CHAPTER 31

HOLISTIC ASSESSMENT FOR SELECTION AND PLACEMENT

Scott Highhouse and John A. Kostek

Holism in assessment is a school of thought or belief system rather than a specific technique. It is based on the notion that assessment of future success requires taking into account the whole person. In its strongest form, individual test scores or measurement ratings are subordinate to expert diagnoses. Traditional standardized tests are seen as providing only limited snapshots of a person, and expert intuition is viewed as the only way to understand how attributes interact to create a complex whole. Expert intuition is used not only to gather information but also to properly execute data combination. Under the holism school, an expert combination of cues qualifies as a method or process of measurement. For example, according to Ruscio (2003), "Holistic judgments are premised on the notion that interactions among all of the information must be taken into account to properly contextualize data gathered in a realm where everything can influence everything else" (p. 1). The holistic assessor views the assessment of personality and ability as an ideographic enterprise, wherein the uniqueness of the individual is emphasized and nomothetic generalizations are downplayed (see Allport, 1962). This belief system has been widely adopted in college admissions and is implicitly held by employers who rely exclusively on traditional employment interviews to make hiring decisions. Milder forms of holistic belief systems are also held by a sizable minority of organizational psychologists—ones who conduct managerial, executive, or special-operation assessments.

In this chapter, the roots of holistic assessment for selection and placement decisions are reviewed and the applications of holistic assessment in college admissions and employee selection are discussed. Evidence and controversy surrounding holistic practices are examined, and the assumptions of the holistic school are evaluated. That the use of more-standardized procedures over less-standardized ones invariably enhances the scientific integrity of the assessment process is a conclusion of the chapter.

HISTORICAL ROOTS

The traditional testing and measurement tradition is associated with people such as Sir Francis Galton and James McKeen Cattell (see DuBois, 1970, for a review). The holistic assessment tradition for selection and placement, however, was developed by psychologists outside of this circle. The intellectual forefathers of holistic assessment were influenced by gestalt concepts and were concerned with personality diagnosis for the purposes of selecting officers and specialists during World War II. The most prominent of these were Max Simoneit of Germany, W. R. Bion of England, and Henry A. Murray of the United States.

Max Simoneit

Max Simoneit was the chief of German military psychology during World War II. The Germans believed that victory depended on the superior leadership and intellect of the officer (Ansbacher, 1941). Simoneit, therefore, believed that psychological diagnosis (i.e., character analysis) of officer candidates and specialists should be the primary focus of

DOI: 10.1037/14047-031
APA Handbook of Testing and Assessment in Psychology: Vol. 1. Test Theory and Testing and Assessment in Industrial and Organizational Psychology, K. F. Geisinger (Editor-in-Chief)

military psychology. Assessments were qualitative rather than quantitative and subjective rather than objective (Burt, 1942). Simoneit believed that intelligence assessment was inseparable from personality assessment (Harrell & Churchill, 1941) and that an officer candidate needed to be observed in action to assess his total character. Although little is known about Simoneit, it is believed that he studied under the psychologist Narziss Ach (Ansbacher, 1941). Ach believed that willpower could be studied experimentally using a series of nonsense syllables as interference while a subject attempted to produce a rhyme (Ach, 1910/2006). As with Ach, assessment of will power was a central theme in Simoneit's work (Geuter, 1992). He devised tests such as obstacle courses that could not be completed and repeated climbs up inclines until the candidate was beyond exhaustion (Harrell & Churchill, 1941). These tests were accompanied by diagnoses of facial expressions, handwriting, and leadership role-plays. Simoneit's methods were seen as innovative, and the use of multiple and unorthodox assessment methods inspired officer selection practices used in Australia, Britain, and the United States (Highhouse, 2002).

W. R. Bion

W. R. Bion was trained as a psychoanalyst in England and became an early pioneer of group dynamics (Bion, 1959). He was enlisted to assist the war effort by developing a method to better assess officers and their likelihood of success in the field. According to Trist (2000), the British War Officer Selection Board was using a procedure in which psychiatrists interviewed officer candidates, and psychologists administered a battery of tests. This procedure created considerable tensions concerning how much weight to give to psychiatric versus psychological conclusions. Bion replaced this process with a series of leaderless group situations—inspired by the German selection procedures—to examine the interplay of individual personalities in a social situation. Bion believed that presenting candidates with a leaderless situation (e.g., a group carrying a heavy load over a series of obstacles) indicated their capacity for mature social relations (Sutherland & Fitzpatrick, 1945). More specifically, Bion believed that the pressure for the candidate to look good individually was put into competition with the pressure for the candidate to cooperate to get the job done. The challenge for the candidate was to demonstrate his abilities through the medium of others (Murray, 1990). Candidates underwent a series of tests and exercises over a period of 2.5 days. Psychiatrists and psychologists worked together as an observer team to share observations and develop a consensus impression of each candidate's total personality.

Henry A. Murray

Henry A. Murray was originally trained as a physician but quickly abandoned that career when he became interested in the ideas of psychologist Carl Jung. He developed his own ideas about holistic personality assessment while working as assistant director, and later director, of the Harvard Psychological Clinic in the 1930s. During the war, Murray was enlisted by the Office of Strategic Services (OSS) to develop a program to assess and select future secret agents. Murray's medical training involved grand rounds, in which a team of varied specialists contributed their points of views in arriving at a diagnosis. He believed that one shortcoming of clinical case studies was that they were produced by a single author rather than a group of assessors working together (Anderson, 1992). Accordingly, Murray and his colleagues assembled an OSS assessment staff that included clinical psychologists, animal psychologists, social psychologists, sociologists, and cultural anthropologists. Conspicuously absent from his team were personnel psychologists (Capshew, 1999). Murray developed what he called an organismic approach to assessment. The approach, described in detail in *Assessment of Men* (OSS, 1948), involved multiple assessors inferring general traits and their interrelations from a number of specific signs exhibited by a candidate engaged in role plays, simulations, group discussions, and in-depth interviews—and combining these inferences into a diagnosis of personality. Murray's procedures were the inspiration for modern-day assessment centers used for selecting and developing managerial talent (Bray, 1964).

