list, while not exhaustive, does indicate the extent of information about an individual that may be collected. The following lists important practices with regard to assessee rights:

- a) Assesseees are to receive feedback on their assessment center performance and should be informed of any recommendations made.
- b) Assesseees who are members of the organization have a right to read any formal, summary, written reports concerning their own performance and recommendations that are prepared and made available to management. Applicants to an organization should be provided with, at a minimum, the final recommendation made with regard to their individual case, and if possible and requested by the applicant, the reason for the recommendation.
- c) To ensure test security, assessment center exercises and assessor reports on performance in particular exercises are exempted from disclosure, but the rationale and validity data concerning ratings of behavioral constructs and the resulting recommendations should be made available upon request of the individual assessee.
- d) The organization must inform the assessee what records and data are being collected, maintained, used, and disseminated. Assesseees must be informed if their activities in the assessment center are being recorded as well as if such recordings or other personal data will be transferred across national borders or over the Internet.
- e) If the organization decides to use assessment results for purposes other than those originally announced and that can have an impact on the assessee, then the assessee must be informed and consent obtained.

3. Copyrights and Intellectual Property—In addition, assessment center materials (e.g., simulations and other exercises, rating scales, assessor training materials) often are intellectual property protected by international copyright laws. Respect for copyrights and the intellectual property of others must be maintained under all circumstances.

4. Data Protection—The assessment center program must also comply with any relevant data protection laws governing the regions in which assessment is being carried out (e.g., the UK Data Protection Act; the U.S. Freedom of Information Act; the European Union Directive on Data Protection; South Africa’s Protection of Personal Information Bill; The U.S. Safe Harbor Privacy Principles). See also Section X.

5. Compliance with Relevant Employment Laws and Regulations—As stated in Sections VI and X, assessment center design, validation, implementation, and documentation must be carried out in compliance with laws and statutes existing both in an organization’s/agency’s home locale,

state, province, or nation—as well as in the local, state, province, or nation where the assessment center program is delivered. This includes preventing unfair discrimination against protected groups (such as native/aboriginal people, racial groups, religious groups, and those protected on the basis of age, gender, disability, sexual orientation, etc.).

XII. Conducting Assessment Centers Across Cultural Contexts

1. It is common for single assessment center programs to cross both cultural and national boundaries. In some situations it may be necessary to adapt many assessment center practices to the local culture in which an assessment center is deployed. In other situations it may be decided that an assessment center program requires standardization across all regions to which it is being deployed.

Practitioners using assessment center methods beyond the boundaries of the country/region from which the assessment center program originated, or with members of multiple cultural groups, must determine the extent to which cultural accommodations may be necessary. Such an analysis includes the collection of evidence that the validity and applicability of the assessment center have not been compromised at either the design or implementation phase.

2. A range of contextual factors must be considered during such a process, including:

- a) The extent of commonality in the cultural, business, legal, and socio-political environments between countries (e.g., cultural beliefs and behaviors, local business laws)
- b) Differences in national guidelines set by local professional associations (see Section XIII)
- c) Commonality of behavioral constructs critical for job success
- d) Commonality in performance standards/behaviors required for job success
- e) The extent of commonality of the business models between the organizations across which the assessment center/method is being adapted (i.e., overall business strategy, vision, values, and practices)
- f) Degree of centralized vs. decentralized (i.e., local) control across branches of the organization
- g) Whether comparison statistics (e.g., normative comparisons) are required to interpret the results across locations
- h) Where applicable, the extent to which personnel need to be transferred across countries/cultures

3. When assessment center programs are designed to be culturally specific, the following aspects may be considered for modification based on each culture in which the assessment center is used:

