WHY DO ASSESSMENT CENTERS WORK? THE PUZZLE OF ASSESSMENT CENTER VALIDITY

RICHARD KLIMOSKI
The Ohio State University

MARY BRICKNER
University of Akron

The finding that assessment centers are successful in making valid predictions of managerial success is well established in the literature. Unfortunately, it is not clearly understood why assessment centers work. For example, construct validity of assessment center dimensions is rarely obtained. In this paper we affirm the evidence for predictive validity of assessment centers and conclude that assessment centers can work for a variety of purposes and in numerous contexts. But we also assert that we do not know why they work. The bulk of the paper goes on to raise possible explanations for the predictive validity observed in assessment centers and raises implications for practice and guidance of future research.

Assessment centers appear to be the modern enigma in human resource practices. The use of assessment center procedures for assessments of managerial potential is prevalent in corporations, large and small. Although assessment centers are generally considered valid predictors of managerial success, the nature of those predictions and the underlying dynamics of assessment center practices remain a puzzle. A number of questions surrounding assessment centers have been raised. In this paper we will try to piece the assessment center picture together by considering questions that surround assessment center practices.

Do Assessment Centers Work?

for different types of performance predictors. Their reported mean predictive validity for assessment centers was .407. Meta-analyses by Hunter & Hunter (1984) and Gaugler and her colleagues reported validities in the range of 37–43.

Given the predictive validities consistently reported in reviews, we would have to conclude that indeed assessment centers do work. Assessment centers are useful tools for predicting the future success of potential managers.

For Whom Do Assessment Centers Work?

Available studies demonstrate the usefulness of assessment centers for predicting managerial success regardless of educational level (Huck, 1973), prior assessment center experience (Struth, Frank, & Amato, 1980), race (Huck & Bray, 1976; Moses, 1973), or gender (Moses, 1973; Moses & Boehm, 1975). These studies supported the usefulness of assessment centers in predicting managerial success fairly regardless of membership in subgroups.

Where Do Assessment Centers Work?

Assessment centers have been used in a wide variety of organizational settings. This selection tool has been effectively utilized in manufacturing companies (Turnage & Muchinsky, 1982), government (Struth et al., 1980), military services (Borman, 1982, Tziner & Duran, 1982), utility companies (Schmitt, 1977), oil companies (Norton, 1977), educational institutions (Schmitt, Nee, Merritt, & Fitzgerald, 1984) and by the FBI (Neidig, Martin, & Yates, 1979).

For What Purposes Do Assessment Centers Work?

Assessment centers have proven to be useful for a variety of purposes beyond promotion and selection (Cascio & Silbey, 1979). They are useful in training and development (Lorenzo, 1984), for career planning (Gaugler, Rosenthal, Thornton, & Bentson, 1985), and in improving important managerial skills in assessors (Lorenzo, 1984). However in the few comparative studies they have been most predictive of advancement criteria (Klimoski & Strickland, 1981; Turnage & Muchinsky, 1984).

Why Do Assessment Centers Work?

The studies reviewed above establish that the assessment center is a useful tool for predicting managerial success, across organizations and types of employees and for a number of purposes. Given the extensive literature, however, we believe it appropriate (even long overdue) to ask “Why?”.

As straightforward as this question must seem, it appears to us that no firm answer is yet available. Quite the contrary, given the nature of assessment center research and the evidence generated in the last ten years, there appear to be several plausible explanations.

The traditional explanation. Assessment centers have been designed to predict managerial success by providing raters with an opportunity to infer personal qualities and traits that have been determined, through careful job analysis, to be relevant to success (Byham, 1980). That is, assessment centers are standardized devices to allow assessments of traits, which are then used for predicting future success on the job. The traditional argument is that assessment centers work because they do a good job of measuring and integrating information regarding an individual’s traits or qualities (Byham, 1980). But do they?

The evidence for the construct validity of the dimensions used in assessment centers is not encouraging. The bulk of the reported literature shows little support for the view that assessment center procedures do in fact produce scores that serve as valid representations of separate constructs or that those constructs are used in evaluation decisions in the manner proposed by assessment center designers.