The three figures discussed in this section were mavericks who rejected the prevailing wisdom that consistency is the key to good measurement.

Examiners were often encouraged to vary testing procedures from candidate to candidate and to give special attention to tests they preferred. In other words, there was little appreciation for the concepts of reliability and standardization. Although many celebrated the fresh approach brought about by the holistic pioneers, others questioned the appropriateness of many of their practices (Eysenck, 1953; Older, 1948).

APPLICATIONS IN SELECTION AND PLACEMENT

Much has been written on the application of holistic principles in clinical settings (see Grove, Zald, Lebow, Snitz, & Nelson, 2000; Korchin & Schuldberg, 1981), but their application to selection and placement decisions has received considerably less attention (cf. Dawes, 1971; Ganzach, Kluger, & Klayman, 2000; Highhouse, 2002). It is notable, however, that one of the earliest debates about the use of holistic versus analytical practices involved the employee selection decision-making domain (Freyd, 1926; Viteles, 1925). Morris Viteles (1925) objected to the then-common practice of making decisions about applicants on the basis of test scores alone. According to Viteles,

> It must be recognized that the competency of the applicant for a great many jobs in industry, perhaps even for a majority of them, cannot be observed from an objective score any more than the ability of a child to profit from one or another kind of educational treatment can be observed from such a score. (p. 134)

Viteles (1925) believed that the psychologist in industry must integrate test scores with clinical observations. According to Viteles, "His judgment is a diagnosis, as that of a physician, based upon a consideration of all the data affecting success or failure on the job" (p. 137). Max Freyd (1926) responded that psychologists are unable to agree, even among themselves, on a person's abilities by simply observing the person. Rather than diagnosing a job candidate, Freyd argued that the psychologist should make subjective impressions objective by incorporating them into a rating scale. According to Freyd,

> The psychologist cannot point to the factors other than test scores upon which he based his correct judgments unless he keeps a record of his objective judgments on the factors and compares these records with the vocational success of the men judged. Thus he is forced to adopt the statistical viewpoint. (p. 353).

Most modern-day organizational psychologists share Freyd's (1926) view of assessment for selection, but those who practice assessment at the managerial and executive level are less likely to do so (Jeanneret & Silzer, 1998; Prien, Schippman, & Prien, 2003).

The most common applications of holism in assessment and selection practice are discussed next. These include (a) college admissions decision making, (b) assessment centers, and (c) individual assessment.

College Admissions

Colleges and universities have continuously struggled with how to select students who will be successful while at the same time ensuring opportunity for underrepresented populations (see Volume 3, Chapters 14 and 15, this handbook, for more information on this type of testing). Standardized tests provide valuable information on a person's degree of ability to benefit from higher education. Admissions officers, however, are charged with ensuring a culturally rich and diverse campus and accepting students who will exhibit exceptional personal qualities such as leadership and motivation. In 2003, the U.S. Supreme Court (*Gratz v. Bollinger*, 2003) ruled that it is lawful for admissions decisions to be influenced by diversity goals, but that holistic, individualized selection procedures, not mechanical methods, must be used to achieve these goals (see McGaghie & Kreiter, 2005). This decision was in response to the University of Michigan's then practice of awarding points to undergraduate applicants based on, among other things, their minority status. These points were aggregated into an overall score, according to a fixed, transparent formula. Justice Rehnquist argued

that consideration of applicants must be done at the individual level rather than at the group level. Race, according to the majority decision, is to be considered as one of many factors, using a holistic, case-by-case analysis of each applicant. In his dissenting opinion, Justice Souter argued that such an approach only encourages admissions committees to hide the influence of (still illegal) racial quotas on their decisions (*Gratz v. Bollinger*, 2003).

In 2008, Wake Forest University became the first top-30 U.S. university to drop the standardized test requirement for undergraduate admissions. Wake Forest moved to a system in which every applicant is eligible for an admission interview (Allman, 2009). The Wake Forest interviews do not follow any specific format, and interviewers are free to ask different questions of different applicants. Although the Wake Forest interviewers make overall interview ratings on a scale ranging from 1 to 7, the admissions committee explicitly avoids using a numerical weight in the overall applicant evaluation (Hoover & Supiano, 2010). Wake Forest is a clear exemplar of what Cabrera and Burkum (2001) referred to as the holistic era of college admissions in the United States.

Assessment Centers

The notion that psychologists could select people for higher level jobs was not widely accepted until after World War II (Stagner, 1957). The practices used to select officers in the German, British, and U.S. militaries were seen as having considerable potential for application in postwar industry (Brody & Powell, 1947; Fraser, 1947; Taft, 1948). Perhaps most notable was Douglas Bray's assessment center (Bray, 1964). Bray, inspired by the 1948 OSS report *Assessment of Men*, put together a team of psychologists to implement a program of tests, interviews, and situational performance tasks for the assessment of the traits and skills of prospective AT&T managers. Although the original assessment center was used exclusively for research, the procedure evolved into operational assessment centers still in use today. Unlike the original, clinically focused center, the operational assessment centers of today focus on performance in situational exercises, and they commonly use managers as assessors. The focus on standardization and objective rating is in contrast to the earlier holistic practices advocated by the World War II psychologists (Highhouse, 2002). One similarity that remains between the modern and early assessment centers, however, is the use of rater consensus judgments. The consensus judgment process is predicated on the notion that observations of behavior must be intuitively integrated into an overall rating (Thornton & Byham, 1982). This consensus judgment process involves discussion of everyone's ratings to arrive at final dimension ratings and ultimately an overall assessment rating for each candidate. The group discussion process can take several days to complete and does not involve the use of mechanical or statistical formulas.

