Industrial and Organizational Psychology, **4** (2011), 270–296.

FOCAL ARTICLE

Individual Psychological Assessment: A Practice and Science in Search of Common Ground

ROB SILZER
HR Assessment & Development, Inc. and
Baruch College, City University of New York

RICHARD JEANNERET
Retired I–O Psychologist

Abstract

During the past 30 years, individual psychological assessment (IPA) has gained in use and in value to organizations in the management of human resources. However, even though IPA is considered a core competency for industrial–organizational (I–O) psychology, its practice is not without critics. This article is written not only to address several criticisms of IPA but also to discuss a variety of issues that must be taken into consideration if IPA is to advance as a major component of the I–O scientist–practitioner model. We rely upon a working definition of IPA in general but, when possible, focus on executive assessment in particular, given its high level of complexity and growing popularity. We discuss the effectiveness of assessment practice, including the ongoing statistical versus clinical prediction argument and the difficulties with establishing validity. Although we are confident that IPA has many strong research and practice underpinnings, we also propose some important research questions, training guidelines, and opportunities for assessing psychologists to improve their practices.

Individual psychological assessment (IPA[1]) has been part of the field of industrial–organizational (I–O) psychology for many years. It has gained in popularity over the past 30 years as assessment has become a central function in human resource departments and as I–O psychologists have gained influence and credibility in organizations (McPhail & Jeanneret, in press; Prien, Schippmann, & Prien, 2003; Silzer & Jeanneret, 1998). Howard (1990) estimated that approximately one-third of Society for Industrial and Organizational Psychology (SIOP) members conduct individual assessments. We suspect that over the last 20 years the number of I–O psychologists conducting individual assessments has noticeably increased. IPA is now routinely offered by consulting firms and independent consultants as part of assessment services.

However, there have been criticisms of individual assessment. Some have provided useful critical reviews (Ryan & Sackett, 1987, 1998), whereas others have

Correspondence concerning this article should be addressed to Rob Silzer,
E-mail: robsilzer@prodigy.net
Address: HR Assessment and Development Inc., 14 West 11th St., New York, NY 10011

We would like to recognize the outstanding literature search contribution of Mr. Chad Parson.

1. We will use the terms ''individual psychological assessment,'' ''individual assessment,'' and ''IPA'' interchangeably.

https://doi.org/10.1111/j.1754-9434.2011.01341.x Published online by Cambridge University Press

questioned the reliance on clinical judgment and integration in assessments (Highhouse, 2002; Kuncel, Klieger, Connelly, & Ones, 2008). Guion (1998) questioned the job relatedness of the assessment constructs, assessor reliability, the overdependence on a few primary indicators, unsupported job relevance for personality dimensions, inadequate measurement of interpersonal skills, and potential privacy issues; but noted that all of his criticisms could be addressed by proper development and design of the assessment procedures. Some of these criticisms diminish the central role of psychology and psychological judgment in conducting psychological assessments. We think that some of these arguments are misplaced.

We believe that we are in fact *psychologists* and that psychological judgment is critical to IPA, particularly at the senior management and executive level. Individual psychological assessment is a core competency in industrial–organizational psychology (SIOP, 1999). It is recognized that to be in compliance with state licensure statutes individuals who identify themselves as psychologists conducting individual assessments must be licensed. In response, some in our profession argue against requiring I–O psychologists to be licensed as psychologists (Campbell, Levy, Murphy, Schmitt, & Weiss, 2009; Locke, 2009). However, we believe that I–O psychologists conducting assessments need to maintain a legal identity as psychologists and be licensed (see Jeanneret, 1998, 2009).

Our focus here is to discuss why psychological judgment is central to management and executive-level assessments and to address some concerns that have been raised. We will discuss the definition of IPA, what benefit it brings, why it is a complex process, and what recent research has found. We will also provide some suggestions for further research, for training assessment psychologists, and for improving assessment practice. We offer a perspective based on our experience of conducting individual assessments for the past 30 years and a review of the relevant literature. Our focus when possible is on management and executive-level assessments, which we find to be the most complex and challenging individual assessment practices.

What Is Individual Psychological Assessment?

Over 10 years ago we proposed a definition of individual psychological assessment that we think is still valid: ''a process of measuring a person's knowledge, skills, abilities, and personal style to evaluate the characteristics and behavior that are relevant to (predictive of) successful job performance'' (Jeanneret & Silzer, 1998b, p. 3).

This definition allows for some variance in practice rather than requiring a rigid adherence to a strict protocol. A more inclusive definition is appropriate to allow for variations that might improve the effectiveness of the assessment process. However, we have made clear distinctions between IPA and other types of evaluation in organizations, such as performance appraisals, multirater feedback surveys, and group selection testing (Jeanneret & Silzer, 1998b). Individual assessment ''brings a type of reliable psychological measurement, a degree of objectivity, and an understanding of underlying psychological constructs to the process that cannot be achieved by organizational evaluations'' (Jeanneret & Silzer, 1998b, p. 5).

We see value in identifying a few core characteristics of individual assessment. One core characteristic is defining ''individual'' as the focus on assessing ''one person.'' This would exclude assessments of a group of participants as a whole or a large group selection process. However, we do not think this should be extended to *requiring* ''singleness'' in other assessment features (i.e., single assessor, single integrator, or individually administered tools). A second core characteristic is direct contact between the assessor psychologist and the individual, usually through an interview. This contact allows the assessor to make direct observations about the individual's behavior in

https://doi.org/10.1111/j.1754-9434.2011.01341.x Published online by Cambridge University Press

a one-on-one context and further distinguishes individual assessment from other assessment methods. A third characteristic is the use of multiple assessment methods when possible to measure psychological constructs and work-related behavior. Admittedly, there are some special situations, such as assessment of individuals in some cultures, where a long structured interview might be the sole assessment tool.

Others have used different definitions that more narrowly define IPA. For example, Ryan and Sackett (1998) found it difficult to identify *typical* assessment practices and defined IPA as ''one psychologist making an assessment decision for a personnel-related purpose about one individual'' (p. 50). More recently Roller and Morris (2008b) eliminated studies from their IPA meta-analysis that included a group exercise in the assessment process. We think these are unnecessary restrictions. Specifically, we believe that individual assessment should:

- *Not be restricted by the type of assessment methods used*. In some situations role-play exercises, 360-feedback data, task-force exercises, or even group discussions can be useful methods for collecting relevant assessment data on an individual. However, there should be at least one assessment method that allows direct contact with the individual (such as an interview or a simulation exercise).
- *Not be restricted by involvement of only one psychologist (and no more) in the process*. Although often there may be only one assessor psychologist involved, additional exercise assessors, test evaluators, and other assessors can add to the assessment effectiveness and accuracy by participating in assessment data collection and decisions. For example, including an integration discussion to reach final judgments might be a useful addition.
- *Not be restricted by data integration only by a pure clinical judgment integration process*. Certain data such as personality dimension scores are now derived by statistical algorithms developed by the test creators, and in some instances several dimension scores are statistically linked to identify patterns or broad constructs or to support overall interpretations. A variety of integration methods should be explored to find the most useful integration approaches, but this does not detract from our belief in the core role of psychological judgment in the interpretation and integration of assessment data.

It is useful to distinguish individual assessment from two other related assessment methods: test battery assessments and assessment centers. Test battery assessments (TBAs) rely solely on data collected through psychological tests and inventories, and do not typically include direct observations of the individual through interviews, simulations, or other assessment methods (although some online TBAs may now include automated interviews and simulations that permit indirect observations of the assessees). The test battery is usually based on a job analysis and a validation study, and the process is typically focused on screening large numbers of candidates for either one job or a very small number of jobs, often at lower levels in the organization. In our view, TBAs that are missing direct interaction with the individual should be considered a distinct assessment method and should not be considered as representative of individual assessments in general (see Roller & Morris, 2008a, b).

Assessment centers are also a related but different assessment method when contrasted with IPA. Both assessment methods measure similar constructs in evaluating behaviors that are relevant to successful job performance. They often have the similar selection and development objectives and are often seen as equally effective by providers and clients, although there are usually differences in implementation logistics, costs, and staff requirements. There are some obvious method differences such as the required and dominant

https://doi.org/10.1111/j.1754-9434.2011.01341.x Published online by Cambridge University Press

use of simulations and multiple assessors (not necessarily psychologists) in assessment centers, whereas individual assessments often include a different, and often broader, mix of assessment tools. However, we do believe that knowledge about assessment centers and related research can provide useful insights and guidance on how individual psychological assessments can be further improved. In fact, individual assessment can benefit by leveraging some of the advantages of assessment centers and designing ''assessment centers for one.''

Why Does Individual Assessment Survive?

Assessment practices have existed for centuries; however, individual psychological assessment became widely used in the 20th century, particularly in the 1970s, as business organizations recognized the added value of IPA in helping selected individuals into jobs (Jeanneret & Silzer, 1998a; Prien et al., 2003; Silzer, 1984). The practice has survived and even thrives because it provides useful benefits to organizations, individual assessees, and assessor psychologists.

Organizations find significant value in individual assessment because it (a) provides an objective, external view of an individual's knowledge, skills, abilities (KSAs), and personal style that are related to successful job performance; (b) evaluates psychological constructs and behavior that are difficult to measure by other methods; (c) is based on sound psychological principles, theory, and research; (d) provides professional insight into the individual's future job performance, potential for higher-level positions, and potential to be successful in changing organizational demands; (e) gauges the individual's fit with the immediate manager, the peer management team, the existing direct reports and organizational structure, the organizational culture, the company values, and the country culture; (f) provides clear and objective comparisons between candidates for the same position; (g) identifies potential performance risks in hiring (or promoting) an individual and helps to avoid significant performance failures in high-level positions; (h) can be adapted to changing organizational contextual variables and job demands; and (i) can provide a road map for effectively managing, developing, and coaching the individual in the future.

These are not trivial benefits and many organizations and hiring mangers now rely on individual assessment as a reliable source of candidate information, particularly for external candidates. Given the cost of hiring and then turning over a poor performer, many organizations see individual assessment as a sound and useful investment. Russell (2001) estimated that an executive assessment can generate an additional $3 million in annual profit per candidate selected. Veneziano (2010) reported that clients found individual assessments ''insightful and accurate'' and that they added ''significant value to the selection process, particularly at higher organizational levels'' (slide #19). Assessment can also meet other needs by quickly and efficiently providing valuable information on individuals to help organizations meet changing demands for talent (Cooper-Thomas & Trayes, 2008; Klimoski & Zukin, 2003). Our experience is that once organizations start using individual assessments, there is often a pattern of continued use over time. Organizations come to rely on assessment information and consider it valuable in supporting sound business decisions.

Individual assessees benefit from individual assessment because it (a) provides an objective evaluation of their KSA's and personal characteristics, independent of organizational politics and manager biases; (b) is conducted by a professional assessment psychologist; (c) includes assessment tools and methods that have sound psychometric properties; and (d) provides professional feedback on their match to the position and the organization (and possibly to a global assignment). Our experience is that individuals typically give high ratings to their individual assessment experiences and

https://doi.org/10.1111/j.1754-9434.2011.01341.x Published online by Cambridge University Press

appreciate the valuable feedback and the professional treatment.

Assessment psychologists benefit from conducting individual assessments because it (a) is based on sound psychological principles and research, (b) allows them to use the full range of their psychological knowledge and professional skills, (c) gives them an opportunity to fully understand an individual on key psychological constructs and job requirements, (d) helps them to understand an organizational culture and what is required to be effective in that environment, (e) helps them to build a base of assessment psychology knowledge to better understand behavioral patterns and how behavior relates to underlying constructs, (f) provides them with an ongoing sustained client relationship that allows them to increase their contributions over time, (g) helps them to build a professional assessment practice and offer psychological services to clients, and (h) guides them in training new assessment psychologists.

