A SURVEY OF ASSESSMENT CENTER PRACTICES IN ORGANIZATIONS IN THE UNITED STATES

ANNETTE C. SPYCHALSKI, MIGUEL A. QUIÑONES
Rice University
BARBARA B. GAUGLER
Personnel Decisions, International
KATIA POHLEY
University of Hamburg

Two hundred fifteen organizations in the United States provided information about multiple aspects of their assessment centers, including design, usage, and their adherence to professional guidelines and research-based suggestions for the use of this method. Results reveal that centers are usually conducted for selection, promotion, and development purposes. Supervisor recommendation plays a sizable role in choosing center participants. Most often, line managers act as assessors; they typically arrive at participant ratings through a consensus process. In general, respondents indicate close adherence to recommendations for center design and assessor training. Recommendations involving other practices (e.g., informing participants, evaluating assessors, validating center results) are frequently not followed. Furthermore, methods thought to improve predictive validity of center ratings are underutilized. Variability in center practices according to industry and center purpose was revealed. We encourage practitioners to follow recommendations for center usage, and researchers to work to better understand moderators of center validity.

Since their first industrial application in 1956, assessment centers have been used in numerous organizations, both in the public and private sector. Assessment centers are conducted for various purposes, including employee selection, early identification of managerial talent, development planning, identification of training needs, promotion, and management succession. Assessment center practices in the United States vary dramatically; some think that a "typical" assessment center does not exist (Bender, 1973).

We would like to thank Ginny Boehm, Doug Bray, Bill Byham, Donna Denning, Bill Howell, Vicki Kaman, Elaine Sloas, George Thornton, and Alan Wolfson for their helpful comments on an early draft of the survey. We especially appreciate Cindy Benson's assistance in designing the original survey and developing our mailing list.

Correspondence and requests for reprints should be addressed to Annette Spychalski or Miguel Quiñones at Rice University Psychology Department, P.O. Box 1892, Houston, TX, 77251.

COPYRIGHT © 1997 PERSONNEL PSYCHOLOGY, INC.
Despite the lack of uniformity in assessment center practices, professional guidelines regarding assessment center development and use, the *Guidelines and Ethical Considerations for Assessment Center Operations* have been written (Task Force on Assessment Center Guidelines, 1989). These guidelines, developed and endorsed by practitioners who specialize in the use of the assessment center method, are based on theory, research, and practice in the field. The *Guidelines* delineate the essential elements and activities of an assessment center and describe the kinds of activities that do not constitute an assessment center. Detailed recommendations for specific aspects of assessment center operations are also provided, such as the content and length of assessor training, the information assesses should receive before participating, data usage and storage, and methods for ascertaining the validity of an assessment center. However, the *Guidelines* only suggest how assessment centers should be developed and used; adherence to the prescriptions put forth in the *Guidelines* by organizations is unknown.

In addition to suggesting ways of constructing and using assessment centers, researchers have identified attributes of assessment centers that affect the validity of the data gathered with this method. For example, characteristics such as type of assessor, assessor familiarity with assesses, and source of performance rating have been found to moderate assessment center validity (Gaugler, Rosenthal, Thornton, & Benton, 1987; Schmitt, Schneider, & Cohen, 1990). The way in which users of assessment centers incorporate each of these factors into their centers is not well understood.

Previous surveys about assessment center usage have addressed organizations' adherence to the then-current *Guidelines* to a limited extent. The last published survey about assessment center practices in the United States appeared in 1982 (Fitzgerald & Quaintance, 1982). The authors found that assessment centers were widely used by state and local government agencies, most frequently for promotional purposes. The most typical exercises used across the centers were in-baskets, leaderless group discussions, role play exercises, and interviews. In general, very few women and minorities participated in assessment centers. Another general practice was evaluating assessment centers with content validation studies. The authors' greatest concern based on the respondents' data involved the brevity of some of the assessment centers (2 hours in some cases) and their assessor training programs (1.5 hours in some cases).

Prior to Fitzgerald and Quaintance's 1982 study, Bender (1973) published a survey of assessment center practices in the United States and Canada. He highlighted a need for more frequent criterion-related validation studies, more thorough training for assessors, and more thought
to the way in which data collected in the center were used. Another survey published by Alexander (1979) found that assessment center results were employed for short-term rather than long-term uses (e.g., for promotion decisions rather than for long-range development). Furthermore, despite the time and expense involved in evaluating candidates, assessment center data were given less weight in making promotional decisions than were supervisor recommendations and regular performance appraisal instrument results. Alexander concluded that, despite their potential usefulness, the future for assessment centers was uncertain.

