A Person Perception Explanation for Validation Evidence from Assessment Centers

Robert G. Jones

Department of Psychology, Southwest Missouri State University
Springfield, MO 65804

Consonant with recent person perception accuracy findings, recent longitudinal results from assessment center evaluations suggest that assessors may be accurate in understanding how emotional and motivational tendencies predict target behaviors in work contexts. The features of assessment center exercises and the characteristics of assessors which may explain these findings are discussed and suggestions for future research and practice provided. Contrary to some current arguments, it is suggested that certain types of trait-based dimensions may be essential to effective assessment center practice.

Assessment center (AC) research findings (Bray & Howard, 1983; Howard & Bray, 1989; Jones & Whitmore, 1995) suggest that the predictive power of assessment centers (ACs) is based partly on assessors' accurate perceptions of assessees' motivation. This is good news for person-perception-based methods of inquiry, but seems to fly in the face of substantial evidence that people are not especially accurate in their appraisals of one another generally (see Funder, 1987; Heneman, 1986; Murphy & Cleveland, 1995), and in assessment centers particularly (Russell, 1985). The suggestion that people can accurately assess each other under certain conditions, and regarding certain characteristics, leads to some significant questions about findings which have demonstrated our inability to be accurate in appraising each other.

In this paper, I will use existing conceptual and operational definitions of accuracy to attempt to reconcile these findings. The concept of "motive traits," which appear to be the basis of assessor judgment accuracy, will be developed to explain the phenomenon of assessment center prediction. In addition, I will suggest ways that motive traits can be used to explain anomalous findings from assessment center research. Specific characteristics of assessment centers which may provide the context for accurate judgments will be discussed, and suggestions for research and practice offered.

Author's Note: E-mail may be sent via Internet: rgj63f@vmu.smu.edu.

DEFINITIONS OF ACCURACY

The primary difference between assessment center findings and other research findings regarding person perception is based on the conceptual and operational definitions of accuracy. Following Kruglanski's (1989) conceptual distinctions, the vast majority of research in person perception accuracy has relied on correspondence and consensus definitions of accuracy. Here, observers' judgments and memories are compared with objective features of the stimulus situation and the judgments of others. The others against whom observers' judgments are compared usually include experts and targets themselves. These comparisons between observers' reports and other criteria are transformed into statistics which index level and ranking differences (Sulsing & Balzer, 1988), and which tend not to relate to one another.

Operationally, these indices of accuracy rely almost exclusively on laboratory settings, where observation and evaluation of accuracy are accomplished within a short time of each other. This leaves much of the research that has shown the inadequacy of person perception open to the problem of single operations (Greenwald, Pratkanis, Leippe, & Baumgardner, 1986; Russell, 1994). Put simply, the findings may be a result of the narrow range of operational definitions of the phenomena of interest. Other findings might be obtained in other contexts, using other operations. For assessment centers, the temporal context may be a problem in this regard.

Evaluations of assessment centers rely on what Kruglanski (1989) refers to as "utility" accuracy. This is the comparison of observer judgments to later outcomes of importance. In the case of assessment centers, there is ample evidence to support the "utility" accuracy of assessor judgments (see Gaugler, Rosenthal, Thornton & Benton, 1987). Of particular interest are the findings of Howard and Bray (1989) and Bray and Howard (1983) that assessor judgments predicted a significant group of important target behaviors more than thirty years after assessment. These differences in conceptual and operational definition of accuracy can account for the contradictory findings of person perception and assessment center research.

It is also likely that the nature of the judgments made in assessment centers differs from judgments in person perception research. While both assessment centers and social psychological laboratories control the stimulus situation to a significant extent, the nature of this control differs. Specifically, assessment centers are designed to present targets and observers with a high fidelity simulation of the stimuli experienced in jobs (usually manager jobs) for which assessees are being judged. The complexity of this situation, both in terms of its cognitive and social demands, differs from the relatively simple situation of many laboratories. Thus, assessors are able to observe a wider range of possible behaviors within a relatively constant setting. This "richer" opportunity for observation may help to provide more accurate judgments (see Bray & Howard, 1983).