Given the reasons above, we believe that individual psychological assessment is alive and well and provides clear and meaningful value to organizations, individual assessees, and assessing psychologists. We do, however, have reservations about how the process is designed and used in some situations and by some assessors. For example, some assessors are poorly trained or follow assessment practices that are not consistent with professional standards, ethical codes, or psychological principles; use questionable assessment tools or methods that have little or no supporting construct validity; do not appropriately control the confidential assessment information that is collected; and have a limited understanding of work psychology and organizations. Some assessments are poorly administered, and assessment data may be misused. Moreover, we have reservations about executive-search consulting firms using individual assessments to advocate for and promote their candidates to client organizations. Therefore, we support the establishment of professional guidelines and the certification or licensure of assessment psychologists. The development of the International Organization for Standardization standards currently underway (ISO, 2010) is a valuable undertaking in this regard.

What Makes Management and Executive Individual Assessments So Complex?

Over the past 30 years each of us has completed thousands of management and executive-level individual assessments for many client organizations. We appreciate the complexity related to the assessment context, identifying an individual's assessment profile, and predicting future performance (Day, 2009; Hollenbeck, 2009; Kristof-Brown, Zimmerman, & Johnson, 2005; Silzer, 2002a, b; Silzer & Church, 2009; Silzer & Davis, 2010; Silzer & Dowell, 2010). It is far more complex than just trying to predict college grade point averages for entering college students (Kuncel et al., 2008); the two situations should not simplistically be seen as comparable.

At each step in the assessment process, psychologists face complex choices and decisions. When specifically addressing executive and management assessment, the days are long gone when the job requirements were clearly defined, the candidates were easy to profile, and the selection decision was straightforward. Today the position requirements and organizational expectations placed on an individual are more numerous and often more ambiguous, the candidates are more experienced but also more assessment savvy, and the selection decisions are more complicated and carry much higher risk to the organization.

Assessors often face a unique set of circumstances for each senior-level assessment, including constantly changing position descriptions and performance expectations, increasing latitude that each incumbent has to structure and shape their own work, broader management authority to change the focus and expectations of the position (often on short notice), decreasing likelihood that the more senior manager in place has held the position to be filled or

https://doi.org/10.1111/j.1754-9434.2011.01341.x Published online by Cambridge University Press

has been in the organization very long, frequently changing organizational strategies and goals, and the growing impact of the local country culture on the performance expectations. There is often great pressure on new senior managers to quickly show their effectiveness and do something significantly different from the last incumbent. We clearly disagree with those who have a simplistic view of executive roles (Ones & Dilchert, 2009).

It can be challenging to collect enough job information to identify key performance dimensions and person requirements. In addition, there are organizational pressures and influences from many directions that impact performance, such as expectations from the immediate and even higher-level managers (often reflecting the manager's own personal biases or changing work goals); from human resources on what is now needed in the position (frequently quite different from what was needed in the past); that the individual will avoid the specific performance issues of the last incumbent (a repeat of similar issues will be seen as a selection failure); from the organization about expected performance related to future career moves; from the organization that the individual will fit specific cultural norms, values, and success profiles of the particular country and geographical culture (Dunnette, 1998, attributed the success of the clinical judgments in the AT&T assessment studies to the experts' extensive knowledge of the determinants of success at that organization); from the organization related to expected changes in the organizational culture, values, or strategies; and on whether the individual will be expected to significantly change, turn around, maintain, or grow the function or work group.

One of the most challenging assessor tasks is to make an overall recommendation if the assessment is being completed as part of a selection process. It requires that an assessor simultaneously consider a number of critical issues, such as:

- What are the performance dimensions most related to performance effectiveness in this position?
- What are the contextual factors and influences that may impact how well this individual or any individual can perform in this role?
- How valid is my assessment of these key dimensions and factors and how accurate is the individual's assessment profile?
- What are high-risk issues either in the individual or in the organizational context that I need to carefully consider?
- How effective can I be in predicting future performance in a changing organization where many variables are unknowable?
- Can I thoughtfully balance all these factors and arrive at a single recommendation that fairly and accurately reflects the individual's skill and abilities and the organization's needs?
- How will the organization use the recommendation, and how much weight does it give the assessment recommendation?

Some assessors try to avoid giving a recommendation (Guion, 1998) and see their primary responsibility as description of the individual and not prediction of effectiveness in a future work environment. Ryan and Sackett (1987) found that approximately 35% of the assessors in their sample provide dimension ratings, and approximately 60% provide a specific recommendation. There are many difficulties in providing an overall assessment rating or recommendation (Arthur, Day, & Woehr, 2008; Arthur, Woehr, & Maldegen, 2000; Lance, 2008; Stratton, 1991). Assessment psychologists need to consider how the recommendation will be used, what factors and conditions could affect later performance, and what consequences could occur if the recommendation is used to make comparative judgments across individuals. Often later performance failures at the executive level can be spectacular

https://doi.org/10.1111/j.1754-9434.2011.01341.x Published online by Cambridge University Press

events, and most assessors understand that significant risks are involved. The key predictors are likely to be somewhat different for each organization and when predicting short- versus long-term effectiveness (Russell, 2001; Silzer, 2010). It is critical to understand how assessment fits into a particular organization (Guion, 1998; Muchinsky, 2004).

How Effective Is Individual Psychological Assessment?

As professionals, I–O psychologists are interested in how research provides evidence for the effectiveness of assessment practice. We consider three assessment components: What skills and limitations do the assessment psychologists bring to the process?, How accurate are assessor integration and judgment approaches?, and What is the validity evidence for individual assessment?

Assessor Skills

Clinicians for sometime have been seen as more effective collectors and interpreters of psychological information than mechanical methods (Holtzman, 1960; Westen & Weinberger, 2004) but less effective at combining and integrating large data sets. We think Holtzman's view is still valid:

> As a collector of information the computer has little to offer that can not be done better by the clinician or his assistant. As a processor of information, machines greatly surpass the human brain once the primary coding of information has been done by the clinician or his assistant. As an interpreter of information, once again the clinician has a definite edge over the computer. (Holtzman, 1960, p. 122)

Effective assessors have ''collection and interpretation'' skills that enable them to bring unique data and insights to the assessment process. Assessors also have other useful skills.

First, they are accurate observers of behavior (Holt, 1975; Jeanneret & Silzer, 1998b; Westen & Weinberger, 2004). Through natural and trained skills, assessors can see and hear behavior in their observations of an individual that can provide useful and sometimes critical pieces of information to rating the individual on key dimensions. An example is an assessor's skills in observing an individual's interpersonal and communication behavior in an assessment interview. It requires well-developed visual observation, listening skills, and interpretative capabilities to capture the extensive behavioral data that individuals usually demonstrate in interviews.

Assessors can also formulate and test hypotheses about the individual. Using an analytical approach, they can probe and collect additional information relevant to a concern or a dimension. They look for behavioral patterns and probe for additional information that either supports or conflicts with a hypothesis about the individual's related behavior. This flexibility allows the assessor to pursue behavioral leads in an interview. They can adapt to new information and changing dynamics in the interview. Meehl (1954) concluded that the clinician's ''genuine creative act'' of generating a ''structural-dynamic hypothesis'' (p. 57) is the distinguishing feature between statistical and clinical prediction.

Assessors can understand specific behavioral data points while also seeing larger behavioral patterns and psychological constructs. Particularly in the assessment interview, capable assessors can synthesize the interview information at both levels (data and construct) simultaneously and understand the relationships between the two levels. Similarly, assessors can complete both normative and ipsative interpretations for the same variables for the same assessee that leads to a fuller understanding of that individual (Jeanneret & Silzer, 1998b; McPhail & Jeanneret, in press; Meyer, 1998). This is a process that would be virtually impossible to complete in some mechanical or statistical manner.

https://doi.org/10.1111/j.1754-9434.2011.01341.x Published online by Cambridge University Press

Assessors can accurately sort behavior into key performance-related dimensions. We disagree with Highhouse (2002) who criticized individual assessment for taking a ''holistic approach.'' Most assessors conscientiously, and we would argue systematically, focus on collecting and assessing information specifically related to important performance dimensions and constructs or contextual and organizational variables. We think most assessors are able to arrive at a coherent understanding of the person based on the individual's assessment profile across the critical assessment dimensions (Jeanneret & Silzer, 1998b; McPhail & Jeanneret, in press; Prien et al., 2003). For most assessors, this is based on dimension evaluations rather than from a clinical whole person approach, although there are some I–O psychologists who reject a dimension or competency approach (see Hollenbeck, McCall, & Silzer, 2006; Silzer & Hollenbeck, 1992).

Assessors can integrate information and accurately rate an individual on specific performance dimensions. Silzer found that assessors were as accurate as mechanical integration methods (and in some cases more accurate) when predicting specific performance dimensions (Silzer, 1984). However, several researchers have noted that human judgment in combination with some mechanical processes can improve prediction accuracy in some situations (Ganzach, Kluger, & Klayman, 2000; Silzer, 1984; Whitecotton, Sanders, & Norris, 1998; Wunder, Thomas, & Luo, 2010). Prien, Schippmann, and Prien (2003) contend that configural interpretation is an effective way to understand the relationships between personal characteristics, competencies, and contextual variables.

Finally, assessors can consider a range of behavior and determine how relevant the behavior is to later performance effectiveness. Meehl (1954) pointed out the need for data to be as clear as a ''broken leg'' to override a mechanical integration prediction. Often with senior managers and executives there are many pieces of assessment data that seem as clear as a ''broken leg,'' that indicate the likelihood of later performance issues that either appear fairly quickly (and result in turnover) or take longer to surface (and often lead to quiet moves to the side). For example, often there are a variety of factors related to fitting in with the organizational culture, the immediate manager or CEO, the peer team, or the country culture (see Silzer, 2002a, b, 2005 for discussions of executive fit). The list of potential areas that can lead to performance problems is extensive in some organizations, and some of our colleagues report a high level of executive failure (Hollenbeck, 2009). We suggest that many of the potentially problematic behaviors can be identified by an expert assessment psychologist and that the executive success rate over time is higher for companies that regularly use well-designed and implemented executive individual assessment.

Some assessors are more skilled at individual assessment than others. They have abilities not shared by other assessors, and there is evidence of individual differences among clinicians that affect judgment accuracy (Garner & Smith, 1976; Kaplan, 1975; Levenberg, 1975; Shanteau, 1988; Taft, 1955). Some assessment psychologists may just be better natural observers of behavior than others and have a deeper theoretical and conceptual understanding of behavioral patterns and constructs. Silzer (1984) found that assessors differed in their assessment validity and the best clinician achieved higher validity levels than the most accurate statistical integration methods. Years ago, Holt (1958, 1970) argued that clinicians ''vary considerably in their ability to do the job, but the best of them can do it very well'' (p. 348). We would encourage research that would identify those individual difference variables that effect assessment skills and accuracy. Others have found some variables that impact the ability to judge others (Ambady & Rosenthal, 1993; Taft, 1955). Early on, Taft identified the main attributes of the ability to judge others as possessing appropriate judgment norms,

https://doi.org/10.1111/j.1754-9434.2011.01341.x Published online by Cambridge University Press

judging ability, and motivation. Another consideration is whether individual differences affect the validity of assessment evaluations (Dougherty, Ebert, & Callender, 1986; Dreher, Ash, & Hancock, 1988; Shanteau, 1988).

Assessor Limitations

We also recognize that assessors can make judgment errors and have difficulty with particular assessment tasks and in specific situations (Jeanneret & Silzer, 1998b; Prien et al., 2003). For example, assessment psychologists are not especially effective at integrating large amounts of data that cover a broad spectrum of capabilities into a single overall rating or recommendation. Mechanical combination methods are often as accurate (and sometimes more accurate) in these situations (Dawes, 1979; Nystedt & Magnusson, 1972; Silzer, 1984).