Individual Assessment

One area of managerial selection practice that has maintained the holistic school's emphasis on considering the whole person and intuitively integrating assessment information into a diagnosis of potential is commonly referred to as *individual assessment* (see Ryan & Sackett, 1992). Although the label is not very descriptive, it does emphasize the focus on idiographic (as opposed to nomothetic) assessment. Practices vary widely from assessor to assessor, but individual assessment typically involves intuitively combining impressions derived from scores on standardized and unstandardized psychological tests, information collected from unstructured and structured interviews, a candidate's work and family history, informal observation of mannerisms and behavior, fit with the hiring organization's culture, and fit with the job requirements. The implicit belief behind the practice is that the complicated characteristics of a high-level job candidate must be assessed by a similarly complicated human being (Highhouse, 2008). According to Prien, Shippmann, and Prien (2003, p. 123), the holistic process of integration and interpretation is a "hallmark of the individual assessment practice" (see also Ryan & Sackett, 1992).

Next, the evidence and controversy surrounding the use of holistic methods for making predictions about success in educational and occupational domains are reviewed. A summary of studies that have directly contrasted holistic versus analytical approaches in college admissions and employee selection is also provided.

EVIDENCE AND CONTROVERSY

The first study to empirically test the notion that experts could better integrate information holistically than analytically was conducted by T. R. Sarbin in 1942. Sarbin followed 162 freshmen who entered the University of Minnesota in 1939. Using a measure of 1st-year academic success, Sarbin compared the earlier prediction of admissions counselors with a statistical formula that combined high school rank and college aptitude test score. The counselors had access to these two pieces of information as well as information from additional ability, personality, and interest inventories. The counselors also interviewed the students before the fall quarter of classes. Of interest to Sarbin was the performance of the simple formula against the counselors' predictions. The results showed that the counselors, who had access to all of the test data and interview observations, did significantly worse in predicting 1st-year success than the simple (high school rank plus aptitude test score) formula. Subsequent studies on college admissions have supported the idea that a simple combination of scores is not only effective but is in many instances more effective than holistic assessment for predicting success in school.

Table 31.1 provides a summary of research comparing holistic to analytical approaches to college admissions. The table shows that in almost every case, holistic evaluations based on test scores, grades, and other personal evaluations (e.g., interviews, letters, biographical information) were equaled or exceeded by simple combinations of standardized tests scores and grades.

Organizational psychologists took note of findings like Sarbin's (1942), which were documented

TABLE 31.1

Empirical Comparisons of Holistic and Analytical Approaches to College Admissions

Source	Method	Results
Alexakos (1966)	Guidance counselors made predictions of college GPA on the basis of information collected from testing and interviews over a 4-year period.	The holistic judgments of counselors were slightly outperformed by a statistical combination of high school GPA, standardized test scores, and demographic variables.
Dawes (1971)	Psychology faculty predicted the success of incoming graduate students on the basis of a standardized test, undergraduate GPA, and letters of recommendation.	A mechanical model of the committee's judgment process predicted faculty ratings of graduate success better than the committee itself.
Hess (1977)	A medical school admissions committee predicted success in 1st-year chemistry on the basis of standardized tests, interviews, transcripts, and biographical data.	The holistic judgments of the committee did not predict success in 1st-year chemistry, whereas high school performance data alone were successful.
Rosen & Van Horn (1961)	A scholarship award committee predicted 1st-year GPA using high school rank, standardized test scores, biographical information, and letters of recommendation.	The holistic judgment of the award committee was equal to the use of only high school rank in predicting 1st-semester GPA in college.
Sarbin (1943)	College admissions officers predicted success on the basis of high school rank, a standardized test, and an intensive interview.	The holistic judgments of the admissions officers were inferior to a simple combination of high school rank and standardized test score.
Schofield (1970)	A medical school admissions committee predicted success on the basis of college GPA, a standardized test, biographical data, and letters of reference.	The holistic judgment of the admissions committee was equal to a statistical combination of only college GPA and standardized test scores.
Watley & Vance (1964)	Guidance counselors made predictions of college GPA and participation in activities, using high school rank, standardized tests, and biographical information.	The holistic judgments of counselors equaled a mechanical formula that included high school rank and test scores.

Note. Only studies using actual counselors, faculty, or admissions officers as assessors are included in this table.

in Paul Meehl's classic 1954 book *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*. Moreover, early organizational studies seemed to support Meehl's findings that holistic integration of information was not living up to the claims of the personality assessment pioneers (e.g., Huse, 1962; Meyer, 1956; Miner, 1970). However, an influential review of judgmental predictions in executive assessment—dismissing the relevance of this controversy to the organizational arena (Korman, 1968, p. 312)—eased the mind of many industrial psychologists involved in assessment practice. Also, assessment center research was showing impressive criterion-related validity, suggesting that an approach with many subjective components could be quite useful in identifying effective managers (Howard, 1974).