Sackett and Hakel (1979) examined how individuals used assessment center information in forming overall ratings. They found that assessors used only a small number of dimensions although they had been instructed to use all of them in making their judgments. Three dimensions (leadership, organizing and planning, and decision making) accurately predicted overall ratings. Considered pessimistically, their results imply the existence of only general and diffuse measurements of behaviors. Alternatively, it may be that the dimensionality of effectiveness is not that complex. In either case, this raises real questions about both the need and the potential for centers to discriminate among a large number of dimensions.

Turnage and Muchinsky (1982) reported that assessment center trait ratings gave little help beyond what could be obtained from a global rating. The average correlation between dimension ratings and the overall rating was .91. They reported a lack of discriminant validity and high levels of convergent validity across traits. This was considered an indication that assessors were making global evaluations rather than differentiating among traits.

---

1. The above findings should not be used to conclude that assessment centers are performing better than alternatives in predicting these different criteria. The limited number of comparative studies which contrast alternative predictions suggest that they are not (cf. Borman, 1982; Hinrichs, 1978; Hunter & Hunter, 1984; Klimoski & Strickland, 1981).
Sackett and Dreher (1982) studied the interrelationships among dimensions between and within assessment center exercises in three different organizations. In all three organizations they found within-exercise ratings correlated more highly with each other than did dimensional ratings across exercises. In two of the organizations there was no convergence among the various measures of a dimension (the average correlation was zero). Method variance predominated over the shared variance of measures of a single trait. In the third organization, all ratings were highly correlated with all other ratings. This indicates a lack of discriminant validity for the dimensions considered. They concluded that there was little evidence that assessment center ratings accurately reflect the complex traits that they purport to measure (p. 401).

Russell (1985, 1986) sought to further our understanding of the decision-making processes of assessors in assessment centers. His findings were consistent with previous research (Sackett & Hakel, 1979): assessors did not use specific dimensions in making their judgments. In his study, assessors’ perceptions of center participants were strongly affected by an underlying factor (either interpersonal skills or problem-solving skills). In his later study, halo across the six dimensions assessed in two exercises was great. Russell suggests that “the best guesses of assessment center architects and job analysts should not be expected necessarily to exhibit rigorous evidence of construct validity” (Russell, 1985, p. 743). He too emphasizes that the bases for the predictive validity of assessment centers are not understood.

The available research consistently demonstrates a lack of evidence for the construct validity of assessment center dimension ratings. Moreover, it convinces us that assessment centers are not working as designed. If they have predictive validity, it is not because they are effectively measuring and using traits (Zedeck & Cascio, 1984). We must look elsewhere for the answer.

Some alternative explanations. A variety of alternative explanations for why assessment centers appear to work can be gleaned from the literature. The following are presented as possibilities. In many cases they appear as suggestions made by various writers. Few have been carefully evaluated.

Actual criteria contamination explanation. The apparent predictive validity of assessment centers could arise out of the unintended but real possibility that promotions in organizations (or other criterion decisions) are partially based on assessment center judgments. Thus, individuals who get favorable ratings are considered for promotion over those who do not. A future analysis of the relationship between center assessments and promotions would then appear to show a correspondence between assessment center ratings and success criteria.

In the early literature (e.g., Kraut & Scott, 1972), this dynamic was popularly referred to as the “crown prince (princess)” system of promotions. In true predictive studies (e.g., Bray & Grant, 1966), of course, this type of criterion contamination is not possible (unless the assessment center ratings are allowed to be “leaked out”). However, in many studies (e.g., Klimoski & Strickland, 1981) data are gathered from operational centers. That is, assessment center evaluations are obtained and used for purposes of administrative action (e.g., selecting individuals for promotions). Under these circumstances criterion contamination is quite possible.