Assessors may not be highly reliable in their judgments across assessments and may not consistently follow even their own integration and decision guidelines or rules (Karelaia & Hogarth, 2008; Ryan & Sackett, 1998). They can remain open to seeing new behavioral cues, often to avoid jumping to premature conclusions or relying on stereotypes about behavior patterns. As psychologists, we look for individual differences and are open to an individual's uniqueness (which may help or hinder later performance). But in doing so, we may overlook some basic core behavior patterns and predictors and overweigh more tangential behavioral data. This relates to the decision to use a structured or an unstructured assessment interview format. We suggest that both these choices do not work as well as a semi structured interview approach that follows a standard format and uses similar questions for individuals considered for similar positions, while also giving the assessor the flexibility to probe into issues or behavior that the individual introduces in the interview or to test working hypotheses about the individual during the interview. Garb (2005) found that the use of semi structured interviews can improve prediction accuracy.

Assessors are susceptible to the same judgment errors as others (Garb, 1989, 1996; Kahneman, Slovic, & Tversky, 1982). ''These include ignoring base rates, assigning nonoptimal weights to cues, failure to take into account regression toward the mean, and failure to properly assess covariation. Heuristics, such as representativeness (which leads to belief in the law of small numbers) or availability (leading to over-weighting vivid data), can similarly reduce clinicians' accuracy'' (Grove, Zald, Lebow, Snitz, & Nelson, 2000, p. 25). Others have pointed out additional judgment limitations, such as biases in retrospection, hindsight bias, and availability (Hastie & Dawes, 2001); biased hypothesis-testing strategies (Pfeiffer, Whelan, & Martin, 2000; Strohmer, Shivy, & Chiodo, 1990); and using stereotypes (Casas, Brady, & Ponterotto, 1983; Ridley, Li, & Hill, 1998; Wampold, Casas, & Atkinson, 1981; Wisch & Mahalik, 1999). Einhorn and Hogarth (1978) noted other judgment biases and errors that apply to assessment psychologists, including the unreliability of one's memory for events, ignoring or underweighting evidence that disconfirms a hypothesis, and the development of illusory correlations when a good fit between predicted outcomes and input information produces unwarranted confidence.

Some assessors may not learn from experience (Einhorn & Hogarth, 1981). They may settle into assessment routines and not seek out or receive adequate feedback on their judgment accuracy (Einhorn & Hogarth, 1978), so they have little opportunity to change bad judgment habits. However, it is unclear whether feedback actually improves judgment accuracy for clinicians (Garb, 1998, 2005; Karelaia & Hogarth, 2008; Meehl, 1954; Spengler et al., 2009; Westen & Weinberger, 2004). Schmitt (1978) suggests that outcome feedback can result ''in confusion which is reflected both in lower consistency and lack of insight or confidence into their weighting strategy'' (p. 186).

And finally, some assessors are not sufficiently systematic in their assessment

https://doi.org/10.1111/j.1754-9434.2011.01341.x Published online by Cambridge University Press

approach to insure that they collect data on the core assessment dimensions in consistent and systematic ways. Meehl pointed this out years ago and identified the need to take a more systematic and structured approach to interview data (Meehl, 1959). This consistency issue remains a challenge.

Clinical Integration and Judgment Accuracy

Over the past 50 years there has been considerable discussion about the effectiveness of clinical integration of data and the accuracy of clinical judgment. There is an extensive relevant literature on psychological judgment from the fields of clinical and counseling psychology (Ægisdóttir et al., 2006; Dawes, Faust, & Meehl, 1989; Einhorn, 1986; Garb, 1998; Garb & Grove, 2005; Meehl, 1954; Westen & Weinberger, 2004; Wiggins, 1973) and judgment and decision-making literature (Shanteau & Stewart, 1992; Smith, Shanteau, & Johnson, 2004). For many years clinicians were seen as very effective in diagnosing individuals and making predictions about future behavior (Allport G.W., 1937; Thorndike, 1918), although even years ago some characterized the clinical approach as intuitional and therefore unscientific (Allport F.H., 1937; Cattell, 1937; Sarbin, 1944). Some arguments never seem to change!

Meehl (1954) completed a rational review of 20 studies on clinical versus statistical prediction and famously concluded that ''in about half of the studies the two methods are equal and in the other half the clinician is definitely inferior'' (p. 119) and described the clinician as a ''second rate accounting machine.'' In response, Holt (1958) proposed a ''sophisticated clinical decision-making'' approach and argued that ''under a great show of objectivity or at least bipartisanship, Professor Meehl has actually sold the clinical approach up the river'' (p. 1).

Over the years there have been many challenges to the use of clinical integration methods (Dawes, 1979; Dawes, Faust, & Meehl, 1989; Garb, 1994; Grove & Meehl, 1996; Sawyer, 1966; Wiggins, 1973, 1981) and some defenders (Holt, 1958, 1970; Holtzman, 1960; Hunt, 1959; Jeanneret & Silzer, 1998b; Korman, 1968; Levinson, 1998; Zedlow, 2009). Sawyer's review (1966) clarified some issues by making a distinction between data collection and data combination. However, the research regularly concluded that statistical combination methods were at least as accurate, and often more accurate, than clinical combination methods in predicting various outcomes (also see reviews by Ægisdóttir et al., 2006; Dawes, Faust, & Meehl, 1989; Grove et al., 2000; Marchese, 1992; Silzer, 1984; Sines, 1971). It should be noted that when the two methods were equally accurate, the results were often seen as favoring mechanical integration by the researchers, perhaps not a fair conclusion. Some argue that having equal prediction outcomes favors the statistical approach because of lower costs associated with mechanical combination (Grove et al., 2000). However, we would argue that the costs associated with validating and regularly updating a statistical prediction formula for different positions and different organizations is significant and much higher than using clinical integration methods.

More recently, greater attention has focused on ways of improving the accuracy of clinical judgment (e.g., Arkes, 1981, 1991; Dawes et al., 1989; Faust, 1986; Garb, 1989, 1998, 2005; Shanteau, 1988; Spengler, Strohmer, Dixon, & Shivy, 1995; Spengler et al., 2009) and on identifying the situations and moderators that influence or improve accuracy.

Issues related to assessor judgment and the clinical integration of data are central concerns in individual assessment (Jeanneret & Silzer, 1998b; Ryan & Sackett, 1987, 1992, 1998). Some seem to reject clinical judgment out of hand, such as Highhouse who dismisses judgment as ''intuitive and subjective'' (2002, 2008) and argues that people have ''an inherent resistance to analytical approaches'' (2008, p. 337). We disagree and suggest that his criticism is neither theoretically nor factually true for many

I–O assessment psychologists. We and others make a distinction between ''effortless intuition and deliberate reasoning'' (Kahneman, 2003, p. 697). Although we agree that assessors need to work at improving their accuracy, research evidence suggests that clinicians can show sound judgment in specific situations (Dawes, 1994; Ericsson & Lehman, 1996; Faust, 1991; Garb, 2005; Holt, 1958; Karelaia & Hogarth, 2008; Kaufmann & Athanasou, 2009; Lichtenberg, 1997; Shanteau & Stewart, 1992; Spengler, 1998; Spengler et al., 2009; Westen & Weinberger, 2004).

Certainly one of the important issues in individual assessment concerns how assessment information is integrated. This has been a concern raised by numerous assessment researchers (Jeanneret & Silzer, 1998b; Kuncel et al., 2008; Ryan & Sackett, 1998; Silzer, 1984; Silzer & Jeanneret, 1998). There has been limited research specifically in the assessment area. Silzer (1984) found that mechanical combination methods were generally more accurate than clinical strategies in predicting an *overall general effectiveness* performance criterion, although a clinical synthesis approach was equally accurate to mechanical methods (clinical synthesis gives the statistical prediction to the clinician before making a final recommendation). However, there were no significant differences in accuracy between clinical and statistical methods when prediction is focused on specific performance areas (such as work motivation). He also found that the most accurate clinicians achieved higher validity coefficients than any of the other data combination methods (including all the mechanical integration approaches). Kuncel et al. (2008) in a meta-analysis on selection decisions (across a broad range of situations and decision tasks) found that for work outcomes the average validity coefficient for mechanical data combination methods was .36 and for clinical combination methods it was .25.

We think that there are four different stages of information integration in an individual assessment: (a) integrating information *within* one assessment method or source (i.e., within an interview, an exercise or a personality inventory); (b) integrating information *across* assessment methods and sources at the dimension or performance factor level; (c) integrating *across* dimensions to understand and reach an integrated assessment profile; and (d) integration of all assessment information on the individual with knowledge about the position requirements, performance expectations, and organizational context into an overall recommendation. We would encourage researchers to study assessor integration to determine whether these are, in fact, operationally distinct stages and whether assessors are more effective at some stages than others. Our suspicion is that assessors will be most effective in the first three stages but struggle with the last stage.

There have been two recent and relevant meta-analytic studies completed on clinical versus mechanical prediction (Ægisdóttir et al., 2006; Grove et al., 2000). The earlier Grove et al. study (2000) included 136 studies from the fields of medicine, psychology, forensics, and education. The more recent Ægisdóttir et al. study (2006) focused on 67 studies involving predictions in the field of mental health.

The two studies reached some similar conclusions. Grove et al. (2000) found that ''mechanical prediction techniques were *about 10% more accurate* than clinical predictions. Mechanical prediction substantially outperformed clinical prediction in 33%–47% of studies examined. In half of the studies we analyzed, the clinical method is approximately as good as mechanical prediction,... in only a few studies (6%–16%) were they substantially more accurate'' (p. 19). Ægisdóttir et al. (2006) concluded that ''the overall effect of clinical versus statistical prediction ... indicated a *13% increase in accuracy* using statistical versus clinical methods.... clinical decisions are accurate 47% of the time..., whereas statistical decisions are accurate 53% of the time'' (p. 359).

Given the different study samples, it is important to note the similarity of a 10%–13% accuracy advantage in general

https://doi.org/10.1111/j.1754-9434.2011.01341.x Published online by Cambridge University Press

for mechanical methods. This is a very modest effect size and might be influenced by many other variables. For example, the specific methodology decisions (e.g., coding of data) made when conducting a meta-analysis may bias the results. Grove et al. (2000) note that for their study the inclusion/exclusion rules required that the ''clinician and the mechanical procedure had to have access to the same (or almost the same) predictor variables and predict a common criterion'' (p. 20). But this seems to benefit the mechanical approach and clearly limits the clinical approach. They also tried to use ''that judge (or group of judges) . . . with the most experience with the particular prediction task at hand. If the study did not describe the experience or training of the judges well enough to make such distinctions, then we took the *median* of all judges'' (p. 20). However, on the mechanical side they ''preferred cross-validated rules to ones that were not; among cross-validated rules, we took the one with the highest accuracy'' (p. 20). The use of cross-validated mechanical formulas has long been viewed as a likely source of bias in these studies. The researchers acknowledged that some of their research procedures ''might bias the results''(p. 20).