Table 31.2 provides a summary of research comparing holistic to analytical approaches to selection and placement in the workplace. Although the studies vary in rigor and sometimes do not provide fair comparisons of holistic and analytical approaches, some broad inferences can be drawn from this compilation:

- There are surprisingly few studies on the relative effectiveness of holistic assessment for employee selection, especially as it regards individual assessment.
- Only one study clearly favored holistic assessment (i.e., an assessor with knowledge of a cognitive ability test score did better than the score alone; Albrecht, Glaser, & Marks, 1964), compared with at least five that clearly favored analytical approaches, and at least seven that were a draw.
- The few studies to examine the incremental validity of holistic judgment have not provided encouraging results.

Our summary shows that evidence for the superiority of holistic judgment is quite rare in educational and employment settings. A meta-analysis comparing clinical to statistical predictions in primarily medical and health diagnosis settings found that statistical methods were at least equal to clinical methods in 94% of the cases and significantly superior to them in as much as 47% of studies (Grove et al., 2000). Despite the fact that clinicians often had access to more information than the formulas, the statistical methods were estimated to be approximately 10% better in overall accuracy.

Recall that advocates of the holistic school have suggested that experts may take into account the interactions among various pieces of assessment evidence and understand the idiosyncratic meaning of one piece of information within the context of the entire set of information for one candidate (Hollenbeck, 2009; Jeanneret & Silzer, 1998; Prien et al., 2003). Given such expertise, one might expect that holistic judgments—which consider all of the information at hand—should unequivocally outperform dry formulas based on ratings and test scores. This has not been the case.

The existing research on selection and placement decision making has provided disappointingly little evidence that subjectivity and intuition provide added value. Traditional employment interviews provide negligible incremental validity over standardized tests of cognitive ability and conscientiousness (Cortina, Goldstein, Payne, Davison, & Gilliland, 2000; see Chapter 27, this volume, for more information on employee interviews). Research has also unequivocally shown that the more the interview is structured or standardized to look like a test, the greater its utility for predicting on-the-job performance (Conway, Jako, & Goodman, 1995; McDaniel, Whetzell, Schmidt, & Maurer, 1994). Having assessors spend several days discussing job candidates in assessment centers, and arriving at an overall consensus rating for each, provides no advantage over taking a simple average of each person's ratings (Pynes, Bernardin, Benton, & McEvoy, 1988). Assessing a candidate's fit with the job—a common practice in individual assessment—also appears to provide little advantage in predicting a candidate's future job performance (Arthur, Bell, Villado & Doverspike, 2006). Taken together, this research has suggested that considering each candidate as a unique prediction situation has not resulted in demonstrably better prediction. As Grove and Meehl (1996) noted in their review of the debate between ideographic versus nomothetic views of prediction: "That [debate] is clearly an empirical question rather than a purely philosophical one decidable from the armchair, and empirical

TABLE 31.2

Empirical Comparisons of Holistic and Analytical Approaches to Employee Selection and Placement

Source	Method	Results
Albrecht, Glaser, & Marks (1964)	Psychologists ranked ~~of~~ managers on the basis of an intensive interview, cognitive ability tests, and projective tests.	The holistic judgments of psychologists outperformed the cognitive ability test score alone.
Borman (1982)	Military recruiters provided assessment center exercise effectiveness ratings and consensus overall assessment ratings.	A mechanical combination of unit-weighted exercise ratings slightly outperformed the holistic discussion-based judgments.
Feltham (1988)	Assessors provided exercise scores and consensus overall assessment ratings in a police assessment center.	A unit-weighted composite of exercise scores outperformed the holistic discussion-based judgments.
Ganzach, Kluger, & Klayman (2000)	Judgments of interviewers from the Israeli military were used as predictors of military transgressions.	Adding a holistic interviewer rating to mechanically combined interview dimension ratings slightly increased the prediction of the criterion.
Huse (1962)	Psychologists made final ratings on the basis of an intensive interview and standardized and projective tests.	The validities of holistic ratings based on complete data were not higher than validities based solely on standardized (paper-and-pencil) tests.
Meyer (1956)	Manager judgments were made on the basis of interview and standardized test scores.	Four of the five validity coefficients for holistic judgments were below the validity of a cognitive ability test alone.
Mitchel (1975)	Assessors provided overall potential ratings on the basis of exercise performance and test scores.	The multiple correlation of the predictors strongly outperformed the holistically derived overall assessment, but the two converged over cross-validation.
Pynes, Bernardin, Benton, & McEvoy (1988)	Assessors provided preconsensus and postconsensus dimension ratings and overall consensus ratings in a police assessment center.	The mechanically and holistically derived dimension ratings were indistinguishable ($r = .83$) and correlated strongly with the overall holistic judgment ($r = .71$ for both).
Roose & Doherty (1976)	Manager judgments were made on the basis of 64 cues from personnel files, including test, biographical, and objective interview data.	The mean increase in R^2 achieved by adding the holistic combination of cues by the judges over a linear combination of cues was 0.7%.
Sackett & Wilson (1982)	Assessors provided preconsensus and postconsensus dimension ratings and overall consensus ratings in a managerial assessment center.	A simple average of dimension ratings predicted postdiscussion ratings 93.5% of the time.
Trankell (1959)	One psychologist made predictions of Swedish airline pilot success in training on the basis of observations and standardized test scores.	The holistic evaluations slightly outperformed each of the test scores alone.
Tziner & Dolan (1982)	Assessors subjectively combined ratings, cognitive ability tests, and exercise ratings into an overall assessment.	The *R* of the predictors outperformed the holistically derived overall assessment.
Wollowick & McNamara (1969)	Assessors subjectively combined tests, dimension ratings, and exercise ratings into an overall assessment.	The *R* of the predictors strongly outperformed the holistically derived overall assessment.

evidence is, as described above, massive, varied, and consistent" (p. 310).

As such, the holistic approach to selection and placement as commonly practiced in hiring and admissions is not consistent with principles of evidence-based practice (Highhouse, 2002, 2008).