- a) Selection of performance criteria
- b) Criteria for occupational success
- c) How behavioral constructs are defined
- d) Types of exercises used and their content
- e) Selection of behavioral cues
- f) Selection of assessors
- g) Level of directness and confrontation employed by assessors and role players
- h) Assessor training processes and content, possibly including explicit training on cultural differences and norms when the assessors may not share a common culture with the assesseees
- i) Reporting practices
- j) Feedback delivery process, format, medium (i.e., written or spoken), and level of detail
- k) Changes in currency; kilometers vs. miles; people and place names – even if the language of the assessment center remains the same

4. In contrast, several aspects of the assessment center process must remain standardized, even when the process has been culturally adapted. Features that must remain the same across cultures include:

- a) Inclusion of behavioral observation
- b) Training of assessors in the process of behavioral observation
- c) Classification and rating of behavior
- d) A systematic process of integrating evaluations across exercises, behavioral constructs, and assessors

5. In adapting assessment centers for use across regions or for different cultural populations, it might also be necessary to translate all program documents, stimuli, rating tools, and report templates to one or more alternate languages. Materials must be translated, back translated, and quality checked according to professional standards (e.g., *International Test Commission*

Guidelines for Translating and Adapting Tests). If an assessment center operates in multiple languages, evidence needs to be collected and documented as to the equivalence of the alternate forms. The assessors employed in such multi-cultural assessment centers must also receive appropriate training in dealing with relevant cross-cultural issues.

6. It is also important to comply with regulations surrounding the transfer of data across national boundaries (see Sections X and XI).

XIII. National Assessment Center Guidelines

In addition to the International Guidelines presented herein, some countries have developed local guidelines to guide (in parallel with the International Guidelines) assessment center practices in their specific national contexts. Examples of national standards include:

1. **Germany**—Arbeitskreis Assessment Center. (2004). *German Standards for Assessment Center Operations*. http://www.arbeitskreis-ac.de/images/attachments/149_german-ac-standards-2004-usl.pdf

2. **Indonesia**—Indonesian Task Force on Assessment Center Guidelines. (2002). *Ethical Guidelines for Assessment Center Operations*. Daya Dimensi Indonesia.

3. **Russia**—Personnel Assessment Federation. (2013). *Russian Standard for Assessment Centers*. http://hrdevelopment.ru/federations/f_assess/03.pdf

4. **South Africa**—Assessment Centre Study Group. (2007). *Guidelines for Assessment and Development Centres in South Africa* (4th ed.). http://www.acsg.co.za/ac_information/guidelines/acdcguidelines.pdf

5. **Switzerland**—(2007). *AC Standards Swiss Assessment*. http://www.swissassessment.ch/files/ac_standards_swiss_assessment.pdf

6. **United Kingdom**—British Psychological Society’s Division of Occupational Psychology. (2015). *The Design and Delivery of Assessment Centres*.

The International Guidelines present broad, universal guidance to which all nations (and national standards) need comply, whereas the national standards provide more detailed recommendations based on local legal and cultural environments. The national standards listed above have been reviewed and deemed consistent with the spirit and content of these International Guidelines.

Appendix A

Past Taskforce Members

1975 1st Edition

Albert Alon	Miracle Food Mart
Douglas W. Bray	AT&T
William C. Byham	Development Dimensions International, Inc.
Donald L. Grant	AT&T and University of Georgia
Lowell W. Hellervik	University of Minnesota
James R. Huck	AT&T—Michigan Bell Telephone Company
Cabot L. Jaffee	Assessment Designs, Inc.
Alan I. Kraut	International Business Machines
John H. McConnell	American Management Association
Joseph L. Moses (Chair)	AT&T
Leonard W. Slivinski	Public Service Commission (Canada)
Thomas E. Standing	Standard Oil of Ohio
Edwin Yager	Consulting Associates