These studies did find some interesting moderator variables that influenced effect size:

- *Type of prediction.* Mechanical prediction had a much greater advantage in medical and forensic settings (Grove et al., 2000). Predictions of violence or academic performance were much more accurate with statistical techniques, whereas treatment length was predicted equally well by both methods (Ægisdóttir et al., 2006).
- *Familiarity with the prediction setting* (Ægisdóttir et al., 2006). Familiarity did not help clinicians do better than statistical methods; in fact, clinicians fared worse.
- *Predictor data* (Grove et al., 2000). Mechanical prediction was favored more when the results of an interview are used as predictive data than when no interview data were available. Counterintuitively, clinical predictions were outperformed by a greater margin when interview data were available to the clinician. Use of medical data (quantitative data) as predictors is associated with smaller differences. But regarding the use versus nonuse of psychological tests, trait ratings or behavioral observations did not impact effect size.
- *Amount of information* (Ægisdóttir et al., 2006). When clinicians were given the same or more information than the statistical formula, the formula did better. Increasing the amount of information decreased clinicians' judgment accuracy. More information may not be better.
- *Information about base rates* (Ægisdóttir et al., 2006). Having base rate information available resulted in clinicians approaching the prediction accuracy of statistical methods.
- *Clinical expertness* (Ægisdóttir et al., 2006). Clinicians who were considered experts in a prediction task did better than nonexperts and did as well as statistical methods. When judgments are made by expert clinicians, the difference between clinical and statistical methods seems to disappear.
- *Type of statistical formula* (Ægisdóttir et al., 2006). All statistical types, except logically constructed rules, did better than clinicians. Logical rules are unexpectedly no better and no worse than clinical methods.

There were several variables that did not have any impact on effect sizes: background of clinical judges (e.g., medical versus psychological), judges' level of experience, relative amount of data available to the clinicians versus mechanical formulas, whether the mechanical algorithm was cross-validated or not (Grove et al., 2000), and clinician access to the statistical formula (clinical synthesis, Ægisdóttir et al., 2006).

https://doi.org/10.1111/j.1754-9434.2011.01341.x Published online by Cambridge University Press

There has been a long running discussion on whether these ''horse race'' studies are a fair comparison. Many years ago Holt (1975) noted:

> There is no magic in clinical intuition that enables a clinician to predict a criterion about which he knows little, from data the relationship of which to the criterion he has not studied and to do so better than an actuarial formula based on just such a prior study of predictor—criterion relationships. (p. 120)

Overall, we believe that the results are encouraging for assessment psychologists for several reasons. First, clinical prediction methods *were as accurate* as other methods about 50% of the time. These studies suggest some situations where clinicians do as well or even better than other methods, such as predicting treatment length, having information about base rates, and using clinicians considered experts. Of course there are some moderators that seem to favor mechanical methods such as clinicians' familiarity with prediction setting, when interview results were used as predictors, and having an increasing amount of available information.

These results do focus the discussion on finding those factors that favor or improve clinical judgment. Early on, Meehl (1959) suggested situations that favor clinical prediction, such as open endedness, insufficient time, and highly configural functions; but warned against clinical errors related to the Barnum effect (glittering generalities) and stereotypes. Later he identified two situations where mechanical prediction will not surpass clinical prediction: (a) where the predictor and criterion variables are very similar and (b) when faced with unique idiographic circumstances (Meehl, 1981, personal communication). Pritchard (1980) indicated two situations where clinical judgment should be used: (a) when configural patterns are important and (b) when there is a concern about an individual's predictability. Management and executive individual assessments, in particular, face several of these circumstances, such as configural patterns, unique idiographic circumstances, and concern about an individual's predictability. In a literature review, Korman (1968) argues that ''the relative usefulness of judgmental prediction methods is exemplified by executive assessment procedures'' (p. 316).

Impact of Experience on Accuracy

Early on in the debate over prediction methods, Allport (1961) encouraged putting a greater emphasis on improving clinical accuracy, whereas Holt (1958, 1970) argued for the development of a ''sophisticated clinician approach.''

One factor that may affect the accuracy of clinical judgment is the experience of the assessor. Wiggins (1973) stirred things up by asserting that ''there is little empirical evidence that justifies the granting of 'expert' status to the clinician on the basis of his [or her] training, experience, or information-processing ability'' (p. 131). This has led to a heated debate (see Garb & Grove, 2005; Westen & Weinberger, 2005) with some arguing that experience does not help and may make things worse (Brodsky, 1998; Faust, 1991; Faust et al., 1988; Garb, 1989, 1996; Ziskin, 1995). Others have argued that judgment accuracy should improve with experience (Falvey & Hebert, 1992; Gambrill, 2005; Shanteau, 1988). Wedding and Faust (1989) went so far as to state, ''There are virtually no data suggesting judgmental accuracy is related to experience'' (p. 249). However, a recent APA task force on evidence-based practice gave experience a central role in determining best clinical practice and concluded that ''clinical expertise allows the types of complex decisions that result in well-conceptualized evidenced-based practice'' (Levant, 2005, pp. 10–11).

A recent meta-analysis examined the impact of educational experience and clinical experience on clinical judgment accuracy (Spengler et al., 2009). They analyzed 75 studies that assessed the accuracy of clinical judgments about mental health

https://doi.org/10.1111/j.1754-9434.2011.01341.x Published online by Cambridge University Press

(e.g., diagnosis, prognosis, treatment) and psychological issues (e.g., vocational, personality). The focus was on diagnosing accuracy and not prediction. They found:

- Experienced counselors and clinicians gain almost a 13% increase in their judgment accuracy. Experience (both education and clinical experience) ''improves the ability to predict behaviors and correctly classify client conditions'' (p. 384). Experienced counselors and clinicians are reliably but modestly more accurate than less experienced clinicians.
- Three moderator variables were found. More experienced clinicians had even greater accuracy than less experienced clinicians when (a) diagnosing and formulating treatment recommendations consistent with guidelines (type of judgment task), (b) the criterion was seen as more ambiguous and had low validity (criterion validity), and (c) the study was published in APA journals as opposed to non-APA journals (publication source).
- No moderating effects were found for whether the experience was from education or clinical practice, the breadth of the clinicians' experience, or whether the clinician was given feedback about accuracy.

Spengler et al. (2009) identified some research design features that may have limited the effect sizes. They noted the inherent problem of defining an ''expert'' (Strohmer & Spengler, 1993) and suggested it should be ''based on actual performance related to outcomes and not just peer nominations'' (Spengler et al., 2009, p. 382). They mention the limited range of experience that is used, often comparing ''some experience'' with ''more experience,'' rather than ''little experience'' with ''substantial experience.'' Skovholt, Rønnestad, and Jennings (1997) suggested that 10 to 15 years of clinical experience might be needed to develop expertise, similar to other domains where the ''the highest levels of human performance can only be attained after around 10 years of extended, daily amounts of deliberate practice activities'' (Ericsson & Lehmann, 1996, p. 273). And finally they suggested that the best predictor of judgment accuracy may be individual differences in the assessors, as has been found for therapist individual differences (APA, 2002; Crits-Christoph et al., 1991).

The finding that experience has a modest but reliable affect on diagnostic accuracy and with ambiguous criterion supports our view that assessors are probably more accurate in understanding and describing individuals than in predicting that person's future performance in a complex and ambiguous context. But the effect size seems rather small. It was surprising that feedback did not improve accuracy, but the researchers note that the finding of no effect is based only on two studies. There are mixed views on whether feedback should improve accuracy, and some suggest that feedback can actually be misleading for a clinician (e.g., Garb, 1996). On the whole, however, we are encouraged that experience does, in fact, improve accuracy and suggest that assessor individual differences as moderator variables are worth research attention.

Validity Evidence for Individual Assessment

The discussion on validity of individual assessment needs to first consider the objective or criterion for individual assessment. For some assessment situations, the client is interested in predicting job performance in a specific position for both the short and the long term. For other clients, the primary interest is in knowing as much as they can about the individual so they can make a sound human resource decision (e.g., selection, promotion, high-potential identification, special development opportunities, and mid career development planning).

At the risk of over simplification, we suggest that there are two main objectives for most individual psychological assessments at the senior management or executive

https://doi.org/10.1111/j.1754-9434.2011.01341.x Published online by Cambridge University Press

level: description and performance prediction. Description requires assessors to accurately understand and describe the individual against the assessment factors. This approach relies more on a construct validity argument. Spengler et al. (2009) found that experienced clinicians can make accurate judgments about individuals. We believe experienced assessors can also accurately describe senior managers and executive candidates and provide client organizations with many useful insights and understandings about the individual that would not otherwise be available.

Performance prediction relies more on criterion-related validity. Performance prediction is complex when trying to identify future management and executive job performance. This is partly because executive roles are complex and partly because there are numerous organizational variables that can influence an executive's performance. To predict performance in the short term, the individual's fit with the position, the organizational culture, the CEO or immediate boss, the CEO's expectations, the peer team, the direct report team, the business strategies, and the country culture need to be considered (Silzer, 2002b, 2005). Some of these factors are the contextual performance variables identified by Borman and Motowidlo (1993, 1997), although we would suggest that the context is significantly more complex at the executive and senior management levels (Silzer, 2002b; Silzer & Adler, 2003). It should be noted that other types of assessment, such as assessment centers, also find it more difficult to predict performance than supervisor potential ratings (Gaugler, Rosenthal, Thornton, & Bentson, 1987).

Over the longer term there are additional issues to consider such as the change orientation of the organization and the likely strategic and cultural changes over time. Most of the variables mentioned as considerations for the short term may change significantly over time for some organizations, and the most likely changes must be considered as potential factors that could impact executive performance. Many years ago Ghiselli (1956) alluded to these complexities when he distinguished three types of criteria: static dimensions, dynamic dimensions, and individual dimensions (individual differences). We should also keep in mind that I–O psychology has not progressed very far in explicating or mapping contextual dimensions, especially for executive and senior management positions. Furthermore, we have limited understanding of the interactions of individual differences with context dimensions. Silzer and others have advocated for developing a theoretical structure of context variables to support a better understanding of which ones have the most impact on job performance (Hollenbeck, McCall, & Silzer, 2006). For example, what types of individual profiles work best with demanding and harsh CEOs or in highly collaborative, low risk-taking cultures? This is where the experience and insights of expert assessors add value when matching individuals to context.

Much of the most relevant criterion-related research was conducted in the 1950s and 1960s when individual psychological assessment was beginning to be used more widely in organizations (see Prien et al., 2003; Ryan & Sackett, 1998; Silzer, 1984 for more thorough reviews). Ryan and Sackett (1998) identified a number of core issues with this validation research and concluded that much more research is needed. As a start, they wanted to see more ''detailed descriptions of the individual assessment process used to enable the accumulation of information about how variations in practice relate to effectiveness'' (p. 72). In complementary fashion, Prien et al. (2003) concluded:

> So can an assessment practitioner accurately forecast how an individual will perform a future job?... the studies that directly evaluate this question are dated and few. Furthermore for the most part, the work that has been done represents casual, opportunistic, contaminated, psychometric evaluation. In spite of all of this, the results are still supportive

https://doi.org/10.1111/j.1754-9434.2011.01341.x Published online by Cambridge University Press

of the process. Therefore the answer would appear to be ''Yes.'' (p. 17)

More recently, Roller and Morris (2008a, b) completed a meta-analysis of the relationship between individual psychological assessment and job performance. They analyzed 25 samples from 18 studies, after excluding 4 studies either because the assessment included a group exercise (although we think this is a legitimate individual assessment data source) or because there was not enough information to derive a validity coefficient. They also did not include any studies that used a statistical or mechanical integration approach and, for almost all of the included studies, the assessor was a psychologist.

They distinguished between studies that predicted overall ratings of job performance and those that predicted administrative decisions (such as change in job level or salary). They found considerable variability among the studies such as ''the content of the assessment battery, the standardization of the assessment procedures, characteristics of the assessors and the degree of structure in the process'' (p. 8). They found two main effects:

- The estimated mean corrected validity of individual assessments *predicting overall ratings of job performance* is .26 (with a 95% confidence interval of .15–.31).
- The estimated mean validity of individual assessments predicting *administrative decisions* (change in job level or salary) is .17 (with a 95% confidence interval of .05–.30).