ASSUMPTIONS OF HOLISTIC ASSESSMENT

Given that the early promise of the holistic approach has not held up to scientific scrutiny, it is reasonable to ask why many people continue to hold this point

of view. Some common assumptions held by holistic assessors are outlined here.

Assessors Can Take Into Account Constellations of Traits and Abilities

Advocates of holistic assessment have argued that the expert combination of information is a sort of nonlinear geometry that is not amenable to standardization in some sort of simple formula (Prien et al., 2003, p. 123). This argument implies that holistic assessment is a sort of mystical process that cannot be made transparent. Ruscio (2003) compared it with the arguments of astrologers who, when faced with mounds of negative scientific evidence, reverted to whole-chart interpretations to render their professional judgments. Aside from the logical inconsistencies involved with the claim that assessors can take into account far more unique configurations of data than can be cognitively processed by humans, considerable evidence has shown that simple linear models perform quite well in almost all prediction situations faced by assessors (e.g., Dawes, 1979).

It has long been recognized that it is possible to include trait configurations in statistical formulas (e.g., Wickert & McFarland, 1967). However, very little research on the effectiveness of doing so in selection settings has been conducted, likely because predictive interactions are quite rare. Dawes (1979) noted that relations between psychological variables and outcomes tend to be monotonic. In contrast to conventional wisdom, nonmonotonic interactions (e.g., certain types of leaders are really good in one situation and really bad in another situation) are quite rare. Furthermore, the evidence has suggested that assessors could not make effective use of such interactions, even if they existed.

Assessors Can Identify Idiosyncrasies That Formulas Ignore

Meehl (1954) described the "broken leg case" in which a rare event may invalidate a prediction made by a formula. Meehl used the example of predicting whether Professor X would go to the cinema on a particular Friday night. A formula might take into account whether the professor goes to the movie on rainy or sunny days, prefers romantic comedies to action movies, and so forth. The formula may not, however, take into account the fact that Professor X broke his leg on the previous Monday. A human assessor could take into account such broken-leg cues. Although the example is compelling and is commonly used to justify the use of holistic assessment procedures, evidence has not supported the usefulness of broken-leg cues (see Camerer & Johnson, 1991). The problem seems to be that assessors overrely on idiosyncratic cues, not distinguishing the useful ones from the irrelevant ones. Assessors find too many broken legs.

Assessors Can Fine Tune Predictions Made by Formulas

A related argument is the idea that assessors may use their experience and wisdom to modify predictions that are made mechanically (Silzer & Jeanneret, 1998). The problem with this argument is that it assumes that a prediction can be fine tuned. As noted by Grove and Meehl (1996),

> If an equation predicts that Jones will do well in dental school, and the dean's committee, looking at the same set of facts, predicts that Jones will do poorly, it would be absurd to say, "The methods don't compete, we use both of them." (p. 300)

If a mechanical procedure determines that an executive is not suitable for a position as vice president, then fine tuning the procedure involves overruling the mechanical prediction. Certainly, intuition could be used to alter the formula-based rank ordering of candidates. We have yet to find evidence that this results in an improvement in prediction of job performance.

Some Assessors Are Better Than Others

There are experts in many domains, but evidence for expertise in intuitive prediction is lacking. The renowned industrial psychologist Walter Dill Scott concluded long ago, "As a matter of fact, the skilled employment man probably is no better judge of men than the average foreman or department head" (Scott & Clothier, 1923, p. 26). Subsequent research on assessment centers has found few differences among assessors in validity (Borman, Eaton, Bryan, & Rosse,1983). Similar findings have emerged for

the employment interview (Pulakos, Schmitt, Whitney, & Smith, 1996). After reviewing research on predictions made by clinicians, social workers, parole boards, judges, auditors, and admission committees, Camerer and Johnson (1991) concluded, "Training has some effects on accuracy, but experience has almost none" (p. 347). The burden of proof is on the assessor to demonstrate that he or she can predict better than someone with rudimentary training on the qualities important to the assessment.

Candidates for High-Level Jobs Do Not Differ Much on Ability and Personality

One common assumption of holistic assessment is that variability of test scores is restricted for people being selected at the highest levels of organizations. As one example, Stagner (1957) contended about executive assessment that

> simple, straightforward tests of intelligence and other objective measures seem not to have too much value, largely because an individual is not considered for such a position until he has already demonstrated a high level of aptitude in lower level activities. (p. 241)

Large-scale testing programs at Exxon and Sears in the 1950s, however, demonstrated that using a psychometric approach to identifying executive talent can be quite effective (Bentz, 1967; Sparks, 1990). Personality tests better predict behavior for jobs that provide more discretion (Barrick & Mount, 1993), and the validity of cognitive ability measures increases as the complexity of the job increases (Hunter, 1980). Research has also shown that managers and executives are more variable in cognitive ability than conventional wisdom would suggest (Ones & Dilchert, 2009). Test scores can predict for higher level jobs.

Formulas Become Obsolete

A final assumption to consider is the idea that formulas are static and inflexible and thus are not useful for making predictions about performance in the chaotic environments of the marketplace. According to Prien et al. (2003), "Economic conditions and circumstances and the nature of client businesses might be evolving, dynamic and in flux, changing so that any particular algorithm, no matter how carefully developed, could be obsolete" (p. 128). The problem with this argument is that assessors are somehow assumed to be more flexible and attuned to subtle changes in effectiveness criteria. In fact, assessors are likely to rely on implicit theories developed from past training and experience. Moreover, these implicit theories have likely become resistant to change as a result of positive illusions and hindsight biases (Fisher, 2008). Formulas may be updated on the basis of new information and empirical research.