As of 1974, Howard listed ten studies reporting validity data in which center ratings had been used for promotions. Not surprisingly, all of these showed significant relationships. She argued, however, that while such contamination could indeed artificially inflate the apparent validity of center predictions, it was not a major problem because center ratings would be used less frequently as a basis for advancement in later years. Kraut and Scott (1972) were similarly reassuring. They maintained that a favorable rating was not sufficient to be promoted, with the “possible exception of the top 5% who receive the very highest rating.” (p. 128) If some companies use center ratings for promotions, however, it is only for the initial move. Beyond this, other dynamics operate to determine who gets ahead (Stumpf & London, 1981). Silzer (1985) reported that 73% of those providing on-the-job ratings had never seen assessment reports; 9% had seen them three to nine years before providing criterion information. Finally, Gaugler et al. (1985) found little support for the actual contamination hypotheses. In their meta-analysis, study design did not moderate assessment center validity. Specifically, reports from operational centers did not show higher validities than those where center ratings (as potential predictors) were thought not to be used. Nevertheless, in the face of so little research to test this notion specifically, it cannot be ruled out as a potential explanation for apparent center validity.

Subtle criterion contamination. Klimoski and Strickland (1977) proposed the notion that assessment center validity might be affected by a form of criterion contamination that operates in an indirect or subtle fashion. Their reasoning was based on the fact that a great deal of evidence for the validity of assessment centers was based on predictions of promotions or promotion-linked criteria (e.g., salary growth). They proposed that, instead of looking for behavioral evidence of specific traits or personal qualities, assessment center staff in fact observe and evaluate candidates on the basis of their knowledge of those factors needed to get ahead in the company. Instead of systematically evaluating each person on the dimensions created for the center, they attempt to perform a policy-capturing function and to mimic what future decision makers might do in making a promotion decision. This may or may not be based on performance. Thus,
judgments made by center staff would tend to correlate with judgments made by managers in the field.

Once again there is little definitive research on this potential explanation; most evidence is circumstantial. For example, if this dynamic were operating, assessment center validities would be higher for promotion criteria than for other success indicators (e.g., performance in grade). As mentioned, this does appear to be the case. Similarly, the explanation implies that assessment center staff characteristics would affect the magnitude of the validities obtained. That is, staff who come from or who have intimate knowledge of the corporate setting into which center participants will go after being assessed should be better able to predict (or anticipate) the promotion criteria and processes. They should produce judgments with higher validities than outsiders or consultants. Silzer (1985) did find higher validities for staff who had greater familiarity with norms in one versus another company. In fact, he agrees with Holt (1970) that such knowledge is an indispensable feature of a good clinical prediction paradigm. In contrast, Gaugler et al. (1985) reported in their meta-analysis that validities were higher when assessors were psychologists rather than managers. Presumably the former would have less familiarity with organizational decision making about promotions than the latter. Still other studies have reported equivocal results (e.g., Borman, Eaton, Bryan, & Rosse, 1983).

One other line of evidence relevant to the argument that subtle criterion contamination may be operating has been raised by Dunnette & Borman (1979). They observed that assessments of overall performance or potential in assessment centers tend to correlate more highly with organizational success criteria than do ratings of specific dimensions. Moreover, the average validity coefficients for dimension ratings are much lower than those obtained for overall ratings (e.g., Turnage & Muchinsky, 1984). They reason that overall assessment ratings are thus likely to be influenced by factors that are presumably linked to success in the company but not reflected in the dimensions (e.g., proper background, appearance, etc.). In our opinion, the subtle criterion contamination hypothesis of Klamoski and Strickland cannot be ruled out as a significant contributor to assessment center validity.

Self-fulfilling prophecy explanation. Another possible explanation for the apparent validity of assessment centers relates to a self-fulfilling prophecy dynamic operating for assesses. It could be argued that being selected to participate in an assessment center may reinforce the feelings of self-efficacy for competent managerial candidates. Bandura (1982) notes that judgments of self-efficacy affect how much effort a person will direct toward a goal and how long he or she will persist in striving to attain a goal. He suggests that if a person has a strong sense of self-efficacy, that person will exert more effort and be more persistent in meeting a challenge. Moreover, competent managerial candidates, after experiencing success on assessment center tasks, may also have increased feelings of self-efficacy. This state results in the candidate directing more effort toward the development of skills and abilities that are important to managerial success. In this way, the selection of competent people for participation in the assessment center becomes a self-fulfilling prophecy. Even if individuals are put through centers as a matter of course (as a result of seniority or a court ruling), receiving favorable feedback on performance in the center would reinforce an "effective-manager" self-image. This, in turn, would result in greater effort and persistence in developing the managerial skills and abilities important to managerial success. Such increased effort in the development of managerial skills should result (given the assessed capacity), in greater managerial capability. Finally, higher levels of managerial skill would increase the probability of selection or promotion to a managerial position. In a sense, the assessees can make the staff prediction come true.