They also found some moderator variables that influenced overall effect size:

- *Single versus multiple assessors*. Validities were higher for processes that used multiple assessors (.33) than for those using single assessors (.18).
- *Same versus different assessor across candidates*. Validities were higher when the same assessor or assessors were used across all candidates (.41) than when different assessors were used (.17).
- *Occupational group*. Corrected mean validities were higher when individual assessment was used for manager positions (.47), than for professional positions (.24) and for all other occupational groups (.16).
- *Research design. Primary studies* (i.e., studies that required direct interaction between the participant and the assessor who made the prediction and in which assessors had primary access to all standardized test data and interview information upon which a clinical judgment was made for each candidate) had a mean corrected validity of .20. *Secondary studies* (i.e., studies in which predictions were made based on a final report provided to the hiring organization and in which validation study assessors had only secondary access to the assessment report and did not have any direct contact with study participants) had an average corrected validity of .19. *Noninteractive primary studies* (i.e., studies that included only standardized tests in the assessment procedure upon which a clinical assessment was made and in which no interview may have occurred between any assessors and candidates) had a mean corrected validity of .65. This is commonly called a test battery assessment (TBA).

No moderating effects were found for the number of assessment tools and methods used (using six or more assessment tools or methods vs. using fewer) or for the degree of procedure standardization across candidates (using same vs. different procedures).

These are very intriguing results. The overall validity coefficient for predicting job performance ratings is very encouraging. We concur with the researchers that ''individual assessments can produce high validities when they are designed so that assessor bias is limited as much as possible'' (p. 36). It is not surprising that job

https://doi.org/10.1111/j.1754-9434.2011.01341.x Published online by Cambridge University Press

performance is better predicted than job level or salary changes, as there are many tangential variables that can affect these later criteria independent of the individual's performance. In addition, using the same assessor(s) across candidates makes logical sense and is typically practiced in many assessment programs. It reduces the additional interrater reliability issues that Ryan and Sackett explored in their 1989 study. It makes sense that having multiple assessors involved in each assessment could reduce assessor bias and errors. We support having second assessors review assessment conclusions before reports are finalized or even having regular assessment case reviews to discuss and provide feedback on specific assessments.

As our focus has been on management and executive-level individual assessment, it is satisfying to see the higher validities for management assessments, although it is important to point out that most of the managerial samples were first-line supervisors in manufacturing firms with some department managers and general division managers (generally not the organizational level we have focused on but closer than the other occupational groups in the meta-analysis).

Perhaps one of the most telling findings is that test battery assessments did far better than assessments where the assessor had direct interaction with the individual. This is not a surprise to us who have experience working with both types of assessments. However, at the management and executive level we think that using only a test battery would be far less effective than for other positions because the higher-level jobs are much more complex and nuanced (see Hollenbeck, 1994; Silzer, 2002b; Silzer, Fulkerson & Hollenbeck, 2002). Test batteries would only evaluate individuals on very general factors (mostly cognitive factors) that are far too broad to differentiate specific executive success. We think test batteries are best suited for less complex positions (e.g., sales and individual contributors) and for general screening of a large pool of lower level candidates rather than for assessment of a few final candidates for an executive role.

As Roller and Morris (2008b) have pointed out, ''the validity of individual assessment probably depends on the validity of the assessment methods that are used'' (p. 5). Selecting well designed and construct-valid tests and assessment methods is a critical step to building a valid assessment process. We refer the reader to other sources for discussions about identifying and using valid assessment tools and methods (Jeanneret & Silzer, 1998a; McPhail & Jeanneret, in press; Prien et al., 2003; Ryan & Sackett, 1998; Scott & Reynolds, 2010), as well as on the construct validity argument itself (Landon & Arvey, 2007).

At the executive level, collecting criterion-related validity evidence is almost impossible for several reasons: Each executive position is complex and often unique, the key contextual factors that need to be considered are numerous and usually different for each assessment and organization, and an assessment psychologist may only see the final one or two candidates (so there is significant range restriction). Other problems also exist such as criterion contamination. Despite the debate going on in the assessment center field (Arthur, et al., 2008; Lance, 2008; Rupp, Thornton, & Gibbons, 2008), we do not think a job sampling approach to establish job relatedness (as suggested by Lance, 2008) is feasible, given the complexity of executive jobs and the significant influence of contextual factors. Others agree with us on this point (Sackett & Arvey, 1993). Sackett and Arvey suggested that ''the bottom line is that the psychologist is making a judgment about a construct labeled something like job suitability'' (p. 439). They further state that:

> The psychologist needs to justify rationally three components of the assessment process: the constructs to be assessed, the linkage between the measurement tools and these constructs and a causal model relating the constructs to the overall decision. . . . All of these and their

https://doi.org/10.1111/j.1754-9434.2011.01341.x Published online by Cambridge University Press

> rationale should be made explicit. In essence the assessor is being asked to articulate a model of job performance, and a model that cannot be articulated cannot be critiqued. The assessor who either cannot or will not articulate his or her model of information integration cannot, in our opinion, claim to be making valid inferences. (p. 439)

We fully agree and believe that this perspective fits executive individual assessments. But it takes some effort to identify and articulate those integration models. Often individuals do not have sufficient insight into the policies they use to integrate information (Shanteau & Stewart, 1992) and even if they are aware, they may have difficulty putting them into practice (Ryan & Sackett, 1998). We, however, do support a construct validity approach by identifying and measuring relevant performance-related constructs and articulating a model of performance with a clear integration approach.

What Are the Pressing Research Questions and Policy Issues?

Unfortunately, individual assessment has had little research attention from the I–O psychology researcher community. Needless to say there are many unanswered research questions including questions about assessment tools and methods, assessor judgment and integration criterion issues, providing results and feedback, and assessment design and evaluation. Our list of key research questions and policy issues can be found in Table 1.

Enhancing the understanding and practice of individual psychological assessment requires efforts from both researchers and practitioners. Research on individual assessment requires systematically collecting and retaining assessment data. Assessment psychologists need to build assessment databases in their own assessment practices or in client organizations, which can be made available to researchers. Research–practitioner partnerships need to be established to study some of the key research issues and policy issues mentioned above.

How Can We Effectively Teach and Train Assessment Skills?

Recent research suggests that clinical experience and educational experience ''improves the ability to predict behaviors and correctly classify client conditions'' (Spengler et al., 2009, p. 384). This experience effect was even greater under certain conditions, such as when making diagnoses and treatment recommendations and when predicting a criterion with lower validity. Other research has found that experience improves judgment accuracy. Spengler et al. noted that

> Experienced counselors have been found to differ from novice counselors on a number of cognitive dimensions, including (a) broader knowledge structures, (b) better short- and long-term memory for domain-specific information, (c) efficiency in terms of time spent on client conceptualizations, (d) number of concepts generated, and (e) the quality of the their cognitive schemata about case material (Cummings, Hallberg, Martin, Slemon, & Hiebert, 1990; Holloway & Wolleat, 1980; Kivlighan & Quigley, 1991; Martin, Slemon, Hiebert, Hallberg, & Cummings, 1989; Mayfield, Kardash, & Kivlighan, 1999; O'Byrne & Goodyear, 1997, p. 352).

Others have found judgment accuracy differences across assessors (Dougherty et al., 1986; Graves, 1993; Graves & Karren, 1992; Kwaske & Morris, 2008a; Shanteau, 1988; Silzer, 1984; Spengler et al., 2009). We also know that there are some hurdles, such as judgment biases and decision errors, that inhibit greater judgment accuracy (Garb, 1989, 1996). But we have limited understanding of the specific individual difference variables that affect judgment accuracy.

https://doi.org/10.1111/j.1754-9434.2011.01341.x Published online by Cambridge University Press

Table 1. *Individual Psychological Assessment Research Questions and Policy Issues*

Tools, Methods, and Process

- What assessment tools are more valid and effective in evaluating an individual on specific competencies and performance dimensions? For example, what tools can best assess the construct of strategic thinking skills or leadership in a collaborative culture?
- How can we identify sound assessment models that identify the most crucial pieces of assessment data from all the available data? Can we identify useful configural patterns in assessment data (such as the interaction of dominance and sociability scales on the CPI [California Psychological Inventory])?
- How can we better conceptualize and operationalize assessment constructs, performance dimensions, and organizational variables that contribute to performance effectiveness (Arthur, Day, & Woehr, 2008; Klimoski & Zukin, 2003).
- What are the core variables that determine an executive's organizational fit (Silzer, 2002b)?
- Are some tools and methods more effective than others for certain assessment objectives such as career planning, development, or assessing potential?
- How can an assessment interview be designed so it improves assessor judgment accuracy? What is the right balance between having sufficient structure and having the flexibility to pursue critical assessment information?
- What are the skills that lead some assessors to achieve higher accuracy with less structured formats (Dougherty et al., 1986; Dreher et al., 1988)?

Assessor Judgment and Integration

- What are the critical components, core process steps, and likely judgment errors in the data integration process? How can we leverage the judgment and decision-making literature (Dalal et al., 2010) to improve assessor decisions.
- Are there more effective ways that assessors can leverage the construct validity of assessment tools and methods and identify more user-friendly models to make more accurate judgments (Katsikopoulos, Machery, Pachur, & Wallin, 2004)?
- How can we more effectively sort out organizational context variables in order to focus the integration process on those variables that are most influential on performance?
- How can we guide assessors to be more consistent and reliable in how they integrate assessment data (see Weiss & Shanteau, 2004, on the virtues of consistency over consensus)?
- Under what conditions does assessment experience and expertise improve accuracy? Do we need to introduce a second assessor into the assessment process to improve judgment reliability?
- What are the key assessor skills, abilities, knowledge, and individual differences that contribute to judgment accuracy (Kwaske & Morris, 2008a, b; Prien et al., 2003; Spengler et al., 2009)?
- How can we select for and train for these assessment KSAs? What courses or experiences should be included in PhD graduate programs that prepare graduates for conducting individual assessments?
- Under what conditions can assessors effectively determine an overall assessment recommendation (OAR) and what decision methods are most effective to do this?

Criterion Issues

- How can we better describe and structure the myriad of context variables in an organization and identify how they impact the individual's performance success and the validity of individual assessments? Can we find a core set of contextual variables that should be considered in at least most senior management assessments?
- What are the key outcome variables that need to be considered when evaluating the contributions of individual assessment (in addition to job performance)?

Providing Results and Feedback

- What are the most effective ways of communicating assessment results to the organization and the individual to facilitate understanding and acceptance?
- How can assessment results be best leveraged for broad organizational change and global impact (Davis, 1998; Fulkerson, 1998; Silzer, 1998)?

https://doi.org/10.1111/j.1754-9434.2011.01341.x Published online by Cambridge University Press

Table 1. *Continued*

Individual Assessment Design and Evaluation
• How should we adapt the assessment process and tools for different assessment objectives?
• How can assessment feedback be used effectively to evaluate and improve assessor accuracy?
• How can we establish professional guidelines for conducting individual assessments? Should we certify or license assessment psychologists in order to establish and implement professional assessment standards?
• How can we more accurately determine the costs and benefits of doing individual assessments in an organization?

We have mentioned some differences among assessors related to their knowledge, skills, and abilities. Having observed and trained assessors for many years, we think there are both naturally occurring individual differences and differences related to training and experience. For example, some assessors have a natural interest and curiosity about people. They are better observers of human behavior and notice subtle actions and people differences. They naturally start assembling their own behavioral norms by observing others. One might say they have a naturally occurring sense about people, but other assessors seem more mechanistic and show weaker behavioral observation skills.

These natural differences interact with experiences (both clinical and educational experience) to multiply and widen the assessor differences. Strong observers of behavior seem to obtain more from assessor training and gain a deeper understanding of assessment constructs and methods. It is most easily observed when training assessors for an assessment center. Within the 2 to 4 days of assessor training, it usually becomes very obvious who is learning the most and who will be a skilled assessor and who will not.