These surveys share a number of shortcomings. First, the sample size on which they were based was somewhat limited (from 6 to 115 companies). Second, their samples tended to heavily represent a particular type of respondent (e.g., government organization, client from a single consulting firm). Furthermore, the surveys were rather modest in the amount of data they collected: Respondents were asked a limited number of questions regarding only certain aspects of assessment center usage. Therefore, it seemed that a thorough update of assessment center practices was badly needed. It was hoped that this study could tap a sample of companies larger and more diverse than had previous surveys, and that it would produce extensive information about multiple aspects of assessment center design and usage. It was also hoped that the results would indicate the extent to which assessment center users had adopted practices that were consistent with those prescribed by the Guidelines.

The major purpose of this study was to describe current assessment center practices in organizations in the United States and to evaluate their conformity to existing guidelines for assessment center operation. Another goal was to describe the unique features of assessment centers used in various industries and those administered for various purposes. Prevalence of characteristics that have been found to contribute to the predictive validity of assessment center data was of particular interest. The results of questions most directly related to these issues are presented in comparison to recommendations stipulated by the Guidelines concerning assessment center development and usage.

Method

Sample

Several sources were used to identify all possible companies using assessment centers: client lists of several consulting firms that design and administer assessment centers for organizations, lists of participants and attendees at the International Congress on the Assessment Center
Method and the International Association of Chiefs of Police Assessment Center conference for the 3 years prior to survey distribution, and authors of articles and books on the assessment center method published during the 10 years prior to survey distribution. Surveys were mailed to the 655 companies and individuals identified. Survey data collection lasted 6 months (i.e., the first half of 1990).

Procedure

A questionnaire was mailed to the representative from each company thought to be in charge of assessment center development and administration. Company representatives were told that their participation in the study would help the authors "find [sic] if there is such a thing as a typical assessment center, and further, explore [sic] the outer bounds of assessment center technology and practice." Respondents were told that the survey results would provide several benefits, including: establishing valuable baseline information on assessment center practices and providing upper management with accurate information about assessment center practices; determining whether organizational practices, standards, and policies regarding assessment centers are consistent with established professional guidelines; identifying where research efforts need to be focused. Company representatives were asked to complete the survey describing the assessment center employed in their organization. In the event that a company used more than one assessment center, respondents were asked to describe the assessment center used to assess the largest number of people. (Due to the length of the survey, the authors felt that it would be too burdensome to request descriptions of multiple assessment centers.) Company representatives were assured that their confidentiality, as well as that of their organization, would be respected: Survey results would be presented and discussed in aggregate form only. Respondents were asked to return the survey in the self-addressed, postage-paid envelope that was enclosed. A reminder note was sent to the company representatives approximately 6 weeks after the initial questionnaire.

Survey Description and Development

In developing the questionnaire, the published and unpublished literature on the assessment center method was reviewed. Then a preliminary draft of the survey was sent to several prominent practitioners and researchers in the field to evaluate the comprehensiveness of its content, appropriateness of response categories, clarity of wording, length,
and so forth. The final survey was 15 pages long and contained 120 questions, both multiple choice and open-ended. The bulk of the questions inquired about assessment center development and evaluation, center usage, or data processing and usage. Many multiple choice questions instructed the respondents to “check all that apply.” In these cases, individuals could endorse up to 44 responses to a single question.

Results

Respondents

Of the 655 surveys that were sent, 47 were returned by the post office for lack of a forwarding address. Of the 291 questionnaires that were received (response rate = 48%), 76 were accompanied by responses that there was no assessment center currently used in the organization. Our findings reflect responses from the 215 individuals who completed the surveys.

The demographic characteristics of the respondents and their organizations are described below. The organizations surveyed represented government (34.6%), manufacturing (18.7%), service (13.6%), education (12.1%), consulting (4.7%), sales (4.7%), and banking (3.3%) industries (8.4% unspecified type of organization or organization that represents more than one type). The size of the organizations ranged considerably, from 1 to 800,000 employees ($M = 16,273$ employees; $SD = 65,946$ employees; Median = 2,600 employees). Some of the organizations had used assessment centers for only a brief period of time (e.g., 1 year), while others had had them in place for up to 35 years ($M = 8.1$ years; $SD = 6.0$ years; Median = 8.0 years). Most of our respondents held multiple roles in their assessment centers: most served as administrators of assessment centers in their organizations (86.8%) and developed their assessment center (58.0%); some conducted research on their assessment centers (29.0%), served as assessors (10.8%) or assessor trainers (3.3%); a few served in other capacities (9.4%). On average, respondents had been involved in assessment center matters (in one way or another) in their organizations for 5.25 years ($SD = 3.89$ years) and had been involved in assessment center matters in general for 7.18 years ($SD = 4.98$ years).