Both Russell (1986) and Gaugler, et al. (1985) refer to these notions and the potential role of self-efficacy. Moreover, we know that expectations for high performance on the part of others can be communicated and do have an impact (Rosenthal & Jacobson, 1968). For example, in a field experiment in a training context, Eden found that staff members who believed in the high potential of their trainees (regardless of their actual capabilities) had classes with disproportionate numbers of high performers. More to the point, this high performance continued even when the original trainers had been transferred for administrative reasons (Eden & Ravid 1982; Eden & Shani, 1982). Thus, instructor expectancies appeared to have been perceived, reinforced, and internalized as part of the trainees' self-images, with positive outcomes as a result (see also Crawford, Thomas, & Fink, 1980). While little data to test this notion in an assessment center context are available, we do know that center participation alone can change or reinforce self-perceptions (Schmitt, et al., 1986).

Performance consistency explanation. Judgments in assessment centers are supposed to be based on trait inferences made from observations of behavior elicited by the center exercises and tasks.

It is possible, however, that a different process might be operating. That is, staff may be evaluating the past and present performance of individuals and basing overall assessments on these, thus bypassing the judgment of traits entirely. High performers in the centers are thus predicted to be high performers in future managerial roles.

This explanation is based on certain features of the typical assessment center. One is that a great deal of achievement-relevant background information in the form of autobiographical information is available to staff before they make assessment evaluations. This information is gathered at a number of points in the center process, most notably in the extensive, in-depth interviews that are usually conducted. The vast literature on the
value of biodata for predicting job success (Campbell, Dunnette, Lawler & Weick, 1970; Childs & Klimoski, 1986; Owens, 1976) makes it clear that the validity of staff predictions could be enhanced by relying on such information. Past history of successes would thus be expected to relate to future success.

As straightforward as this explanation is, it is also true that there seems to be little evidence available to evaluate it. Studies have attempted to examine components of the assessment center for their contributions to the center's prediction of success (e.g., Hinrichs & Haanpera, 1976; Wollowick & McNamara, 1969), but the biodata component has rarely been explicitly considered. Borman (1982) found little evidence of the predictive validity of a structured interview in an assessment center created for evaluating Army recruiters. But it is not clear how much biodata was actually obtained in the interviews conducted for the study.

On the other hand, Hinrichs (1969) noted that data from existing personnel records were highly predictive of overall ratings obtained from a two-day assessment center program. He suggested that a careful evaluation of personnel records and employment history could provide much of the same information (or at least the predictive power) as a lengthy and expensive assessment program (p. 431). Turnage & Muchinsky (1984) did use assessment center data in which there was a biodata component, but the results (with the exception of age) were not that supportive of the past history explanation of center ratings.

Alternatively, the predictive validity of assessment centers may rest on a second feature, the fact that center exercises serve as job samples. To the extent that the exercises (as job samples) reflect behavior and performance that is relevant and representative of future managerial job requirements, evaluations should be related to and predict managerial success. Such evaluations, however, would serve as indications of current levels of performance (a sample) rather than assessments of managerial potential (a sign; Wernimont & Campbell, 1968). To phrase it differently, present performance (in center exercises) predicts future performance (as managers).