We think that this may also be true for graduate students. Certain individual difference variables can probably be measured in graduate applicants or graduate students to indicate who is most likely to be effective at psychologically assessing people. Characteristics such as mental flexibility, interpersonal skills, empathy, tolerance, conceptual and critical thinking, and listening and communication skills may be key predictor variables.

But we equally believe in the importance and value of educational training and clinical experience in developing individual psychological assessment expertise. In 1999, SIOP identified individual assessment as ''area of competence'' to be developed in doctoral-level I–O psychology programs (SIOP, 1999). However, we know of few PhD level graduate programs in I–O psychology that offer individual assessment courses or experience (such as that offered at Baruch College, CUNY). In addition, graduate students can and should be taught decision-making skills that will lead to more accurate judgment and decisions (Dalal et al., 2010; Meier, 1999; Spengler & Strohmer, 2001). A scientist–practitioner model for assessment developed by Spengler et al. (1995) that incorporates methods of scientific hypothesis testing and debiasing techniques also might be included in graduate training.

Why have doctoral graduate programs not followed SIOP guidelines? Part of the reason may be that few faculty members are interested in or feel qualified in individual psychological assessment. The solution is to bring practitioners with assessment expertise onto the doctoral faculty to teach these courses. Instead of ignoring this core area of I–O psychology, doctoral programs should be encouraged to teach this core competency.

However, taking a psychological assessment course or two in graduate school is only one step to becoming an assessment psychologist. It is also important to get supervised assessment experience. I–O

https://doi.org/10.1111/j.1754-9434.2011.01341.x Published online by Cambridge University Press

practitioners need to offer more opportunities to learn assessment skills in a structured and supervised environment. Both of the authors learned individual assessment skills primarily at consulting organizations early in our careers, and approximately 50% of the assessors say they are self taught (Ryan & Sackett, 1987). Consulting internships for graduate students and newly graduated I–O psychologists can also be designed to help them learn effective assessment practices.

Faculty members and practitioners should feel some obligation to provide both educational and practical experience. Here are what we consider to be key components in educating and training assessment psychologists:

- Graduate courses in (a) assessment tools and methods, test interpretation, behavioral observation, interviewing, integration, and judgment; (b) psychometrics and measurement theory; (c) selection; and (d) personality theory. Other courses that would also be beneficial include work analysis (particularly of single incumbent positions), leadership, and social psychology.
- Supervised assessment internships with active assessment psychologists for 6–12 months. In fact, we would suggest that it should be a requirement for certification or licensure as an assessment psychologist. We believe that individual psychological assessment is a core practice area in psychology and if individuals are going to label themselves as a psychologist in practice and conduct individual assessments, then they should comply with the relevant state statutes (i.e., be licensed). We believe that most assessment psychologists would get licensed if the requirement for licensure were more appropriate to I–O psychology (Silzer, Cober, Erickson, & Robinson, 2008; Silzer, Erickson, & Cober, 2009).

We should also note that individual psychological assessment is rarely mentioned in I–O psychology textbooks. Although the I–O psychology academic community has long resisted being required to offer particular courses, they have largely chosen to ignore core competence areas of I–O practice, such as individual assessment. Fortunately, there has been a steady descriptive literature on individual assessment by I–O practitioners that graduate programs can use as a foundation (see Goodstein & Prien, 2006; Hanson & Conrad, 1991; Jeanneret & Silzer, 1998a; McPhail & Jeanneret, in press; Prien et al., 2003).

How Can I–O Psychologists Improve Their Individual Assessment Practices?

It seems that individual assessment is widely used and well regarded by I–O psychology practitioners but generally ignored by I–O researchers and by most academic graduate programs with only a few notable exceptions (Garman, 2002; Ryan & Sackett, 1998).

Research in the area of assessment practice is very limited. Ryan and Sackett (1987, 1992) made an important contribution to assessment practice by surveying several different pools of assessment psychologists about their practice. However, the survey data cannot identify the best practices in assessment. We draw upon more than 60 years of combined experience in individual psychological assessment in sharing a few of our learnings from that experience. This is not a primer on how to do an individual assessment (see other sources, such as Hanson & Conrad, 1991; Jeanneret & Silzer, 1998a; McPhail & Jeanneret, in press; Prien et al., 2003) but rather some suggestions for experienced assessors. These suggestions favor a structured and systematic assessment process that encourages assessors to be more consistent and transparent in their assessment judgments.

1. Be clear on the assessment objective and outcome. In some situations

https://doi.org/10.1111/j.1754-9434.2011.01341.x Published online by Cambridge University Press

organizational fit may be an equally important assessment objective.
2. Obtain a good understanding of the specific position requirements and organizational context variables that will impact performance.
3. Be aware of the construct validity of the assessment tools and methods that are being used and how they relate to the assessment dimensions and constructs that are being assessed.
4. Consider using a semistructured interview process that follows a structured format, but allows opportunities to probe for critical information and to test hypotheses.
5. Structure the assessment integration and judgment process so you have specified in advance which data are considered for which assessment dimensions.
6. Use structured support tools that will help you be more consistent in your judgment, such as interview guidelines, anchored rating scales, assessor integration models, and an integration matrix (a multidimension multimethod matrix similar to an assessment center integration form).
7. Be prepared to defend your integration and judgment decisions to others. Be familiar with typical (and your own) biases and judgment errors; use both confirming and disconfirming hypothesis testing.
8. Cautiously make overall selection recommendations and consider using a structured and consistent decision-making model for similar positions.
9. Have another assessor review your assessment cases and reports and provide feedback to you on the strength of your judgment links. Consider forming an assessment group that periodically reviews and discusses assessment cases, for both evaluative feedback and for training.
10. Seek out regular feedback from client organizations on the later performance of individuals you have assessed and ask for feedback on your assessment accuracy.

Conclusion

Individual psychological assessment is a core competency of I–O psychology. We urge our fellow I–O psychologists to join in efforts to maintain and enhance this competency through better understanding of current IPA processes and practices, more research aimed at improving IPA, increased emphasis in graduate training on IPA-related knowledge and skills, and a commitment by current IPA practitioners to continuous improvement and to finding ways to support critical research.

We realize that there are those who take issue with IPA as a practice domain and argue that it does not meet the level of scientific rigor necessary to be appropriate for the practice of I–O psychology. Other critics are concerned that if IPA gains greater recognition as an acceptable practice it would place significant burden, such as licensing or certification, on I–O psychologists and demand unnecessary (and perhaps from their perspective, inappropriate) credentialing. This article is also a reply to those critics and skeptics, pointing to the benefits organizations and individuals attribute to IPA, the broader research on clinical judgments from which we can build, and the lessons learned from practitioners who have experienced success in this arena. Preparing this article has made us recognize that even after our years of experience there is still a vast amount about IPA that we do not know or fully understand. We look forward to replies that add to or challenge our perspectives.

References

Ægisdóttir, S., White, M., Spengler, P., Maugherman, A., Anderson, L., Cook, R., . . . Rush, J. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist, 34*, 341–382.

Allport, F. H. (1937). Teleonomic description in the study of personality. *Character and Personality, 5*, 202–214.

https://doi.org/10.1111/j.1754-9434.2011.01341.x Published online by Cambridge University Press

Allport, G. W. (1937). *Personality: A psychological interpretation.* New York, NY: Holt.

Allport, G. W. (1961). *Pattern and growth in personality.* New York, NY: Holt, Rinehart and Winston.

Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology, 64,* 431–441.

American Psychological Association. (2002). Criteria for evaluating treatment guidelines. *American Psychologist, 57,* 1052–1059.

Arkes, H. R. (1981). Impediments to accurate clinical judgment and possible ways to minimize their impact. *Journal of Consulting and Clinical Psychology, 49,* 323–330.

Arkes, H. R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin, 110,* 486–498.

Arthur, W., Day, E. A., & Woehr, D. J. (2008). Mend it, don't end it: An alternative view of assessment center construct validity evidence. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1,* 105–111.

Arthur, W., Jr, Woehr, D. J., & Maldegen, R. (2000). Convergent and discriminant validity of assessment center dimensions: An empirical re-examination of the assessment center construct-related validity paradox. *Journal of Management, 26,* 813–835.

Borman, W. C., & Motowidlo, S. J. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 71–98). San Francisco, CA: Jossey Bass.

Borman, W. C., & Motowidlo, S. J. (1997). Task performance and contextual performance: The meaning for personnel selection research. *Human Performance, 10,* 99–109.

Brodsky, S. L. (1998). Forensic evaluation and testimony. In G. P. Koocher, J. C. Norcross, & S. S. Hill (Eds.), *Psychologists' desk reference.* New York, NY: Oxford University Press.

Campbell, J., Levy, P., Murphy, K., Schmitt, N., & Weiss, H. (2009, April). Letter to the SIOP Executive Committee. *The Industrial—Organizational Psychologist, 46,* 11–12.

Casas, J. M., Brady, S., & Ponterotto, J. G. (1983). Sexual preference biases in counseling: An information processing approach. *Journal of Counseling Psychology, 30,* 139–145.

Cattell, R. B. (1937). Measurement versus intuition in applied psychology. *Character and Personality, 6,* 114–131.

Cooper-Thomas, H., & Trayes, J. (2008). Preoccupied with predictive validity: An alternative perspective. *Consulting Psychology Journal, 60,* 286–292.

Crits-Christoph, P., Baranackie, K., Kurcias, J. S., Beck, A. T., Carroll, K., Pery, K., Zitrin, C. (1991). Meta-analysis of therapist effects in psychotherapy outcome studies. *Psychotherapy Research, 2,* 81–91.

Cummings, A. L., Hallberg, E. T., Martin, J., Slemon, A., & Hiebert, B. (1990). Implications of counselor conceptualizations for counselor education. *Counselor Education and Supervision, 30,* 120–134.

Dalal, R. S., Bonaccio, S., Highhouse, S., Ilgen, D. R., Mohammed, S., & Slaughter, J. E. (2010). What if industrial–organizational psychology decided to take workplace decisions seriously? *Industrial and Organizational Psychology: Perspectives on Science and Practice, 3,* 385–405.

Davis, S. L. (1998). Assessment as organizational strategy. In P. R. Jeanneret & R. F. Silzer, (Eds.), *Individual psychological assessment: Predicting behavior in organizational setting* (pp. 363–390). San Francisco, CA: Jossey-Bass.

Dawes, R. (1979). The robust beauty of improper linear models in decision making. *American Psychologist, 34,* 571–582.

Dawes, R. M. (1994). *House of cards: Psychology and psychotherapy built on myth.* New York, NY: Free Press.

Dawes, R., Faust, D., & Meehl, P. (1989). Clinical versus actuarial judgment. *Science, 243,* 1668–1674.

Day, D. (2009). Executive selection is a process not a decision. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 2,* 159–162.

Dougherty, T. W., Ebert, R. J., & Callender, J. C. (1986). Policy capturing in the employment interview. *Journal of Applied Psychology, 71,* 9–15.

Dreher, G. F., Ash, R. A., & Hancock, P. (1988). The role of traditional research designs in underestimating the validity of the employment interview. *Personnel Psychology, 41,* 315–327.

Dunnette, M. (1998). Emerging trends and vexing issues in industrial and organizational psychology. *Applied Psychology: An International Review, 47,* 129–153.

Einhorn, H. (1986). Accepting error to make less error. *Journal of Personality Assessment, 50,* 387–395.

Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgment: Persistence of the illusion of validity. *Psychological Review, 85,* 395–416.

Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology, 32,* 53–88.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology, 47,* 273–305.

Falvey, J. E., & Hebert, D. J. (1992). Psychometric study of clinical treatment planning simulation (CTPS) for assessing clinical judgment. *Journal of Mental Health Counseling, 14,* 490–507.