Assessment Center Development

The Guidelines consider the following components “essential” to an assessment center:
1. The use of job analysis to identify necessary components of effective job performance, and those that can be measured in an assessment center.

2. Classification of candidates' behavior during the assessment center exercises into "meaningful and relevant" categories.

3. Design of the assessment center techniques such that they appropriately measure the dimensions outlined in the job analysis.

4. The use of multiple assessment techniques, pretested for reliability, objectivity, and ability to elicit relevant behavioral information.

5. Use of techniques suited for observing candidates' behavior associated with each dimension that are not biased toward assessees of a particular age, gender, or ethnicity for nonperformance reasons.

6. The use of several assessors for evaluating each assesse (both peer and self-assessment are acceptable), excluding the candidate's immediate supervisor.

7. Thorough training of assessors.

8. A system for assessors to record behavioral observations during the center exercises.


10. A combining of candidate information during the integration session in a valid and reliable way.

Respondents were not asked to complete a "checklist" to determine if their centers contained all of these elements. However, they were asked to describe their centers in terms of these characteristics.

**Job analysis.** Consistent with the Guidelines' recommendation, virtually all (93.3%) of our respondents' centers were based on a job analysis of the target job. Typically, a number of sources were tapped for data, including a job description (82.3%) and an interview with job incumbents (79.9%) and supervisors (79.9%) of the target position. A job analysis questionnaire completed by incumbent(s) of the target job (61.3%) and the corresponding supervisor (61.9%) was also often used. Critical incidents (44.8%) and observation of job incumbents (43.3%) were other typical sources of job information. Furthermore, in most of the respondents' centers (80.4%), a periodic review of the assessment center was conducted to ensure that it was still adequately related to the target job.

Our survey respondents have taken considerable effort to ensure the job relatedness of their centers by basing them on a number of sources of job data.

**Behavior elicitation and classification.** By and large, the product of the integration session in our respondents' assessment centers was a set
of ratings on all dimensions for each candidate (81.4%), providing indirect evidence that behavioral observations were combined into “meaningful and relevant” categories as the Guidelines recommend. These evaluations were based on diverse exercises, most commonly in-basket exercises (81.7%), leaderless group discussions, with (43.6%) or without (59.4%) an assigned role, interviews (57.1%), or simulations thereof (53.5%). Other popular exercises included analysis problems (49.3%), presentations (46.2%), fact-finding exercises (37.6%), and skills and abilities tests (31.0%). The majority of the respondents (64.8%) pretested these exercises before using them, usually by piloting them (49.0%) or by asking experts to review them (40.0%). Our respondents seem to have followed the Guidelines’ suggestion to use multiple exercises in their assessment centers that are pretested before use.

A modest percentage of respondents indicated that it used self-evaluations (31.3%) or peer evaluations (22.0%) in its centers. Gaugler et al.’s (1987) meta-analysis found that assessment centers using peer and self-evaluation produced higher criterion-related validity coefficients than those without these assessment techniques. Therefore, our respondents may be underutilizing a potentially valuable source of information about candidates in their assessment centers.

Behavioral records and combination. Also consistent with the Guidelines’ criteria for bona fide assessment centers, respondents in our survey indicated that assessors used multiple methods to record behavioral observations of their assesses. Almost all assessors took notes (93.2%) of candidate performance. Other typical methods included using check lists (41.3%) and behavior observation scales (24.2%) for this purpose. These observations were combined into evaluations by assessor consensus in a large percentage of our respondents’ assessment centers (84.1%). Less frequently, statistical aggregation (14.0%) and voting (2.4%) were used to perform this function. Because the consensus method has been found to produce assessment ratings that predict both training performance and on-the-job performance (e.g., Borman, 1982; Pyves & Bernardin, 1992), we can reasonably conclude that our respondents’ notion of an “assessment center” was consistent with that of the Guidelines’ authors.

Assessment Center Usage

The Guidelines provide detailed suggestions about virtually every aspect of assessment center usage. The following sections will describe recommendations for positioning the assessment center within the organization’s broader human resources system, training assessors, and for
providing information to the center candidates. Our respondents' practices will then be compared to these suggestions.

Organization Human Resources Policy

The Guidelines urge that a formal policy describing the relationship between the assessment center and the organization's human resource system be written. The policy should address the following areas:

1. The objective of the assessment center
2. The assessee selection, notification, and evaluation process
3. The assessors, including training, certification, evaluation, and usage limitations
4. Use of data
5. Qualifications of consultant(s) or assessment center developer(s)
6. Validation.