1979 2nd Edition

Albert Alon	Miracle Food Mart
Dale Baker	U.S. Civil Service Commission
Douglas W. Bray	AT&T
William C. Byham	Development Dimensions International, Inc.
Steven L. Cohen	Assessment Designs, Inc.
Lois A. Crooks	Educational Testing Service
Donald L. Grant	University of Georgia
Milton D. Hakel	Ohio State University
Lowell W. Hellervik	University of Minnesota
James R. Huck	Human Resources International
Cabot L. Jaffee	Assessment Designs, Inc.
Frank M. McIntyre	Consulting Associates
Joseph L. Moses (Chair)	AT&T
Nicky B. Schnarr	International Business Machines
Leonard W. Slivinski	Public Service Commission (Canada)
Thomas E. Standing	Standard Oil of Ohio
Edwin Yager	Consulting Associates

1989 3rd Edition

Virginia R. Boehm	Assessment & Development Associates
Douglas W. Bray (Co-Chair)	Development Dimensions International, Inc.
William C. Byham	Development Dimensions International, Inc.
Anne Marie Carlisi	BellSouth
John J. Clancy	Clancy & Associates
Reginald Ellis	Canadian National Railway
Joep Esser	Mars B.V.

Fred Frank
Ann C. Gowdey
Dennis A. Joiner
Rhonda Miller
Marilyn Quaintance-Gowing
Robert F. Silzer
George C. Thornton III (Co-Chair)

2000 4th Edition

William C. Byham
Richard Flanary

Marilyn K. Gowing
James R. Huck
Jeffrey D. Kudisch
David R. MacDonald (Chair)
Patrick T. Maher

Jeroen J.J.L. Seegers
George C. Thornton III

2008/9 5th Edition

William C. Byham
Anuradha Chawla
Alyssa Mitchell Gibbons
Sebastien Houde

Dennis Joiner
Myungjoon Kim
Diana Krause
Jeffrey D. Kudisch
Cara Lundquist
David R. MacDonald
Patrick T. Maher

Doug Reynolds (Co-Chair)
Deborah E. Rupp (Co-Chair)
Deidra J. Schleicher
Jeroen J.J.L. Seegers
George C. Thornton III

Electronic Selection Systems Corporation
Connecticut Mutual
Joiner & Associates
New York Power Authority
U.S. Office of Personnel Management
Personnel Decisions, Inc.
Colorado State University

Development Dimensions International, Inc.
National Association of Secondary School
Principals
U.S. Office of Personnel Management
Human Resources International
University of Southern Mississippi
Steelcase Inc.
Personnel & Organization Development
Consultants, Inc.
Assessment & Development Consult
Colorado State University

Development Dimensions International, Inc.
RHR International
Colorado State University
University of Guelph & Royal Military
College of Canada
Dennis A. Joiner & Associates
Korean Psychological Testing Institute
DHV Speyer
University of Maryland
Southern California Edison
Steelcase Inc.
Personnel & Organization Development
Consultants, Inc.
Development Dimensions International, Inc.
University of Illinois at Urbana-Champaign
Purdue University
Right Management Benelux
Colorado State University

Appendix B

Glossary of Relevant Terms

Assessee: An individual who is assessed in an assessment center. Sometimes referred to as “participant,” “delegate,” or “candidate.”

Assessment Center: A process employing multiple assessment components, multiple assessors, and the use of simulation exercises to produce judgments regarding the extent to which an assessee displays proficiency on selected behavioral constructs.

Assessment Center Administrator: The highest-level professional responsible for overseeing all assessment center operations. This individual is responsible for the management of assessment center operations, logistics, assessor (and other staff) training, documentation, risk management, and quality control. Also referred to as the “assessment center manager.”

Assessment Center Component: One of the multiple sub-assessments comprising an assessment center. Assessment center components are most often behavioral simulation exercises. Other components might include tests, interviews, and other forms of measurement/assessment.

Assessment Center Coordinator: An individual who plays an administrative support role, under the assessment center administrator/manager, and is responsible for assessment center processes, scheduling, and logistics and may be responsible for administering simulation exercises and other assessment components, liaising with venue staff, collecting and managing documents and assessor ratings/reports, assembling scores for integration, preparing feedback reports, and other duties as needed. Referred to in some countries as the “center administrator.”