Faust, D. (1986). Research on human judgment and its application to clinical practice. *Professional Psychology, Research, and Practice, 17,* 420–430.

Faust, D. (1991). What if we had really listened? Present reflections on altered pasts. In D. Cicchetti & W. M. Grove (Eds.), *Thinking clearly about psychology: Vol. I. Matters of public interest* (pp. 185–217). Minneapolis, MN: University of Minnesota Press.

Faust, D., Guilmette, T. J., Hart, K., Arkes, H. R., Fishburne, F. J., & Davey, L. (1988). Neuropsychologists' training, experience, and judgment accuracy. *Archives of Clinical Neuropsychology, 3,* 145–163.

Fulkerson, J. R. (1998). Assessment across cultures. In R. Jeanneret & R. Silzer (Eds.) *Individual psychological assessment: Predicting behavior in organizational settings* (pp. 330–362). San Francisco, CA: Jossey-Bass.

Gambrill, E. (2005). *Critical thinking in clinical practice: Improving the accuracy of judgments and*

decisions about clients (2nd ed.). New York, NY: John Wiley.
Ganzach, Y., Kluger, A., & Klayman, N. (2000). Making decisions from an interview: Expert measurement and mechanical combination. *Personnel Psychology, 53*, 1–20.
Garb, H. N. (1989). Clinical judgment, clinical training, and professional experience. *Psychological Bulletin, 105*, 387–396.
Garb, H. N. (1994). Toward a second generation of statistical prediction rules in psychodiagnosis and personality assessment. *Computers in Human Behavior, 11*, 313–324.
Garb, H. N. (1996). The representativeness and past-behavior heuristics in clinical judgment. *Professional Psychological Research Practice, 27*, 272–277.
Garb, H. N. (1998). *Studying the clinician: Judgment research and psychological assessment.* Washington, DC: American Psychological Association.
Garb, H. N. (2005). Clinical judgment and decision making. *Annual Review of Clinical Psychology, 1*, 67–89.
Garb, H. N., & Grove, W. M. (2005). On the merits of clinical judgment. *American Psychologist, 60*, 658–659.
Garman, A. N. (2002). Assessing candidates for leadership positions. In R. Lowman (Ed.), *Handbook of consulting psychology.* New York, NY: Wiley.
Garner, A. M., & Smith, G. M. (1976). An experimental videotape technique for evaluating trainee approaches to clinical judging. *Journal of Consulting and Clinical Psychology, 44*, 945–950.
Gaugler, B., Rosenthal, G., Thornton, G., & Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology, 72*, 493–511.
Ghiselli, E. E. (1956). Dimensional problems of criteria. *Journal of Applied Psychology, 40*, 1–4.
Goodstein, L. D., & Prien, E. P. (2006). *Using individual assessments in the workplace.* San Francisco, CA: Pfeiffer.
Graves, L. M. (1993). Sources of individual differences in interviewer effectiveness: A model and implications for future research. *Journal of Organizational Behavior, 14*, 349–370.
Graves L. M., & Karren R. J. (1992). Interviewer decision processes and effectiveness: A model and implications for future research. *Personnel Psychology, 45*, 313–340.
Grove, W., & Meehl, P. (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.
Grove, W., Zald, D., Lebow, B., Snitz, B., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*, 19–30.
Guion, R. M. (1998). *Assessment, measurement, and prediction for personnel decisions.* Mahwah, NJ: Erlbaum.
Hanson, C. P., & Conrad K. A. (1991). *A handbook of psychological assessment in business.* New York, NY: Quorum Books.
Hastie, R., & Dawes, R. M. (2001). *Rational choice in an uncertain world* (2nd ed.). Thousand Oaks, CA: Sage.
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.
Highhouse, S. (2008). Stubborn reliance on intuition and subjectivity in employee selection. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 333–342.
Hollenbeck, G. P. (1994). *CEO selection: A streetsmart review.* Greensboro, NC: Center for Creative Leadership.
Hollenbeck, G. (2009). Executive selection: What's right... and what's wrong. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 2*, 130–143.
Hollenbeck, G., McCall, M., & Silzer, R. (2006). Do competency models help or hinder leadership development: A debate. *Leadership Quarterly, 17*, 398–413.
Holloway, E. L., & Wolleat, P. L. (1980). Relationship of counselor conceptual level to clinical hypothesis formation. *Journal of Counseling Psychology, 27*, 539–545.
Holt, R. R. (1958). Clinical and statistical prediction: A reformulation and some new data. *Journal of Abnormal and Social Psychology, 56*, 1–12.
Holt, R. R. (1970). Yet another look at clinical and statistical prediction: Or, is clinical psychology worthwhile? *American Psychologist, 25*, 337–349.
Holt, R. R. (1975). Clinical and statistical measurement and prediction: How not to survey its literature. JSAS Catalog of Selected Documents in Psychology, 5(178), No. 837.
Holtzman, W. H. (1960). Can the computer supplant the clinician. *Journal of Clinical Psychology, 16*, 119–122.
Howard, A. (1990). *The multiple facets of industrial-organizational psychology: Membership survey results.* Bowling Green, OH: SIOP.
Hunt, W. A. (1959). An actuarial approach to clinical judgment. In B. M. Bass & I. A. Berg (Eds.), *Objective approaches to personality study* (pp 169–190). Princeton, NJ: D. Van Nostrand.
International Organization for Standardization. (2010). *Assessment service delivery–Procedures and methods to assess people in work and organizational settings, Part 1: Requirements for service providers.* ISO/TC 230 N 116, Geneva, Switzerland.
Jeanneret, P. R. (1998). Ethical, legal, and professional issues for individual assessment. In R. Jeanneret & R. Silzer (Eds.), *Individual psychological assessment: Predicting behavior in organizational settings* (pp. 88–131). San Francisco, CA: Jossey-Bass.
Jeanneret, P. R. (2009, July). Letter to the Editor. *The Industrial—Organizational Psychologist, 47*, 44–45.
Jeanneret, P. R., & Silzer, R. F. (Eds.) (1998a). *Individual psychological assessment: Predicting behavior in organizational settings.* San Francisco, CA: Jossey-Bass.
Jeanneret, P. R., & Silzer R. F. (1998b). An overview of individual psychological assessment. In P. R. Jeanneret & R. F. Silzer (Eds.), *Individual psychological assessment: Predicting behavior in organizational setting* (pp. 3–26). San Francisco, CA: Jossey-Bass.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58*, 697–720.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases.* Cambridge, UK: Cambridge University Press.

Kaplan, M. F. (1975). Information integration in social judgment: Interaction of judge and information components. In M. F. Kaplan & S. Schwartz (Eds.), *Human judgment and decision processes.* New York, NY: Academic Press.

Karelaia, N., & Hogarth, R. (2008). Determinants of linear judgment: A meta-analysis of lens model studies. *Psychological Bulletin, 134*, 404–426.

Katsikopoulos, K., Machery, E., Pachur, T., & Wallin, A. (2004). *The search for models of clinical judgment: Fast, frugal, and friendly in Paul Meehl's spirit.* Unpublished manuscript Retrieved October 15, 2004, from http://jeannicod.ccsd.cnrs.fr/documents/disk0/00/00/05/33/ijn_00000533_00/ijn_00000533_00.pdf.

Kaufmann, E., & Athanasou, J. (2009). A meta-analysis of judgment achievement as defined by the lens model equation. *Swiss Journal of Psychology, 68*, 99–112.

Kivlighan, D. M. Jr., & Quigley, S. T. (1991). Dimensions used by experienced and novice group therapists to conceptualize group processes. *Journal of Counseling Psychology, 38*, 415–423.

Klimoski, R. J., & Zukin, L. B. (2003). Psychological assessment in industrial/organizational settings. In J. R. Graham & J. A. Naglierie (Eds.), *Assessment psychology* (pp. 317–343). Hoboken, NJ: Wiley.

Korman, A. K. (1968). The prediction of managerial performance: A review. *Personnel Psychology, 21*, 295–322.

Kristof-Brown, A. L., Zimmerman, R. D., & Johnson, E. C. (2005). Consequences of individuals' fit at work: A meta-analysis of person–job, person–organization, person–group, and person–supervisor fit. *Personnel Psychology, 58*, 281–342.

Kuncel, N. R., Klieger, D. M., Connelly, B. S., & Ones, D. S. (2008, April). Mechanical versus clinical data combination in selection and admissions decisions: A meta-analysis. In I. L. Kwaske (Chair). *Individual assessment: Does the research support the practice?* Symposium presented at the 23rd Annual Conference of the Society for Industrial-Organizational Psychology, San Francisco, CA.

Kwaske, I. L., & Morris, S. B. (2008a). *Validating individual assessments: A multilevel, multistage validation of individual assessments.* Unpublished paper.

Kwaske, I. L., & Morris, S. B. (2008b, April). Validating individual assessments: A multilevel, multistage validation of individual assessments. In I. L. Kwaske (Chair), *Individual assessment: Does the research support the practice?* Symposium presented at the 23rd Annual Conference of the Society for Industrial-Organizational Psychology, San Francisco, CA.

Lance, C. E. (2008). Why assessment centers do not work the way they are supposed to. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 84–97.

Landon, T. E., & Arvey, R. D. (2007). Practical construct validation for personnel selection. In S. M. McPhail (Ed.), *Alternative validation strategies: Developing new and leveraging existing validation evidence* (pp. 317–345). San Francisco, CA: Jossey-Bass.

Levant, R. F. (2005, July 1). *Report of the 2005 presidential task force on evidence-based practice.* Retrieved September 15, 2005, from www.apa.org/about/president/ initiatives.html.

Levenberg, S. B. (1975). Professional training, psychodiagnostic skill, and kinetic family drawings. *Journal of Personality Assessment, 39*, 389–393.

Levinson, H. (1998). A clinical approach to executive selection. In R. Jeanneret & R. Silzer (Eds.), *Individual Psychological Assessment: Predicting Behavior in Organizational Settings* (pp. 228–242). San Francisco, CA: Jossey-Bass.

Lichtenberg, J. W. (1997). Expertise in counseling psychology: A concept in search of support. *Educational Psychology Review, 9*, 221–238.

Locke, E. A. (2009, October). Why licensing of I–O psychologists is a very bad idea. *The Industrial—Organizational Psychologist, 47*, 53–55.

Marchese, M. C. (1992). Clinical versus actuarial prediction: A review of the literature. *Perceptual and Motor Skills, 75*, 583–594.

Martin, J., Slemon, A. G., Hiebert, B., Hallberg, E. T., & Cummings, A. L. (1989). Conceptualizations of novice and experienced counselors. *Journal of Counseling Psychology, 36*, 395–400.

Mayfield, W. A., Kardash, C. M., & Kivlighan, D. M. (1999). Differences in experienced and novice counselors' knowledge structures about clients: Implications for case conceptualization. *Journal of Counseling Psychology, 46*, 504–514.

McPhail, S. M., & Jeanneret, P. R. (in press). Individual psychological assessment. In N. Schmitt (Ed.), *The Oxford handbook of personnel selection and assessment selection.* New York, NY: Oxford University Press.

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence.* Minneapolis, MN: University of Minnesota Press.

Meehl, P. E. (1959). Some ruminations on the validation of clinical procedures. *Canadian Journal of Psychology, 13*, 102–128.

Meier, S. T. (1999). Training the practitioner-scientist: Bridging case conceptualization, assessment, and intervention. *The Counseling Psychologist, 27*, 846–869.

Meyer, P. (1998). Communicating results with impact. In P. R. Jeanneret & R. F. Silzer (Eds.), *Individual psychological assessment: Predicting behavior in organizational settings* (pp. 243–282). San Francisco, CA: Jossey-Bass.