Although there was no survey question asking whether our respondents had created a formal policy addressing these issues, they did describe their centers in relationship to these areas.

Center objective. The most popular reasons for initially developing an assessment center were selection (50.0%), promotion (45.8%), and development planning (39.2%) purposes. Consistent with these expectations, the most popular decision-making processes that relied on data from assessment centers were promotion (60.8%), selection (54.5%), and development planning (51.2%). Assessment center data were also frequently used to make decisions regarding employee training (34.4%). Assessment center data are used for a number of important human resource decisions in our respondents' organizations.

Assessee selection. The most typical methods of selecting assesses were supervisor recommendation (53.3%), self-nomination (43.9%), or performance on other selection devices (37.7%). The final decision about who participates usually rested with a human resource director, administrator, or coordinator (24.1%) or a middle manager (19.2%). Those selecting assesses typically used supervisor recommendation (47.6%), performance appraisal information (38.3%), test data (26.7%), an interview (22.8%), and/or biographical data (22.3%) to make this determination. Given the highly fallible nature of human judgment (Longenecker, Sims, & Gioia, 1987; Nathan & Alexander, 1985), the heavy reliance on supervisor recommendation for inclusion in a center is somewhat disturbing. Perhaps other, less subjective sources of data should be considered in this decision.

Assessors. Assessors typically held positions that were about two organizational levels above those of the candidates ($M = 1.76$ levels; $SD$
=.71 levels), suggesting that, consistent with the Guidelines' recommendations, direct supervisors usually did not serve as assessors. The greatest proportion of assessors was line (48.5%) or staff (25.6%) management. In only a small percentage of the cases did psychologists serve as assessors (5.7%). These figures are less than encouraging: Gaugler et al. (1987) found higher levels of predictive validity in assessment centers as the usage of assessors who were psychologists increased. In our respondents' centers, assessors usually rotated through center administrations (83.4%), depending on their availability (21.7%), assignment (15.4%), or on a volunteer basis (10.3%). In almost half our respondents' assessment centers, assessors must be certified before acting as assessors (44.3%), which usually entailed the successful completion of a training program (77.8%). The "typical" assessor seems to be a line manager who serves in centers on an intermittent basis.

Assessment center data. Most of the time, candidates received oral feedback from assessment center personnel (70.5%) about their performance in the center for about an hour (M = 64.2 minutes; SD = 37.2 minutes). They also frequently received written reports on their performance (60.5%). In a few of our respondents' centers, no performance feedback was provided to assesses (8.1%). It is encouraging that our respondents adhered to the Guidelines' recommendations regarding providing feedback over 90% of the time. Following feedback delivery, assessment center data were stored for 4.5 years on average (SD = 3.3 years).

Developers and center validation. Consultants, either external (64.8%) or internal (40.0%) to the organization, were usually responsible for assessment center design. Most of our respondents' centers had been evaluated for reliability (51.9%) and/or validity (68.5%). Although the most popular type of validation was content validation (87.9%), predictive validation (36.4%) and construct validation (19.7%) were also used to examine the quality of assessment centers. In our respondents' centers, as the number of exercises in the center and the length of the assessor integration session increased, predictive validation (r = .22, p < .01; r = .31, p < .01, respectively), concurrent validation (r = .19, p < .01; r = .21, p < .01, respectively), and construct validation (r = .17, p < .05; r = .27, p < .01, respectively) were used more often.

For those reporting the type of reliability they used (n = 78), intrarassessor agreement was the most popular (63.9%). This measure was taken for within-exercise dimension ratings and rankings (72.7%), across-exercise dimension ratings and rankings (61.0%), exercise ratings and rankings (55.8%), and the overall assessment center evaluation
(66.2%). Given the consequence of the decisions that are based on assessment center data, it is surprising that a portion of the respondents did not indicate that it evaluated the quality of the data at all (21.4%).

**Assessor Training**

The *Guidelines* describe recommendations around four separate assessor training issues: training content, length of training, performance standards for assessors, and training currency. Variation in training programs based on trainer and instructional design (e.g., trainer expertise, instructional mode), assessor (e.g., familiarization with the organization and target job, frequency of participation), and assessment program (e.g., number of dimensions rated, exercise number and complexity) considerations is expected.

**Training content.** Assessor training programs should cover the following areas: knowledge of the organization and target job; understanding of assessment techniques, dimensions, and typical behavior; understanding of assessment dimensions and their relationship to job performance, knowledge of performance standards; skill in techniques for recording and classifying behavior and in use of the assessment center forms; understanding of evaluation, rating, and data integration processes; understanding of assessment policies and practices; understanding of feedback procedures; skill in oral and written feedback techniques (when applicable); objective and consistent performance in role-play or fact-finding exercises.