FINAL THOUGHTS

As noted by Hogarth (1987), people's intuitive judgments are based on information processed and transformed by their minds. Hogarth noted that there are four ways in which judgments may derail (see Table 31.3): (a) selective perception of information, (b) imperfect information processing, (c) limited capacity, and (d) biased reconstruction of events.

Although humans have limited resources to make judgments, they paradoxically cope with this by adding more complexity to the problem. For example, people often create elaborate stories to make sense of disparate pieces of information, even when the stories themselves are too elaborate to be predictive (Gilovich, Griffin, & Kahneman, 2002; Pennington & Hastie, 1988). This chapter has shown that assessors are not immune to the limitations of human judgment. Indeed, assessment experience may only serve to exacerbate issues such as professional biases and overconfidence (Sieck & Arkes, 2005).

One benefit of the holistic school of thought is that it encourages people to look more broadly at the predictors and the criteria: to consider what the person brings to the educational or work environment as a whole. This broader perspective may encourage one to more thoroughly examine noncognitive attributes of the candidate, along with nontask attributes of job performance. The research does not, however, support the use of a holistic approach to data integration. Assessors are still needed to select the data on which the formulas are based and to assign ratings to the data points that are subjective in nature

TABLE 31.3

Four Ways That Assessor Judgments May Derail

Derailer	Explanation
Selective perception	Assessors are bombarded with information and must selectively choose which things deserve attention. People often see what they expect to see on the basis of initial impressions, background information, or professional biases.
Imperfect processing	Assessors cannot simultaneously integrate even the information on which they choose to focus. The sequence in which a person processes information may bias the judgment.
Limited capacity	Assessors use simple heuristics or rules of thumb to reduce mental effort. Comparing a candidate with oneself, or with people who have already succeeded in similar jobs, are strategies commonly used to simplify assessment.
Biased reconstruction	Memory is formed by assembling fragments of information. Vivid information is more easily recalled, even when it is not representative of the set as a whole. Also, information that supports one's initial impression is more easily recalled than information contrary to it.

(e.g., interpersonal warmth in the interview). If assessors heed the advice of Freyd (1926) by making subjective impressions objective, assessment will move out of the realm of philosophy, technique, and artistry and into the realm of science.

References

Ach, N. (2006). *On volition* (T. Herz, Trans.). Leipzig, Germany: Verlag von Quelle & Meyer. (Original work published 1910) Retrieved from http://www.psychologie.uni-konstanz.de/forschung/kognitive-psychologie/various/narziss-ach

Alexakos, C. E. (1966). Predictive efficiency of two multivariate statistical techniques in comparison with clinical predictions. *Journal of Educational Psychology, 57*, 297–306.

Albrecht, P. A., Glaser, E. M., & Marks, J. (1964). Validation of a multiple-assessment procedure for managerial personnel. *Journal of Applied Psychology, 48*, 351–360.

Allman, M. (August, 2009). *Quintessential questions: Wake Forest's admission director gives insight into the interview process*. Retrieved from http://rethinkingadmissions.blogs.wfu.edu/2009/08

Allport, G. W. (1962). The general and the unique in psychological science. *Journal of Personality, 30*, 405–422. doi:10.1111/j.1467-6494.1962.tb02313.x

Anderson, J. W. (1992). The life of Henry A. Murray: 1893–1988. In R. A. Zucker, A. I. Rabin, J. Aronoff, & S. J. Frank (Eds.), *Personality structure in the life course: Essays on personology in the Murray tradition* (pp. 304–334). New York, NY: Springer.

Ansbacher, H. L. (1941). German military psychology. *Psychological Bulletin, 38*, 370–392. doi:10.1037/h0056263

Arthur, W., Bell, S. T., Villado, A. J., & Doverspike, D. (2006). The use of person–organization fit in employment decision making: An assessment of its criterion-related validity. *Journal of Applied Psychology, 91*, 786–801. doi:10.1037/0021-9010.91.4.786

Barrick, M. R., & Mount, M. K. (1993). Autonomy as a moderator of the relationships between the Big Five personality dimensions and job performance. *Journal of Applied Psychology, 78*, 111–118. doi:10.1037/0021-9010.78.1.111

Bentz, V. J. (1967). The Sears experience in the investigation, description, and prediction of executive behavior. In F. R. Wickert & D. E. McFarland (Eds.), *Measuring executive effectiveness* (pp. 147–205). New York, NY: Appleton-Century-Crofts.

Bion, W. R. (1959). *Experiences in groups: And other papers*. London, England: Tavistock.

Borman, W. C. (1982). Validity of behavioral assessment for predicting military recruiter performance. *Journal of Applied Psychology, 67*, 3–9.

Borman, W. C., Eaton, N. K., Bryan, J. D., & Rosse, R. L. (1983). Validity of Army recruiter behavioral assessment: Does the assessor make a difference? *Journal of Applied Psychology, 68*, 415–419. doi:10.1037/0021-9010.68.3.415

Bray, D. W. (1964). The management progress study. *American Psychologist, 19*, 419–420. doi:10.1037/h0039579

Brody, W., & Powell, N. J. (1947). A new approach to oral testing. *Educational and Psychological Measurement, 7*, 289–298. doi:10.1177/001316444700700206

Burt, C. (1942). Psychology in war: The military work of American and German psychologists. *Occupational Psychology, 16*, 95–110.

Cabrera, A. F., & Burkum, K. R. (2001, November). Criterios en la admisión de estudiantes universitarios en los EEUU: Balance del sistema de acceso a la universidad (selectividad y modelos alternativos) [College admission criteria in the United States: Assessment of the system of access to the university (selectivity and alternative models)]. Madrid, Spain: Universidad Politécnica de Madrid.