Muchinsky, P. M. (2004). When the psychometrics of test development meets organizational realities: A conceptual framework for organizational change, examples, and recommendations. *Personnel Psychology, 57*, 175–209.

Nystedt, L., & Magnusson, D. (1972). Predictive efficiency as a function of amount of information. *Multivariate Behavioral Research, 7*, 441–450.

O'Byrne, K. R., & Goodyear, R. K. (1997). Client assessment by novice and expert psychologists: A comparison of strategies. *Educational Psychology Review, 9*, 267–278.

Ones, D., & Dilchert, S. (2009). How special are executives? How special should executive selection be?

Observations and recommendations. *Industrial and Organizational Psychology, 2*, 163–170.
Pfeiffer, A. M., Whelan, J. P., & Martin, J. M. (2000). Decision-making bias in psychotherapy: Effects of hypothesis source and accountability. *Journal of Counseling Psychology, 47*, 429–436.
Prien, E. P., Schippmann, J. S., & Prien, K. O. (2003). *Individual assessment: As practiced in industry and consulting*. Mahwah, NJ: Erlbaum.
Pritchard, D. A. (1980). Apologia for clinical configural decision making. *American Psychologist, 35*, 676–678.
Ridley, C. R., Li, L. C., & Hill, C. L. (1998). Multicultural assessment: Reexamination, reconceptualization, and practical application. *The Counseling Psychologist, 26*, 827–910.
Roller, R. L., & Morris, S. B. (2008a, April). Individual assessment: Meta-analysis. In I. L. Kwaske (Chair) *Individual assessment: Does the research support the practice?* Symposium presented at the 23rd Annual Conference of the Society for Industrial-Organizational Psychology, San Francisco, CA.
Roller, R. L., & Morris, S. B. (2008b). *Meta-analysis of the relationship between individual assessments and job performance*. Unpublished paper.
Rupp, D. E., Thornton, G. C. III, & Gibbons, A. M. (2008). The construct validity of the assessment center method and usefulness of dimensions as focal constructs. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 116–120.
Russell, C. (2001). A longitudinal study of top-level executive performance. *Journal of Applied Psychology, 86*, 560–573.
Ryan, A. M., & Sackett, P. R. (1987). A survey of individual assessment practices by I/O psychologists. *Personnel Psychology, 40*, 455–488.
Ryan, A. M., & Sackett, P. R. (1989). Exploratory study of individual assessment practices: Interrater reliability and judgments of assessor effectiveness. *Journal of Applied Psychology, 74*, 568–579.
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.
Ryan, A. M., & Sackett, P. R. (1998). Individual assessment: The research base. In R. Jeanneret & R. Silzer (Eds.), *Individual psychological assessment: Predicting behavior in organizational settings* (pp. 54–87). San Francisco, CA: Jossey-Bass.
Sackett, P. R., & Arvey, R. D. (1993). Selection in small N settings. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 418–447). San Francisco, CA: Jossey Bass.
Sarbin, T. R. (1944). The logic of prediction in psychology. *Psychological Review, 51*, 210–228.
Sawyer, J. (1966). Measurement and prediction, clinical and statistical. *Psychological Bulletin, 66*, 178–200.
Schmitt, N. (1978). Comparisons of subjective and objective weighting strategies in changing task situations. *Occupational Behavior and Human Performance, 21*, 171–188.
Scott, J. C., & Reynolds, D. H. (2010). *Handbook of workplace assessment: Evidence-based practices for selecting and developing organizational talent*. San Francisco, CA: Jossey Bass.
Shanteau, J. (1988). Psychological characteristics and strategies of expert decision makers. *Acta Psychologica, 68*, 203–215.
Shanteau, J. & Stewart, T. R. (1992). Why study expert decision making? Some historical perspectives and comments. *Organizational Behavior and Human Decision Processes, 53*, 95–106.
Silzer, R. F. (1984). *Clinical and statistical prediction in a management assessment center*. Unpublished doctoral dissertation. University of Minnesota.
Silzer, R. F. (1998). Shaping organizational leadership: The ripple effect of assessment. In R. Jeanneret & R. Silzer (Eds.), *Individual psychological assessment: Predicting behavior in organizational setting* (pp. 391–444). San Francisco, CA: Jossey-Bass.
Silzer, R. F. (Ed.). (2002a). *The 21st century executive: Innovative practices for building leadership at the top*. San Francisco, CA: Jossey-Bass.
Silzer, R. F. (2002b). Selecting leaders at the top: Exploring the complexity of executive fit. In R. Silzer (Ed.), *The 21st century executive: Innovative practices for building leadership at the top* (pp. 77–113). San Francisco, CA: Jossey-Bass.
Silzer, R. F. (2005). *Playing three dimensional chess: The complexity of executive success*. Presentation at the SIOP Leading Edge Consortium: Leadership at the Top: The Selection, Globalization and Ethics of Executive Talent. St. Louis, Missouri.
Silzer, R. F. (2010, April). Individual psychological assessment: A paradigm shift new approaches to meet organizational and professional needs. In J. Blanton (Chair), *Individual assessment: Where we are; where should we be?* Symposium conducted at the 25th Annual Conference of the Society of Industrial and Organizational Psychology, Atlanta, GA.
Silzer, R. F., & Adler, S. (2003). Selecting managers and executives: The challenge of measuring success. In J. Edwards, J. Scott, & N. Raju (Eds.), *The human resources program evaluation handbook* (pp. 130–152). Thousand Oaks, CA: Sage.
Silzer, R. F., & Church, A. H. (2009). The pearls and perils of identifying potential. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 2*, 377–412.
Silzer, R. F., Cober, R. T., Erickson, A. R., & Robinson, G. (2008). *Practitioner needs survey: Final survey report*. Bowling Green, OH: Society for Industrial-Organizational Psychology.
Silzer, R. F., & Davis, S. (2010). Assessing the potential of individuals: The prediction of future behavior. In J. C. Scott & D. H. Reynolds (Eds.), *Handbook of workplace assessment* (pp. 495–532). San Francisco, CA: Jossey-Bass.
Silzer, R. F., & Dowell, B. E. (Eds.). (2010). *Strategy-driven talent management: A leadership imperative*. San Francisco, CA: Jossey-Bass.
Silzer, R. F., Erickson, A. R., & Cober, R. T. (2009). Licensing and industrial–organizational psychologists. *The Industrial—Organizational Psychologist, 46*, 89–99.
Silzer, R. F., Fulkerson, J. R., & Hollenbeck, G. (2002, April). *Innovative practices for building 21st century executives*. Workshop presented at the Annual Conference of the Society for Industrial and Organizational Psychology, Toronto, Ontario, Canada.
Silzer, R. F., & Hollenbeck, G. P. (1992, April). *Executive assessment*. Workshop presented at the 17th

Annual Conference of the Society for Industrial Organizational Psychology, Montreal, Canada.
Silzer, R. F., & Jeanneret, P. R. (1998). Anticipating the future: Assessment strategies for tomorrow. In R. Jeanneret & R. Silzer (Eds.), *Individual psychological assessment: Predicting behavior in organizational setting* (pp. 445–478). San Francisco, CA: Jossey-Bass.
Sines, J. O. (1971). Actuarial versus clinical prediction in psychopathology. *British Journal of Psychiatry, 116*, 129–144.
Skovholt, T. M., Rønnestad, M. H., & Jennings, L. (1997). Searching for expertise in counseling, psychotherapy, and professional psychology. *Educational Psychology Review, 9*, 361–369.
Smith, K., Shanteau J., & Johnson, P. (2004). *Psychological investigations of competence in decision making*. Cambridge, UK: Cambridge University Press.
Society for Industrial and Organizational Psychology. (1999). *Guidelines for education and training at the doctoral level in industrial/organizational psychology*. Bowling Green, OH: Author.
Spengler, P. M. (1998). Multicultural assessment and a scientist-practitioner model of psychological assessment. *The Counseling Psychologist, 26*, 930–938.
Spengler, P. M., & Strohmer, D. C. (2001, August). *Empirical analyses of a scientist-practitioner model of assessment.* Paper presented at the annual meeting of the American Psychological Association, San Francisco.
Spengler, P. M., Strohmer, D. C., Dixon, D. N., & Shivy, V. A. (1995). A scientist-practitioner model of psychological assessment: Implications for training, practice, and research. *The Counseling Psychologist, 23*, 506–534.
Spengler, P., White, M., Ægisdóttir, S., Maugherman, A., Anderson, L., Cook, R., . . . Rush, J. (2009). The meta-analysis of clinical judgment project: Effects of experience on judgment accuracy. *The Counseling Psychologist, 37*, 350–398.
Stratton, S. A. (1991). *Judgment theory applied to data combination in individual psychological assessment.* Unpublished doctoral dissertation, University of Houston.
Strohmer, D. C., Shivy, V. A., & Chiodo, A. L. (1990). Information processing strategies in counselor hypothesis testing: The role of selective memory and expectancy. *Journal of Counseling Psychology, 37,* 465–472.
Strohmer, D. C., & Spengler, P. M. (1993). Studying mental health counselor clinical judgment: A response to Falvey and colleagues. *Journal of Mental Health Counseling, 15*, 465–474.
Taft, R. (1955). The ability to judge people. *Psychological Bulletin, 52*, 1–23.
Thorndike, E. L. (1918). Fundamental theorems in judging men. *Journal of Applied Psychology, 2*, 67–76.
Veneziano, J. (2010, April). Accurate? Useful? Relevant? Individual assessment research with client partners. In J. Blanton (Chair), *Individual assessment: Where we are; where should we be?* Symposium conducted at the 25th Annual Conference of the Society for Industrial and Organizational Psychology, Atlanta, GA.
Wampold, B. E., Casas, J. M., & Atkinson, D. R. (1981). Ethic bias in counseling: An information processing approach. *Journal of Counseling Psychology, 28*, 498–503.
Wedding, D., & Faust, D. (1989). Clinical judgment and decision making in neuropsychology. *Archives of Clinical Neuropsychology, 4*, 233–265.
Weiss, D. J., & Shanteau, J. (2004). The vice of consensus and the virtue of consistency. In K. Smith, J. Shanteau, & P. Johnson (Eds.), *Psychological investigations of competence in decision making.* Cambridge, UK: Cambridge University Press.
Westen, D., & Weinberger, J. (2004). When clinical description becomes statistical prediction. *American Psychologist, 59*, 595–613.
Westen, D., & Weinberger, J. (2005). Clinical judgment in science. *American Psychologist, 60*, 659–661.
Whitecotton, S., Sanders, D., & Norris, K. (1998). Improving accuracy with a combination of human intuition and mechanical decision aids. *Organizational Behavior and Human Decision Processes, 76*, 325–348.
Wiggins, J. S. (1973). *Personality and prediction: Principles of personality assessment.* Reading, MA: Addison-Wesley.
Wiggins, J. S. (1981). Clinical and statistical prediction: Where are we and where do we go from here? *Clinical Psychology Review, 1*, 3–18.
Wisch, A. F., & Mahalik, J. R. (1999). Male therapists' clinical bias: Influence of client gender roles and therapist gender role conflict. *Journal of Counseling Psychology, 46*, 51–60.
Wunder, R. S., Thomas, L. L., & Luo, Z. (2010). Administering assessments and decision-making. In J. L. Farr & N. T. Tippins (Eds.), *Handbook of employee selection* (pp. 377–398). New York, NY: Taylor & Francis Group.
Zedlow, P. (2009). In defense of clinical judgment, credentialed clinicians, and reflective practice. *Psychotherapy Theory, Research, Practice, Training, 46*, 1–10.
Ziskin, J. (1995). *Coping with psychiatric and psychological testimony* (5th ed., Vols. 1–3). Los Angeles, CA: Law and Psychology Press.