Our results suggest that most assessor training program designers adhered closely to these recommendations. Content of the training was generally quite comprehensive and uniform across organizations: It included introduction to the assessment center method and information regarding the assessment center procedure (98.0%), demonstration (87.7%) and explanation of the exercises to be used (99.0%), as well as explanation of the dimensions measured in the center (98.0%). It usually included information regarding typical rating errors, such as halo and leniency (87.7%). Also included were demonstrations of observation and evaluation (84.2%) and integration (70.4%) processes, as well as practice and feedback in classifying and evaluating behavior (82.8%). There was also usually practice and feedback in integrating evaluations (67.5%) and in writing evaluation reports (52.2%). Frequently, there was a simulation of exercises in which assessors assumed the roles of the candidates (68.0%).

**Training length.** Although length and quality of training are not synonymous, (Dugan, 1988; Gaugler et al., 1987; Task Force on Assessment
Center Guidelines, 1989), the Guidelines suggest that training of inexperienced assessors should last at least 2 days for every day of assessment center exercises. The average length of the assessment centers surveyed was 2.48 days (range: 2 hours–23 days). Comparison of this figure with the mean length of assessor training (3.84 days) and examination of the correlation between the length of assessor training program and the number of days of the assessment center ($r = -0.03, n.s.$) suggests that most organizations did not follow this recommendation.

Length of training program increased as the number of performance dimensions increased in our respondents’ centers ($r = .60, p < 0.01$). Longer refresher training programs were also described as the length of training program increased ($r = .40, p < 0.01$).

Assessor performance. The Guidelines indicate that assessor skills should be evaluated to ensure that the trainee is qualified to serve as an assessor upon completion of the training program. As a minimum, the assessor must demonstrate adequate skill in administering relevant exercises and in observing, classifying, and rating behavior. Our respondents indicate that as the length of the training program increased, the use of assessor evaluation ($r = .28, p < 0.01$) and assessor certification ($r = .36, p < 0.01$) increased. Assessors were evaluated at the end of the training in about half of our respondents’ assessor training programs (52.0%) by the assessment center administrator, trainer, or a consultant (25.5%) based on assessors’ performance during the training session (27.4%). It is discouraging that almost half of the survey respondents did not indicate that they evaluated trainees at the end of the training program, despite the Guidelines’ recommendation.

Training currency. When more than 6 months’ time has elapsed between the trainee’s completion of the training program and the first use of that individual as an assessor or when assessors serve in fewer than two assessment centers over 2 consecutive years, trainers are urged to provide assessors with refresher training (Task Force on Assessment Center Guidelines, 1989). Although the average frequency with which assessors annually served in assessment centers varied widely ($M = 7.09$ times; $SD = 24.63$ times), they usually served twice annually or less (68.9%). Refresher training was generally offered in our respondents’ assessment centers (61.2%). On average, it lasted about 1 day (range: 0.125–5 days), and most often consisted of either a shortened version of the original training session (24.4%), or focused on reviewing the exercises in the assessment center (26.0%).
Information for Participants

The Guidelines describe the information that participants should receive before the center begins (verbally or in writing). Although the content of this message will vary across organizations and assessment centers, it should include:

1. The objectives and purpose of the assessment center.
2. The method of participant selection.
3. The voluntary nature of the center and any consequences of not participating.
4. Training and composition of assessment staff.
5. The materials that are gathered and stored during the center.
6. Usage of center results and length of storage.
7. The type of feedback participants can expect to receive.
8. The policy regarding reassessment.
9. Criteria for who is granted access to candidate performance data.
10. Who is responsible for their data and where it is stored.

Most assesses did not hear much about our respondents' centers before the process began. They were most likely to hear some general information about the assessment center (34.5%), information about exercises (36.5%), or about dimensions (29.1%). In some of our respondents' centers, assesses received no information at all (11.3%). Familiarizing candidates with more aspects of the center and the way in which their performance data will be used seem like pressing areas for improvement. Fewer than 10% of our respondents indicated that they provided assesses with information about the objectives and purpose of the center (7.4%), how assessment center results will be used (7.9%), information about assessors and assessor training (6.4%), and how to prepare for the center (5.9%). The omission of this information is hardly trivial. For example, Joiner’s (1984) experiences with center candidates suggest that many complaints about assessment centers are filed due to a lack of knowledge about the program’s intentions at the outset.