Assessment Center Manager: See Assessment Center Administrator.

Assessor: An individual trained to observe, record, classify, and make accurate and reliable judgments about the behaviors of assessees participating in an assessment center.

Assessor Training: Training for assessors prior to service in an AC, including how to carry out all assessor duties, as well as an evaluation of rating accuracy/reliability.

Behaviorally-Anchored Rating Scale (BARS): Examples of behavioral incidents describing effective, average, and ineffective performance on a behavioral construct, listed as examples for points on a graphic rating scale.

Behavioral Checklist: Lists of behaviors that an assessee must show to demonstrate proficiency in completing an exercise.

Behavioral Construct: Used in these Guidelines to refer more generally to the focal constructs assessed in an assessment center, which may include dimensions; competencies; knowledge, skills, and abilities (KSAs); performance on tasks; or performance in roles.

Behavioral Construct-by-Assessment Component Matrix: A matrix, decided upon after job analysis or competency modeling has been completed and behavioral constructs (e.g., dimensions) identified, which maps what assessment components (e.g., tests, simulation exercises) will assess which behavioral constructs. The matrix should illustrate how each behavioral construct will be assessed in multiple assessment components.

Behavioral Cue: Predetermined statements or stimuli (e.g., statements made by role players, or written statements within provided documentation) that are consistently presented across assesseees to elicit behaviors related to specific job-related behavioral constructs. Also referred to as behavioral prompts.

Behavioral Dimensions: See Dimensions.

Behavior Observation Scale (BOS): Raters indicate the frequency (e.g., on a scale from “almost never” to “almost always”) that an assessee has demonstrated a list of effective and ineffective behaviors related to a behavioral construct.

Center Administrator: See Assessment Center Coordinator.

Competency: See Dimension.

Competency Modeling: Method of collecting and organizing information about the characteristics and qualities individuals need to effectively carry out job duties. Methods may be identical to job analysis methods, although traditionally there is at least some focus on the broader organizational context, including the organization’s strategy, culture, and vision. See Job Analysis.

Component: See Assessment Center Component.

Consensus Discussion: See Integration Discussion.

Development: Improvement in any proficiency set as a desired outcome of the assessment center.

Development Center (DC): A broad term that encompasses *both* assessment centers solely designed to diagnose development needs, and for assessment centers designed to catalyze development within the course of the assessment center program. Not to be confused with the term “developmental assessment center (DAC)”, which is used herein to refer only to the latter.

Developmental Assessment Center (DAC): An assessment center designed for the purpose of directly developing/improving assessees on behavioral constructs of interest.

Diagnosis: An analysis of the strengths and weaknesses of each individual assessee on the behavioral constructs being assessed.

Dimension: A constellation or group of behaviors that are specific, observable, and verifiable that can be reliably and logically classified together and that relate to job success. Sometime used synonymously with “competencies.”

Dimension-by-Assessment Matrix: See Behavioral Construct-by-Assessment Component Matrix.

Dimension-by-Exercise Matrix: See Behavioral Construct-by-Assessment Component Matrix.

Feedback: Information comparing actual performance to a standard or desired level of performance; and the delivery of this information to relevant stakeholders (e.g. the assessee, management, HR).

Fidelity: The extent to which an assessment center simulation requires the assessee to actually display job-relevant behaviors related to one or more select behavioral constructs. Fidelity is related to the realism of the simulation as compared to an actual job situation or task. It also refers to the similarity between the format of the assessment (e.g., computerized) and behaviors carried out on the job.

Frame of Reference Training: Assessor training on the targeted behavioral constructs, aimed at improving the reliability and validity of behavioral construct ratings; focused on ensuring that assessors have the same understanding of the meaning of both behavioral constructs and the level of proficiency expected, demonstrated by inter-assessor agreement and/or agreement with expert ratings.