In addition, the importance of accurate assessment in assessment centers distinguishes assessment centers from most other person perception research. Given the potential importance of accurate judgment to both the organization and the individual assessor, there is probably a greater incentive to provide accurate evaluations in assessment centers than in many other research settings.

Finally, in a very real sense, assessors are expert in this judgment task. Typical assessment centers employ experienced managers from the host organization as assessors. They are therefore likely to be expert about both the ambient social influences of the organization generally, and the discretionary social influences (Hackman, 1992) associated with managers' roles. Assessors are therefore qualified to provide "true scores" (Allen & Yen, 1979) against which other observers' judgments can be compared.

MOTIVE TRAITS

Longitudinal findings (Bray & Howard, 1983; Jones & Whitmore, 1995) suggest another possible reason for the contradictory findings of assessment center and person perception research. These studies have demonstrated that a significant portion of the accuracy of assessment center ratings comes from judgments of assessees motivation. Bray and Howard (1983) demonstrated how a "need for advancement" motive predicted advancement and various life satisfaction measures. Jones and Whitmore's (1995) findings showed "career motivation" judgments predicting advancement and developmental activity when all other ratings were accounted for.

Such motivation judgments are not typical of the sorts of judgments made in other organizational or social psychological person perception research. More commonly, in organizational research, job-related cognitive dimensions such as communication skill, decision-making ability, interpersonal skills and "overall performance" are the objects of judgment. It is striking that a prominent judgment task similar to this in the person perception research is the recognition of emotion task (see Russell,

---

1 Thanks to an anonymous reviewer for this point.
1994, for a recent review). The evidence from this research is somewhat supportive of accurate person perception.

Similarly, only very recently has work by Funder (1995) and others (e.g., Gifford, 1994) in person perception accuracy shown that personality can be accurately inferred from various mediating behavioral information. In particular, information about emotions was indicative of personality in Funder’s research (1995).

It appears therefore, that an essential characteristic which separates assessment center judgment accuracy from most other evaluations of judgment accuracy is the use of motive-based dimensions. Borrowing from various authors (Klinger, 1975; Little, 1989), I suggest that assessor judgments on these dimensions are founded on an awareness of the “motive traits” of assesses. Such motive traits are affective and behavioral dispositions that arise from long-term, persistent goals. As motives, they are predictive of the emotional responses and other motive “states,” which are often characterized as transient (Lang, 1995). Motive traits, though they may be less than permanent, are by definition more persistent than these more transient states. This is because they are based on “personal values, goals, and interests” that “tend to have relatively high stability even over long periods of time” (Mischel, 1990, p. 120).

It may not be obvious to assessors what assesses’/targets’ specific goals are. But by combining relatively constant contextual information and significant observation of behavior in a motive-rich situation, assessors are able to make reasonably accurate judgments of target motive traits. Placed in the larger context of management work in a specific organization, assessors apparently are able to predict whether targets’ motives will lead them to advancement in the organization. This follows closely Klimoski and Brickner’s (1987) discussion of how assessors may use contextual information to reach judgments about “promotability” in this setting. That is, assessors can establish a link between motivated target behavior in the simulation situation and behavior in the “real world” situation of a management job in the organization.

Specific cues for making these judgments probably relate to the emotive responses of assesses to various social stimuli. This is based on two propositions. First, emotions can be thought of as “states of preparedness” with respect to motives (Lang, 1995). Second, emotions serve a social signalling function (Ashforth & Humphrey, 1993; Russell, 1994). Therefore, emotional displays can serve a signaling function regarding motives. For example, if an assessees gets angry with the “subordinate” in a role-playing exercise, the assessor may realize that the assessees’s motives are blocked when their work outcomes are actualized through the work of others. This would make them unlikely to perform well in a management setting. Notice that, even though the specific motives of the assessees are not clear, the assessor is able to reach a meaningful conclusion about how these motives may fit into a given social context.