The designs of most assessment centers are based on the results of job analyses (Byham, 1980). Most specialists would agree that this is a necessary step in the development of any center. Nevertheless, there is some controversy as to just how job analysis information is to be used. Traditionally, it has provided the bases for the identification of managerial job requirements—the traits or qualities to be assessed. In fact, it serves to produce what is usually referred to as the construct validity for assessment centers (Sackett & Dreher, 1984). But job analysis can also be the source of information for creating the actual simulations used. That is, it can be the basis for establishing a center's content validity. For example, Neidig and Neidig (1984) contend that an accurate sampling of relevant work situations in a center is critical for establishing the job relatedness of the assessment center. In this way job analysis can ensure that good performance on the exercises will relate to good on-the-job performance because the candidate is presumed to have the requisite knowledge, abilities, and traits if he/she does well. But it should be stressed that the measurement of performance levels on samples (as predictors) need not even involve the notion of underlying constructs (see, e.g., Campion, 1972).

We will not attempt to restate, much less resolve, the arguments about whether assessment centers should have construct validity or content validity (or both). The interested reader is referred to Brush and Schoenfeldt (1977), Byham (1980), Dreher and Sackett (1981), Neidig and Neidig (1984), Norton (1977, 1981), and Sackett and Dreher (1982, 1984). The point is that most centers do involve the use of apparently content (face?) valid exercises or simulations (i.e., job samples). Assessment centers may work (be valid) because levels of performance on these exercises, not inference with regard to particular traits, form the bases for predicting managerial job success.

A managerial intelligence explanation. A final hypotheses with regard to the reasons for assessment center validity bears some relationship to the subtle criterion contamination notion. It is sufficiently distinct, however, to merit special mention. To put it simply, assessment centers may predict managerial success because the ratings obtained reflect the level of intellectual functioning of candidates.

There seems to be no doubt that intelligence is important for managerial effectiveness. Most analyses of managerial job requirements refer to the importance of verbal skills (manifested in oral and written communications), analytic or reasoning skills, the regular use of short- and long-term reasoning, including well-developed plans or routines (for example, scripts) for the combining of information. Effective managerial functioning is also thought to involve what has been called the application of "tacit knowledge" (Wagner & Sternberg, 1985). Research suggests that tacit knowledge appears to be acquired and developed by intellectually more capable individuals. In short, intelligence can be defined by that old phrase, "what one needs when you don't know what to do". Given the nature of managerial work, intelligence must be viewed as a major determinant of managerial success (Lord, DeVader, & Alliger, 1986; Yukl, 1981).

Actually, the empirical relationship between intelligence and managerial performance has long been established (Ghiselli, 1966, 1971, 1973; McCormick & Tiffin, 1974; Miner, 1957). Ghiselli (1966) reported intelligence tests as the best predictors of future performance for foremen, managers, administrators, and executives. Moderate correlations (.25–.30) between intelligence and performance are frequently obtained. In his (1971) study of managerial talent he was also able to report moderate correlations
between intelligence and performance for managers and personnel officers. Lower correlations (.01-.07), however, were found in studies of line managers in packing plants. Similarly, in 1973, Ghiselli determined that intelligence tests were useful for predicting trainability, job proficiency, and performance among foremen, administrators, and executives (correlation ranging from .28-.36). These data provide evidence for the consistent relationship observed between intelligence and effective managerial performance. Minar also (1957) reviewed a number of studies of intelligence and success in military contexts.

The importance of intelligence has also been recognized explicitly by the assessment center literature. To begin with, it is frequently part of the measurement plan of assessment centers. Thornton and Byham (1982) refer to the number of centers that incorporated intelligence measures. Bray and Grant (1966) used three intelligence tests in their famous Management Progress Study.

Scores from these tests all correlated significantly with staff predictions of success for managers. Moses (1973) found that SCAT scores were significantly correlated with intelligence in a longitudinal study of assessment center evaluations and managerial performance. Huck (1973) reported moderate correlations (.40) between SCAT test scores and overall assessment ratings. And Carleton (1970) found moderate correlations between intelligence test scores and a composite criterion of behavioral ratings in assessment centers. This was also true of a recent study by Wolfson (1985) at IBM. Schmitt (1977) reported high correlations between intelligence test scores (SCAT) and overall dimension ratings.