Integration: Methods for combining behavioral observations and ratings from multiple assessors, behavioral constructs, and/or exercises; may be accomplished by discussion or statistical combination.

Integration Discussion: A method of aggregation in which assessors meet to talk about observations and ratings made within the assessment center.

Job Analysis: The process used to determine the tasks and KSAs linked to success or failure in a job, job role, or job grouping (as well as their linkages). The process typically consists of a combination of techniques to collect job information, such as interviews with and observations of incumbents, interviews with upper-level managers/executives and other subject matter experts, review of existing job documentation (job descriptions, training manuals, etc.), and surveys.

Job Families: Groups of occupations based upon work performed, skills, education, training, and credentials.

Job Role: A pattern of behaviors that is associated with the demands or requirements of a given situation.

Knowledge, Skills, and Abilities (KSAs): An inclusive array of human characteristics, sometimes known as dimensions or competencies.

Overall Assessment Rating (OAR): A summary evaluation of an assessee's overall performance in an assessment center, based on a consensus judgment among assessors or a statistical aggregation of ratings on narrower components of performance such as behaviors, dimensions, tasks, or exercises.

Prediction: A judgment made about the future success of individuals who have been assessed.

Psychometric Tests: Term used in some countries to refer to tests that do not involve direct behavioral observation or naturalistic responding. Often referred to traditionally as "paper-and-pencil" tests, these include measures such as cognitive ability tests and personality inventories.

Reliability: The extent to which a measurement process yields the same scores (given identical conditions) across repeated measurements.

Role Player: An individual responsible for interacting with assessees in applicable behavioral simulation exercises in person, over the phone, or via other forms of communication technology.

Simulation: See Simulation Exercise.

Simulation Exercise: An exercise or technique designed to elicit behaviors related to behavioral constructs of performance on the job, requiring the assessee to respond behaviorally to situational stimuli.

Split Ratings: When assessment center operations allow for the meaningful interpretation of varied performance relevant to a particular behavioral construct across different simulation exercises, operationally shown, for example, by a relatively high rating on a behavioral construct for one type of exercise and a relatively low rating on the same behavioral construct for a different type of exercise.

Task: A segment of work to be accomplished, including the setting, behavior called for, and the outcome desired.

Validity: The extent to which the inferences one desires to make based on scores produced by a measurement tool or process, such as an assessment center, are defensible. Forms of validity evidence might be measured (e.g., construct, content, face, criterion-related, social/consequential) depending upon the questions being explored and the tool or process being investigated.

Appendix C

Relevant Professional Guidelines*

* *These guidelines have been developed to be compatible with the following:*

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. Washington, D.C.: American Educational Research Association.

American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *American Psychologist*, 57, 1060-1073.

American Psychological Association Council of Representatives. (1990). *APA guidelines for providers of psychological services to ethnic, linguistic, and culturally diverse populations*. Boston: American Psychological Association.

American Psychological Association Public Interest Directorate and Council of Representatives. (2002). *Guidelines on multicultural education, training, research, practice, and organizational change for psychologists*. Washington, D.C.: American Psychological Association. Available at:
<http://www.apa.org/pi/oema/resources/policy/multicultural-guidelines.aspx>

Council of the International Test Commission. (2012). *International guidelines for test use: Version 2000*. Leicester, UK: British Psychological Society.

Equal Employment Opportunity Commission, Civil Rights Commission, Department of Labor, & Department of Justice. (1978). *Uniform Guidelines on Employee Selection Procedures*, *Federal Register*, 43 (166), 38290-38309.

International Task Force on Assessment Center Guidelines (2009). Guidelines and Ethical Considerations for Assessment Center Operations. *International Journal of Selection and Assessment*, 17, 243-253.

International Test Commission. (2014). *International guidelines on the security of tests, examinations, and other assessments*. www.intestcom.org.

International Test Commission (2005). *International guidelines on test adaptation*. www.intestcom.org.