Camerer, C. F., & Johnson, E. J. (1991). The process–performance paradox in expert judgment: How can experts know so much and predict so badly? In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 195–217). Cambridge, England: Cambridge University Press.

Capshew, J. H. (1999). *Psychologists on the march: Science, practice, and professional identity in America, 1929–1969*. New York NY: Cambridge University Press. doi:10.1017/CBO9780511572944

Conway, J. M., Jako, R. A., & Goodman, D. F. (1995). A meta-analysis of interrater and internal consistency reliability of selection interviews. *Journal of Applied Psychology, 80*, 565–579. doi:10.1037/0021-9010.80.5.565

Cortina, J. M., Goldstein, N. B., Payne, S. C., Davison, H. K., & Gilliland, S. W. (2000). The incremental validity of interview scores over and above cognitive ability and conscientiousness scores. *Personnel Psychology, 53*, 325–351. doi:10.1111/j.1744-6570.2000.tb00204.x

Dawes, R. M. (1971). A case study of graduate admissions: Application of three principles of human decision making. *American Psychologist, 26*, 180–188. doi:10.1037/h0030868

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist, 34*, 571–582. doi:10.1037/0003-066X.34.7.571

DuBois, P. H. (1970). *A history of psychological testing.* Boston, MA: Allyn & Bacon.

Eysenck, H. J. (1953). *Uses and abuses of psychology*. Harmondsworth, England: Penguin Books.

Feltham, R. (1988). Assessment centre decision making: judgemental vs. mechanical. *Journal of Occupational Psychology, 61*, 237–241.

Fisher, C. D. (2008). Why don't they learn? *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 364–366. doi:10.1111/j.1754-9434.2008.00065.x

Frazer, J. M. (1947). New-type selection boards in industry. *Occupational Psychology, 21*, 170–178.

Freyd, M. (1925). The statistical viewpoint in vocational selection. *Journal of Applied Psychology, 9*, 349–356.

Ganzach, Y., Kluger, A. N., & Klayman, N. (2000). Making decisions from an interview: Expert measurement and mechanical combination. *Personnel Psychology, 53*, 1–20. doi:10.1111/j.1744-6570.2000.tb00191.x

Geuter, U. (1992). *The professionalization of psychology in Nazi Germany* (R. J. Holmes, Trans.). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511666872

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment.* New York, NY: Cambridge University Press.

Gratz v. Bollinger, 539 U.S. 244 (2003).

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. *Psychology, Public Policy, and Law, 2*, 293–323. doi:10.1037/1076-8971.2.2.293

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction. *Psychological Assessment, 12*, 19–30. doi:10.1037/1040-3590.12.1.19

Harrell, T. W., & Churchill, R. D. (1941). The classification of military personnel. *Psychological Bulletin, 38*, 331–353. doi:10.1037/h0057675

Hess, T. G. (1977). Actuarial prediction of performance in a six-year AB-MD program. *Journal of Medical Education, 52*, 68–69.

Highhouse, S. (2002). Assessing the candidate as a whole: A historical and critical analysis of individual psychological assessment for personnel decision making. *Personnel Psychology, 55*, 363–396. doi:10.1111/j.1744-6570.2002.tb00114.x

Highhouse, S. (2008). Stubborn reliance on intuition and subjectivity in employee selection. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 333–342. doi:10.1111/j.1754-9434.2008.00058.x

Hogarth, R. M. (1987). *Judgment and choice: The psychology of decision.* New York, NY: Wiley.

Hollenbeck, G. P. (2009). Executive selection—What's right . . . and what's wrong. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 2*, 130–143. doi:10.1111/j.1754-9434.2009.01122.x

Hoover, E., & Supiano, B. (2010, May). Admissions interviews: Still an art and a science. Retrieved from http://chronicle.com/article/The-Enduring-Mystery-of-the/65545

Howard, A. (1974). An assessment of assessment centers. *Academy of Management Journal, 17*, 115–134. doi:10.2307/254776

Hunter, J. E. (1980). *Validity generalization for 12,000 jobs: An application of synthetic validity and validity*

generalization to the General Aptitude Test Battery (GATB). Washington, DC: U.S. Employment Service.

Huse, E. G. (1962). Assessments of higher-level personnel: The validity of assessment techniques based on systematically varied information. *Personnel Psychology, 15*, 195–205. doi:10.1111/j.1744-6570.1962.tb01861.x

Jeanneret, R., & Silzer, R. (1998). An overview of psychological assessment. In R. Jeanneret & R. Silzer (Eds.), *Individual psychological assessment: Predicting behavior in organizational settings* (pp. 3–26). San Francisco, CA: Jossey-Bass.

Korchin, S. J., & Schuldberg, D. (1981). The future of clinical assessment. *American Psychologist, 36*, 1147–1158. doi:10.1037/0003-066X.36.10.1147

Korman, A. K. (1968). The prediction of managerial performance: A review. *Personnel Psychology, 21*, 295–322. doi:10.1111/j.1744-6570.1968.tb02032.x

McDaniel, M. A., Whetzell, D. L., Schmidt, F. L., & Maurer, S. D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology, 79*, 599–616. doi:10.1037/0021-9010.79.4.599

McGaghie, W. C., & Kreiter, C. D. (2005). Special article: Holistic versus actuarial student selection. *Teaching and Learning in Medicine, 17*, 89–91. doi:10.1207/s15328015tlm1701_16

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence.* Minneapolis: University of Minnesota Press. doi:10.1037/11281-000

Meyer, H. H. (1956). An evaluation of a supervisory selection program. *Personnel Psychology, 9*, 499–513. doi:10.1111/j.1744-6570.1956.tb01082.x

Miner, J. B. (1970). Executive and personnel interviews as predictors of consulting success. *Personnel Psychology, 23*, 521–538. doi:10.1111/j.1744-6570.1970.tb01370.x

Mitchel, J. O. (1975). Assessment center validity: A longitudinal study. *Journal of Applied Psychology, 60*, 573–579.