Finally, the observed predictive validity of assessment centers may be attributable, in part, to assessment centers as measures of intelligence. For example, Tziner and Dolan (1982) compared assessment center results with some traditional forms of evaluation (i.e., tests of verbal intelligence, supervisory evaluation, evaluations for selection interviews). They found that verbal intelligence scores were highly predictive of future performance. In fact, these scores were better than individual exercises and overall assessment ratings for predicting future performance. Intelligence scores accounted for the largest portion of variance. Moreover, Klimoski and Stickland (1981) found that paper and pencil measures of intelligence predicted future managerial performance ratings better than did assessment centers.

All this is not to argue that measures of intelligence are superior to assessment centers as predictors or that assessment ratings are to be characterized as perfect surrogates for intelligence tests. In fact, the evidence is too equivocal to do this. The interested reader can contrast the findings of Schmitt, Gooding, et al. (1984), Turnage and Muchinsky (1984), Russell (1985), and Gaugler et al. (1985).

What does seem clear, however, is that both center behavior and assessor judgments are influenced in part by the level of general intellectual functioning of assesseees. Further, the on-the-job performance of these same individuals is likely to be similarly affected. If assessment centers predict job success, their apparent validity may be due, in part, to this underlying nomological network. At this point, however, we just don’t know the extent of this happening.

In this section we have discussed several possible explanations for the predictive validity that is so characteristic of assessment center predictions. These explanations (or hypotheses) include criterion contamination, policy capturing, self-fulfilling prophecies for assesseees, construct and context capturing capacity, and assessment centers as measures of intelligence. These explanations are not mutually exclusive and are probably not equally complete or plausible. However, they do provide implications for both practice and future research.

Implications For Practice

The various explanations for the mechanisms by which assessment centers appear to obtain their validity would seem to make a difference for practices in this area. That is to say, the beliefs of and assumptions made by managers would guide choices among different assessment center formats and the likelihood of their use.

Managers frequently have a real need to assess or inventory the personal traits or qualities of individuals. For example, they may wish to provide accurate and specific developmental feedback to employees. Thus, they might well turn to assessment centers as the technique of choice. The available research evidence suggests it would be a mistake to do this. They might be better advised to consider alternative devices (e.g., standardized paper-and-pencil instruments, clinical inferences by trained professionals).

If, however, managers believe that the assessment center has the potential to be construct valid, certain steps are warranted. They might start by limiting the number of dimensions or qualities to be assessed. The latter could be selected to reflect higher order constructs or broader attributes than is usually the case (e.g., interpersonal skill vs. behavior flexibility), thus acknowledging both the ecological relationship among human traits and the limits of human (assessor) capacities to discriminate. For example, Barr and Hitt (1986) reported that experienced managers used practically fewer factors yet produced decisions with more explanatory power than did naive subjects.

Furthermore, staff might be given longer periods of training than seem to be the case with centers set up more recently. Similarly, individuals might be given extended "tours of duty" as staff assessors. In this regard,
Implications for Research

While the above advice might be given to practitioners, it must frequently be based on inferences from the available research literature. The fact is, most evidence for the alternative dynamics of center validity is circumstantial. Thus, almost any of the suggestions provided should be considered tentative and the starting point for research.

In terms of priorities, however, given the real need of organizations to assess potential (apart from competencies), it would seem most important to establish if, or under what conditions, assessment centers can be made to produce valid measures of constructs. Given what we now know about human judgment and social cognition, there is some reason to be optimistic about finding a solution. Recent theories in these areas should be applied to the assessment center process (see Zedeck, 1986, for an excellent treatment of this).

Specifically, numerous and potentially relevant variables could be experimentally manipulated to determine their impact on discriminant and convergent validities of staff ratings. For example, Silverman, Dalessio, Woods, and Johnson (1986) predicted on the basis of current models of cognitive processing that there would be differences in the dimensionality and construct validity of assessments made by staff who were instructed to form dimension ratings right after each center exercise (a common practice) and by those who would have to postpone making such judgments until all the exercises had been completed. This was indeed found to be true. Unfortunately, the authors did not go on to see to what extent criterion-related validities were similarly affected. Turnage and Muchinsky (1982) also stressed the potential impact of limiting the number of dimensions to be assessed on construct validity.