Despite the considerable expense associated with assessing a single candidate ($M = $1,730; $SD = $5,192), most of our respondents indicated that candidates may participate in the center again if they were not successful the first time (67.3%). Most often, they were given one additional opportunity (72.1%), provided that they met minimum job qualifications or completed needed development work (28.8%), or if the candidate (12.9%) or organization management (13.7%) requested it. About a third of our respondents' assessment processes included an appeal process for assesses (31.8%), which was activated by contacting the assessment center director or human resources department (29.7%).
using the regular company grievance procedure (20.3%), or by writing a complaint (17.2%).

**Summary**

We have described assessment center practices in a sample of organizations in the United States in a comprehensive way. Assessment centers are generally designed by consultants and are based upon job analysis data. They are most often constructed for selection, promotion, development planning, and (to a lesser extent) training purposes. Designers and users of assessment centers tend to follow professional guidelines concerning the inclusion of components critical to assessment centers.

Supervisor recommendation plays a heavy role in determining who is invited to participate in assessment centers. Center candidates typically get little information about the center before it begins. Multiple exercises (many of which are pretested) are conducted during the center; they are usually observed by line or staff managers who serve as assessors. Frequently through a consensus process, assessors create a set of ratings on a set of performance dimensions for each candidate. Rarely do self- or peer evaluations influence the candidate ratings. Assessee receive on average about an hour’s worth of oral feedback and a written report following the center activities. In many cases, candidates are offered the opportunity to participate in the center more than once if they are not successful the first time.

Assessors are typically sent to comprehensive training programs before serving in a center. However, the training programs may not be as long as they ought to be. Furthermore, the adequacy of assessors’ skills is often not measured at the end of the training program. Similarly, assessment center data are not evaluated in terms of reliability and validity a considerable portion of the time. The assessment center method is too costly and labor intensive to be used without evaluating the quality of the data it produces.

Comparison with the most recent published survey results. In order to update our knowledge of assessment center practices (since Fitzgerald and Quaintance’s 1982 survey), we first isolated respondents from government organizations (this was the population considered in their work). We then examined responses to questions that were similar to the ones that they asked. The following paragraphs describe the similarities and differences in center practices used by their respondents and by ours.

Promotion continues to be the most popular purpose for which assessment centers are administered. In-basket and leaderless group exercises are still among the most frequently used assessment tools. In just
over half of our respondents' centers, interviews, analysis problems, presentations, and interview simulations were also used. Although the average percentage of women and minorities participating in our respondents' centers was just below 20%, there was much variability in the composition of the candidate group: It included anywhere from 0% to 87.5% women and 0% to 95.0% minorities. On average, there seems to be an increase in the participation of women and minorities since the earlier survey.

It is encouraging that about two-thirds of our respondents indicated that they validated their center data. Like respondents to the earlier survey, the most popular method of validation was the content validation method (95.2% of those who validate). However, the use of predictive validation (23.8% of those who validate) and construct validation (16.7% of those who validate) has increased considerably since Fitzgerald and Quaintance's survey was published (1982).

Finally, although the average length of assessor training programs was 3.2 days and the average length of assessment centers was 2.8 days, some respondents indicated that they trained assessors for only .5 day and ran centers for only .3 day. Therefore, we reiterate Fitzgerald and Quaintance's (1982) concern regarding the length of some centers and some training programs.

_**Moderators of Assessment Center Characteristics and Outcomes**_

One objective of this study was to describe differences in assessment center practices in various industries. A second goal was to examine the usage of certain center characteristics that have been found to influence predictive validity. We wanted to explore our respondents' centers with respect to these characteristics in the hope of gaining insight into the way in which they influence the quality of the center data. We have already described characteristics of assessment centers in general. Now we will present differences in assessment centers used by those in various industries. We will then examine differences among centers conducted for various purposes.

_Differences in Assessment Centers According to Industry_

We present a description of certain features of assessment centers that are used in the following types of organizations: educational, consulting, manufacturing, banking, sales, and service. We removed respondents who indicated that they came from an "other" or "unspecified" industry because of the interpretive difficulties that such a response posed. One hundred ninety six cases remained after this adjustment. We have
retained all six industry groups to maximize the richness of our data, despite the relatively small sample size in some cases.

Chi-square (with categorical dependent variables) or ANOVA (with continuous variables) analyses were conducted with industrial groups as the independent variable. The paragraphs below describe the dependent variables that differ significantly among industrial groups and that reflect an interesting and meaningful level of variability among users of assessment centers in the industry groups. (Complete analyses are available from the first author.)