Murray, H. (1990). The transformation of selection procedures: The War Office Selection Boards. In E. Trist & H. Murray (Eds.), *The social engagement of social science: A Tavistock anthology* (pp. 45–67). Philadelphia: University of Pennsylvania Press.

Office of Strategic Services. (1948). *Assessment of men.* New York, NY: Rinehart.

Older, H. J. (1948). [Review of the book *Assessment of men: Selection of personnel for the Office of Strategic Services*]. *Personnel Psychology, 1*, 386–391.

Ones, D. S., & Dilchert, S. (2009). How special are executives? How special should executive selection be? Observations and recommendations. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 2*, 163–170. doi:10.1111/j.1754-9434.2009.01127.x

Pennington, N., & Hastie, R. (1988). Explanation-based decision making: Effects of memory structure on judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 521–533. doi:10.1037/0278-7393.14.3.521

Prien, E. P., Schippmann, J. S., & Prien, K. O. (2003). *Individual assessment: As practiced in industry and consulting.* Mahwah, NJ: Erlbaum.

Pulakos, E. D., Schmitt, N., Whitney, D., & Smith, M. (1996). Individual differences in interviewer ratings: The impact of standardization, consensus discussion, and sampling error on the validity of a structured interview. *Personnel Psychology, 49*, 85–102. doi:10.1111/j.1744-6570.1996.tb01792.x

Pynes, J., Bernardin, H. J., Benton, A. L., & McEvoy, G. M. (1988). Should assessment center dimension ratings be mechanically-derived? *Journal of Business and Psychology, 2*, 217–227. doi:10.1007/BF01014039

Roose, J. E., & Doherty, M. E. (1976). Judgment theory applied to the selection of life insurance salesmen. *Organizational Behavior and Human Decision Processes, 16*, 231–249.

Rosen, N. A., & Van Horn, J. W. (1961). Selection of college scholarship students: Statistical vs clinical methods. *Personnel Guidance Journal, 40*, 150–154.

Ruscio, J. (2003). Holistic judgment in clinical practice: Utility or futility? *Scientific Review of Mental Health Practice, 2*, 38–48.

Ryan, A. M., & Sackett, P. R. (1992). Relationships between graduate training, professional affiliation, and individual psychological assessment practices for personnel decisions. *Personnel Psychology, 45*, 363–387. doi:10.1111/j.1744-6570.1992.tb00854.x

Sackett, P. R., & Wilson, M. A. (1982). Factors affecting the consensus judgment process in managerial assessment centers. *Journal of Applied Psychology, 67*, 10–17.

Sarbin, T. L. (1942). A contribution to the study of actuarial and individual methods of prediction. *American Journal of Sociology, 48*, 593–602. doi:10.1086/219248

Schofield, W. (1970). A modified actuarial method in the selection of medical students. *Journal of Medical Education, 45*, 740–744.

Scott, W. D., & Clothier, R. C. (1923). *Personnel management.* Chicago, IL: A. W. Shaw.

Sieck, W. R., & Arkes, H. R. (2005). The recalcitrance of overconfidence and its contribution to decision aid neglect. *Journal of Behavioral Decision Making, 18*, 29–53. doi:10.1002/bdm.486

Silzer, R., & Jeanneret, R. (1998). Anticipating the future: Assessment strategies for tomorrow. In R. Jeanneret & R. Silzer (Eds.), *Individual psychological*

assessment: Predicting behavior in organizational settings (pp. 445–477). San Francisco, CA: Jossey-Bass.

Sparks, C. P. (1990). Testing for management potential. In K. E. Clark & M. B. Clark (Eds.), *Measures of leadership* (pp. 103–112). West Orange, NJ: Leadership Library of America.

Stagner, R. (1957). Some problems in contemporary industrial psychology. *Bulletin of the Menninger Clinic, 21*, 238–247.

Sutherland, J. D., & Fitzpatrick, G. A. (1945). Some approaches to group problems in the British Army. *Sociometry*, *8*, 205–217. doi:10.2307/2785043

Taft, R. (1948). Use of the "group situation observation" method in the selection of trainee executives. *Journal of Applied Psychology, 32*, 587–594. doi:10.1037/h0061967

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

Trankell, A. (1959). The psychologist as an instrument of prediction. *Journal of Applied Psychology, 43*, 170–175.

Trist, E. (2000). Working with Bion in the 1940s: The group decade. In M. Pines (Ed.), *Bion and group psychotherapy* (pp. 1–46). London, England: Jessica Kingsley.

Tziner, A., & Dolan, S. (1982). Validity of an assessment center for identifying future female officers in the military. *Journal of Applied Psychology, 67*, 728–736.

Viteles, M. S. (1925). The clinical viewpoint in vocational selection. *Journal of Applied Psychology, 9*, 131–138. doi:10.1037/h0071305

Watley, D. J., & Vance, F. L. (1964). *Clinical versus actuarial prediction of college achievement and leadership activity*. Minneapolis: University of Minnesota Student Counseling Bureau.

Wickert, F. R., & McFarland, D. E. (1967). *Measuring executive effectiveness.* New York, NY: Appleton-Century-Crofts.

Wollowick, H. B., & McNamara, W. J. (1969). Relationship of the components of an assessment center to management success. *Journal of Applied Psychology, 53*, 348–352.