Similarly, Zajonc's (1960) concept of cognitive tuning implies that individuals have a tendency to simplify social information and to reduce or minimize attention to discrepant information. This is especially true when we expect to have to pass on impressions to other people. If this is occurring in assessment centers, perhaps centers should be redesigned so that certain staff would provide the behavioral information but others would actually make the dimension rating. To take this one step further, one might speculate on the consequences for better construct validity if separate staff were used to make the overall assessment ratings or predictions of likely success with trait ratings as input from others.

Research on the impact of staff specialization (with regard to dimensions) or exercise specialization (e.g., where exercises are designed to produce behavior relevant to a single dimension) might reveal advantages of these approaches (see Cohen & Sands, 1978, for a parallel example). Finally, alternative methods of reaching overall assessment ratings might be
examined. Instead of having staff follow a majority or consensus-decision rule, dialectical or devil’s advocacy models could be incorporated. There is some evidence that the latter would produce better quality (more construct valid) evaluations (Schweiger, Sandberg, & Ragan, 1986; Zedeck, 1986). Other suggestions for research on the potential impact on group processes on overall assessment ratings can be found in the work of Klimoski, Friedman, and Weldon (1980).

But studies may reveal that it is just not possible to establish assessment centers as valid measures of constructs (at least in the number and variety of interest to managers). If this were to be the case, there would still appear to be any number of programs of research that could be built around some of the alternative explanations for the assessment center’s apparent predictive validity as presented in this paper. The point is, to establish the correctness of one or more of these would lead to sound recommendations for practice. This in turn, would increase our confidence that we actually know what we are doing in the use of whatever selection/promotion devices we end up with. It should also reduce the nagging feeling (as reported by one reviewer of this paper) that when assessment centers actually work, one has just been through some sort of a “voodoo rite.”

Conclusion

For all we do know about assessment centers, we don’t know enough. We know that these procedures are useful for predicting managerial success. Nevertheless, the predictive validity of assessment centers remains a puzzle. We agree with Russell (1985) when he states, “although assessors are apparently not doing what assessment center architects thought they were doing, the reasons behind assessment center predictive validity remain unknown” (p. 743).

In this paper we have attempted to define some of the pieces used in putting the assessment center picture together. By identifying some of the possible explanations for the predictive validity of assessment centers, we believe we provide a framework for a better understanding of the assessment center method.

REFERENCES


Bray DW, Grant DL. (1966). The assessment center in the measurement of potential for business management. Psychological Monographs, 80 (17 whole No. 625).


**Personnel Psychology** publishes empirical applied research dealing with a wide range of personnel problems facing public and private sector organizations. Articles deal with all aspects of personnel psychology, including employee selection, training and development, job analysis, productivity improvement programs, work attitudes, labor-management relations, and compensation and reward systems. Critical state-of-the-art reviews of the research literature in these areas are also published regularly.

**Personnel Psychology** serves a diverse audience: operating personnel managers, psychologists working in organizational settings, and academicians in psychology departments and schools of business with interests in personnel and human resource management. To meet the needs of these various groups, manuscripts are judged on the basis of three criteria:

1. *Technical soundness*—Papers are expected to report adequately designed, well-controlled research.
2. *Readability*—Studies must be well presented, and the implications of the research must be understandable to the personnel executive whose training may have been in fields other than psychology and statistics.
3. *Practicality*—While articles dealing with basic research problems are considered, priority is given to papers whose implications are for immediate and general problems.

Authors are requested to prepare their manuscripts according to the directions given in the *Publication Manual of the American Psychological Association—Third Edition* (1983). Authors who wish to do so are encouraged to submit their work in a form appropriate for blind review.

*Send manuscripts, in triplicate, to:*

**Dr. Paul Sackett,** Editor  
**Personnel Psychology**  
Department of Psychology  
University of Illinois at Chicago  
Box 4348  
Chicago, IL 60680