International Test Commission (2005). *ITC guidelines on computer-based and internet delivered testing*. www.intestcom.org.

Naglieri, J. A., Drasgow, F., Schmit, M., Handler, L., Prifitera, A., Margolis, A., & Velasquez, R. (2004). Psychological testing on the Internet: New problems, old issues. *American Psychologist*, 59, 150-162.

Society for Industrial and Organizational Psychology. (2003). *Principles for the validation and use of personnel selection procedures* (4th ed.). Bowling Green, OH: Author.

U.S.-E.U. Safe Harbor. (2000). *Safe Harbor*. <http://export.gov/safeharbor/index.asp>. Accessed May 20, 2014.

World Wide Web Consortium. (1997). Web Accessibility Initiative. <http://www.w3.org/WAI/>. Accessed May 20, 2014.

Appendix D

Key Sources/Recommended Readings

Books

Jackson, D.J.R., Lance, C.E., & Hoffman, B.J. (2012). *The psychology of assessment centers*. New York: Routledge.

Povah, N. & Thornton III, G.C. (Eds.). (2011). *Assessment centres and global talent management*. Farnham, UK: Gower.

Thornton III, G. C., & Mueller-Hanson, R. A. (2003). *Developing organizational simulations: A guide for practitioners and students*. New York, NY: Psychology Press.

Thornton III, G. C. & Rupp, D. E. (2005). *Assessment centers in human resource management: Strategies for prediction, diagnosis, and development*. Mahwah, NJ: Lawrence Erlbaum.

Thornton III, G. C., Rupp, D. E., & Hoffman, B. (2014). *Assessment center perspectives for talent management strategies*. New York: Routledge.

Key Articles, Including Meta-Analyses

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.

Bray, D. W., & Grant, D. L. (1966). The assessment center in the measurement of potential for business management. *Psychological Monographs: General and Applied*, *80*, 1-27.

Brummel, B., Rupp, D. E., & Spain, S. (2009). Constructing parallel simulation exercises for assessment centers and other forms of behavioral assessment. *Personnel Psychology*, *62*, 135-170.

Byham, W.C. (1970). Assessment centers for spotting future managers. *Harvard Business Review*, *48*, 150-160, plus appendix.

Dean, M. A., Roth, P. L., & Bobko, P. (2008). Ethnic and gender subgroup differences in assessment center ratings: A meta-analysis. *Journal of Applied Psychology*, *93*, 685-691.

Lanik, M., & Gibbons, A. M. (2011). Guidelines for cross-cultural assessor training in multi-cultural assessment centers. *The Psychologist Manager Journal*, *14*, 221-246.

Lievens, F., Tett, R.P., & Schleicher, D.J. (2009). Assessment centers at the crossroads: Toward a reconceptualization of assessment center exercises. In J.J. Martocchio & H. Liao (Eds.),

Research in personnel and human resources management (pp. 99-152). Oxford, UK: JAI Press.

Meriac, J. P., Hoffman, B. J., & Woehr, D. J. (2014). A conceptual and empirical review of the structure of assessment center dimensions. *Journal of Management*. DOI: 0149206314522299.

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.

Other Relevant References

Haertel, E. H. (2006). Reliability. In R. L. Brennan (Ed.), *Educational Measurement, 4th edition* (pp. 65-110). Westport, CT: American Council on Education and Praeger Publishers.

Kirkpatrick, D. L. (1994). *Evaluating training programs: the four levels*, San Francisco, CA: Berrett-Koehler

Quinones, M. A., & Tonidandel, S. (2003). Conducting training evaluation. In J. E. Edwards, J. C. Scott, & N. S. Raju, N. S. (Eds.) *The human resources program-evaluation handbook*. Thousand Oaks, CA: Sage.

Putka, D. J., & Sackett, P. R. (2010). Reliability and validity. In J. L. Farr & N. T. Tippins (Eds.), *Handbook of Employee Selection* (pp. 9-49). New York: Routledge.