Assessment centers used in various industries seem to be characterized by particular features. For example, manufacturing organizations use exercises that are not often used by those in other industries. They also train assessors more thoroughly on integration processes than do other organizations. Sales organizations tap a wider variety of sources for job analysis data (particularly incumbent sources) and for candidate selection data than do other organizations. They also offer candidates feedback from limited sources once the center is over. Banking organizations do not use incumbent sources of job data very often in constructing their centers. They also train their assessors for shorter periods of time than do other organizations. Banking organizations evaluate center data (with respect to reliability and validity) less frequently than do other organizations.

Government organizations seem to stay away from less commonly used exercises. Service organizations tend to use assessment centers heavily for development planning purposes. Educational organizations train assessors thoroughly and evaluate their center data more extensively than do other organizations. However, they also use fewer data sources to select assessee. Consulting organizations tend to do a better job than others at informing candidates about center purposes and objectives before they begin. Perhaps these organizations are more familiar with professional guidelines concerning assessment center administration, and take more care in informing participants as a result.

Moderators of Assessment Center Validity

The bulk of research suggests that assessment centers generally have acceptable predictive validity (corrected $r = .37$; Gaugler et al., 1987). However, there are certain characteristics of assessment centers and the way they are used that influence their validity coefficients. For example, in their meta-analysis of assessment center validity, Gaugler et al. (1987) found that predictive validity increased along with the number of female assessee, psychologist assessors, number of evaluation devices used, and the use of peer evaluation in assessment centers. Predictive validity also
varied with the purpose of the assessment center as well as with the criteria used for the validation process. Specifically, the validity of assessment center data decreased when promotion was used as a criterion and increased when ratings of potential were used as a criterion. With respect to center purpose, centers used for early identification, selection, and research demonstrated higher validities than those used for promotion purposes.

How are these characteristics represented in our respondents' centers? They appear with different frequency in centers used in various industries. For example, respondents from different industries use staff managers as assessors to different degrees: Governmental organizations used the highest proportion of staff managers among their teams of assessors (34.5%); education (12.5% of assessor team) and consulting (9% of assessor team) organizations used this type of individual to serve as assessor the least often. Differences in the use of external psychologists as assessors also existed for respondents from different types of industries. Specifically, respondents from consulting (19.5% of assessors) and banking (16.7% of assessors) organizations used external psychologists significantly more frequently than did respondents from other industries.

Correlational analyses of the survey data revealed differences in the way in which assessors of various types conducted their integration sessions. This may help to explain why centers using psychologists as assessors produce higher validity coefficients than centers using managers as assessors. Specifically, as the percentage of line manager assessors increased, more time was spent reporting the data versus integrating it (average $r = .15$, $p < .05$). Conversely, in assessment centers with more psychologist assessors, more time during the integration session was used for integrating candidate data versus reporting it (average $r = .19$, $p < .05$). Perhaps improvements in center validity can be realized by spending more time integrating candidate data. This strategy might be conveyed during assessor training. Centers used in organizations from different industries include training on integration of assessment results with differing frequency. Most include a demonstration of the integration process. Government organizations include demonstrations of assessor integration processes to a lesser extent (52.1% of the time) than do other organizations. Manufacturing organizations and educational organizations tend to offer assessors more practice in integration than do other organizations (86.8% of the time; 91.7% of the time, respectively). Banking organizations rarely offer practice and feedback in this task (33.3% of the time). Banking organizations include a higher proportion of external psychologists in their assessor teams than do other organizations; perhaps they assume that the assessors are already skilled in the integration task and do not require additional practice.
Finally, the use of peer evaluation was more frequently reported by manufacturing organizations (43.6% of the time) than by respondents from other industries. Sales (10% of the time), government (11.3% of the time), and education (11.5% of the time) organizations used this technique least often. However, in all our respondents' centers, peer evaluations were given minimal weight at best when making decisions about the candidate.

Differences in Assessment Centers According to Purpose

Because center purpose had been shown to affect validity, we examined the moderating effects of this variable on other center characteristics. By performing this investigation, we hoped to shed some light on the drivers of the differences in validity among centers designed for various purposes.

This analysis began by removing respondents who indicated that their centers had been designed for multiple purposes. We eliminated those endorsing the early identification, training, and management succession purposes due to their small sample sizes (i.e., fewer than two respondents). We also removed individuals indicating that their centers had been designed for "other" purposes because of the interpretive difficulties that such a response posed. One hundred twelve respondents remained.

We then conducted ANOVAs when continuous variables were the dependent variables and chi-square analyses when the dependent variables were categorical in nature. We describe the characteristics that differ meaningfully across assessment centers intended for the purposes of selection, development planning, and promotion in the paragraphs below. (Complete analyses are available from the first author.)