Evidence about the prediction of leadership emergence (Hogan, Curphy, & Hogan, 1994) is not inconsistent with this notion. This evidence has shown that leadership emergence (but not necessarily success) can be predicted using a host of personality measures. Following Mischel (1990), it is likely that these measures tap motive traits. This does not diminish the broad usefulness of such measures. Instead, by referring to individual motive traits, researchers allow for greater recognition of the idiosyncratic motives of individuals. Also, rather than "stamping out" using predetermined personality categories, assessment centers maintain the potential integrity of the natural person perception categories of assesses. Thus, the value of assessment centers may rely on the high fidelity of motive trait categories.

**EXPLAINING ASSESSMENT CENTER VALIDATION EVIDENCE**

I have argued elsewhere (Jones, 1992) that assessment center validation evidence requires much closer scrutiny than it has received heretofore. The motive trait explanation for assessment center predictive accuracy helps to inform the ongoing consideration of these issues. First, the tendency (noted by Klimesh & Maucland, 1977) of assessment centers to predict advancement better than performance is probably related to the broad context of assessment center simulation. While I have suggested that the simulation context is somewhat narrow, assessment centers are not typically used to evaluate knowledge and skill dimensions which often differentiate high from low performers in a job. That is, selection into a job is usually based at least partly on more stable “ability and other characteristic” dimensions (Harvey, 1991). Aside from selecting people based on ability dimensions, this would tend to truncate the range of performance on those dimensions in the few people selected into a job. And, if my suggestions are correct, assessment centers primarily evaluate ability dimensions which predict motivation and performance stability, versus job-specific knowledge dimensions. Hence the lack of knowledge and skill prediction and the fidelity of ability prediction.

The lack of discriminant-validity evidence from assessment centers also can be explained by reference to motive traits. Most assessment center designers use “behavior” characteristics as the nominal basis of their dimensions. There are several good reasons which can be given for this choice, but it means that dimensions which explicitly evaluate the fit
of motive traits to management jobs are sometimes not included or are poorly represented in center evaluation dimensions. This lack of explicit inclusion of assesse motive traits in assessment dimensions may not encourage experienced assessors from basing their judgments on motive information. This would introduce noise into other ratings, which would tend to inflate “common method variance.” Hence, one form of construct validation evidence (multitrait-multimethod analyses) is not supportive of dimension-based ratings.

Thus, although some have recommended dropping dimensions in favor of behaviors in assessment centers (e.g., Lowry, 1995), the use of behavioral information may actually obscure the nature of assessor expertise. Specifically, expertise is usually defined in terms of use of higher order categories, rather than specific instances, to form judgments (Glaser & Chi, 1988; Hardiman, Dufresne, & Mestre, 1989). Applying this to behavioral versus categorical (KSAO) information, one should not expect to find behavioral information necessarily to be patterned in the fashion that assessment center architects (who may not themselves be job or job context experts) devised. Instead, assessors may use the more abstract motivational and contextual categories associated with personality measures. As described above, the specific types of dimensions used in the past may have obscured assessor expertise by using knowledge and skill, rather than trait-based dimensions. In addition, when individual behaviors are observed, they are observed in relation to these higher order categories of “good managers” and “management situations.” Though behaviorally rich assessment center exercises are essential, it is not behaviors in themselves that make assessment centers useful.

Finally, although a correlation between assessment center dimension ratings and mental-ability test results has been demonstrated (see Khinoski & Brickenk, 1987), there is by no means a perfect relation between the two. This may also be due to the importance of motivation to assessment center ratings. Mental-ability measures are typically timed, which tends to hold motivation constant (i.e., they are maximum performance measures). As we have seen, assessment center ratings are probably based at least partly on qualitative variability in motivation across stimulus situations. This does not mean that assessment center ratings are “typical” performance measures in the conventional sense. Instead, assessor judgments may be based on limitations in motivation within situational contexts. Consequently, there is a need to compare untimed mental ability measures with assessment center ratings.

WHY DO ASSESSMENT CENTERS "WORK"?

This is analogous to Funder’s (1989) question "Why does the dancing bear dance?" While assessment centers are predictive, and their predictive accuracy appears to derive at least partly from motive traits, it should be noted that correlations between motive trait measures and criteria are only in the 0.30 range. This means that all sorts of possible explanations for the variance that is not accounted for remain. Two suggestions I have made for this (Jones, 1992) are (a) that our current methods are inadequate to tap the complexity of human mental processes (I think most of us would readily agree to this) and (b) that there is still some inaccuracy in our perceptions. So, even though the dancing bear dances, it doesn’t dance all that well.

Additionally, the structures of assessment centers may provide a particularly helpful dance floor. Several aspects of the judgments facilitate accuracy. These include the expertise of the observers (who are generally experienced both as managers and as members of the sponsoring company), the "typicalness" of the assessment center exercises to the management job (leaving plenty of room for observing individual differences, especially in motivation), and the involvement of experienced psychologists in the group process of assessor rating.

What does this mean for the future of assessment centers? First, this approach to assessment centers argues for more explicit and broader use of AO dimensions in evaluating assessment center performance and prediction. Relying on what assessors appear to do well seems like a more viable strategy than relying on person perception of knowledge and skill (or behavioral) dimensions. In fact, further attention could be paid to understanding the expertise of assessors, with the hope of developing and refining such expertise in others (see Lorenzo, 1984). It may be that only very well-trained assessors are capable of great accuracy in perception of motive trait information.

Second, attempts should be made to explicitly define the nature of these categories within exercises. In particular, Funder’s realistic accuracy model (1995) might receive viable tests in assessment center samples. In any case, evaluations of assessment centers would do well to follow this model’s call for multiple criteria across multiple realistic circumstances; assessment centers may provide particularly fertile ground for this. Related to this, the definition of the situational influences on assesse behavior should be given greater attention. Even in a behavior-rating approach, analysis of situational influences through methods like template matching should provide important information about constraints on behavior in exercises. One initial attempt to do this by Highhouse and Harris (1993) met with some success and should be tested further.
Finally, assessment centers should continue to use independent ratings and behavioral reporting as part of the integration of individual ratings. Such use of multiple views to inform situation and person-based ratings has been shown to improve other person-perception-based practices (e.g., Campion, Pursell, & Brown, 1988), and should be researched further in assessment centers.

REFERENCES


Assessing the Assessor: The Relationship of Assessor Personality to Leniency in Assessment Center Ratings

Lynn K. Bartels
Department of Psychology, Southern Illinois University
Edwardsville, IL 62026

Dennis Doverspike
Department of Psychology, University of Akron
Akron, OH 44325

This study examined the relationship between assessor personality and assessment center rating leniency. Middle and upper management assessors took the 16PF and their ratings of assessment center candidates were examined in relation to their own personality profiles. Intelligence, Sensitivity and Tough Poise were significantly related to leniency in assessment center ratings. The results suggest that tenderminded, warmhearted assessors tend to give more elevated ratings. Agreeable personality characteristics may also be related to lenient assessment center ratings.

Despite the large number of studies conducted on assessment centers (e.g., See Gaugler, Rosenthal, Thornton, & Bentson, 1987; Thornton, 1992 for a review), certain areas remain underinvestigated. One important area that needs to be researched involves the people who rate the assessment center candidates—the assessors. Very little is known about how characteristics of the assessor affect the quality of their ratings. Thus, the focus of the current research is on “assessing the assessor.”

In most assessment centers, assessors play a demanding and critical role in the assessment process. Typically, assessors are trained to observe, record, classify and evaluate assessed’ behavior. During the assessment center, assessors observe assessment center candidates performing situational exercises and take notes about the candidates’ behaviors. They may interact with candidates during role plays. After each exercise, assessors write reports summarizing the behaviors observed. After several days of observation and report writing, the assessors may meet to discuss their observations and make evaluations or ratings of potential. In