There are features that appear, to a greater or lesser extent, in assessment centers constructed for various purposes. In selection centers, assessors serve many times annually, thus providing opportunities for them to keep their skills current. Assessors are almost always asked to create overall performance ratings in these centers. Center data are validated more frequently in centers of this type than in others. This may reflect efforts to prepare for possible legal challenges to selection decisions based on the center data. Such challenges may be expected more often in centers of this type than in other centers.

In development planning centers, there are fewer candidate selection mechanisms used, with heavy reliance on supervisor data. There are many female assessees, and varied exercises to complete. Assessors conduct lengthy discussion sessions in these centers, both with other assessors and with candidates in feedback sessions. Perhaps longer feedback
sessions are used to formulate development strategies with the assessees in addition to reporting the results. Development planning centers are infrequently validated.

In centers conducted for promotion purposes, candidates are selected based on a wider variety of data sources than in other centers. Assessors in these centers are asked to observe many more assessees during each exercise than are assessors in other centers. However, assessees in promotion centers are asked to complete fewer types of exercises than are assessees in other centers. This may help make the large candidate load more manageable for assessors.

**Discussion**

We have discovered that general assessment center practices differ from a number of recommendations in the *Guidelines*. Designers and users of assessment centers tend to follow professional guidelines concerning the content of assessor training. They deviate from recommendations regarding length of assessor training and those concerning informing assessors about various aspects of the assessment center process. There are also gaps between recommendation and practice in evaluating various aspects of the center and its use. For example, more frequent pretesting of exercises, evaluation of center validity, and evaluation of assessor performance are needed according to professional guidelines for assessment center usage.

We need to investigate the gaps that this study has identified between assessment center practice and recommendations. Because the *Guidelines* were based upon expert prescriptions for assessment center usage, the usefulness of the process should increase as they are followed more closely. However, it is possible that some of the *Guidelines* prescriptions do not improve either the assessment center process or its outcomes. Thus, it may make sense to ignore those recommendations. Alternatively, perhaps users of assessment centers assume that others have taken steps to ensure that their centers adhere to the *Guidelines* recommendations. For example, an organization may ask a consulting firm to design an assessment center, and assume that the validation and reliability measures were taken by that organization at the outset and that no further measurement is needed. Given the legal ramifications of ignoring the professional guidelines for the assessment center method, it behooves us to evaluate the *Guidelines* impact on center quality and to adjust or remove prescriptions that add little to the process.

The individuals who responded to our survey cared enough about their assessment centers to make a reply; it is likely that others using assessment center technology follow prescriptions for assessment center
usage less closely. This is problematic due to potential legal difficulties associated with assessment center usage. It also suggests a compromised quality of the process.

Research has identified center characteristics that seem to improve its predictive validity (e.g., a higher number of female assesses and psychologist assessors, use of peer evaluation). However, we do not recommend that assessment center users simply begin using these practices by recruiting more female assesses, employing more psychologist assessors, and using peer evaluation more frequently for the sake of increasing the predictive validity of their center data. Instead, we invite practitioners and researchers to investigate the causal dynamics behind relationships of this type so that the truly essential elements can be used to improve the assessment center method. For example, Gaugler et al. (1987) suggest that the inclusion of female assesses may change the center dynamics such that more useful data are gathered from the candidates. Shore, McFarlane Shore, and Thornton (1992) found that peer evaluations accounted for unique variance in job advancement beyond assessor ratings and cognitive ability tests. Reasons behind the moderators' effects on assessment center validity should be investigated more thoroughly.

Another task is to improve assessment center technology and methodology. Common complaints about assessment centers include the time and expense involved. Unless assessment centers meet organization requirements for manageability, there is little hope that they will be utilized, regardless of the benefits that they offer. Therefore, they must be made more "user-friendly."

Study Limitations

There are a number study limitations that should be addressed. The first is the age of the data reported. As mentioned, the data were collected during 1990. Since that time, assessment center technology has advanced. For example, the use of multimedia simulations has become more common (Smiderle, Perry, & Cronshaw, 1994). Unfortunately, no formal comparison of older and newer center characteristics has been published. We do not know how the newer technology affects the assessment center process or the quality of the data. However, we do know that our survey data update our knowledge of usage of assessment center techniques. As discussed, the most recent work that is similar to ours was published in 1982.

A second problem is that we are unable to determine the representativeness of our sample in comparison to the entire group of individuals that was targeted for survey distribution. Unfortunately, there is
no published research that describes the characteristics of organizations that use assessment centers. In our study, the organizations represented most heavily are those in the government. Almost 35% of our respondents came from this source. This should be kept in mind as survey results are interpreted. Our relatively high response rate of almost 50% offers some assurance that our data are representative of organizations that use assessment centers.

REFERENCES


