

# Personnel Selection: Ensuring Sustainable Organizational Effectiveness Through the Acquisition of Human Capital

Robert E. Ployhart

## Abstract

In this chapter I propose a vision for the field of personnel selection: *ensuring sustainable organizational effectiveness through the acquisition of human capital*. This vision is used to organize and review historical approaches to selection, critique contemporary approaches, and identify future research needs. It is argued that while there is a rich history of research on personnel selection practices and techniques, this research falls short of achieving the vision. Rather, scholars will need to take a broader, and frequently more contextualized, orientation if they are to show the organizational benefits of selection procedures.

**Key Words:** Personnel selection, staffing, talent acquisition, assessment, hiring

Visions are powerful ways to stimulate and enact change. Think, for example, of "We will put a man on the moon in 10 years" by John F. Kennedy, "I have a dream" by Martin Luther King, or "Don't be evil" by Google. Visions present an illustration of what should be and what is possible if resources and energy are focused on the vision. Vision statements need not be true today, but instead represent an idealized outcome that will hopefully become true if members work collaboratively toward the vision. In a sense, vision statements are what we want the world to be.

I suggest that the vision of personnel selection should be *ensuring sustainable organizational effectiveness through the acquisition of human capital*. It would be hard to argue for anything else. Selection is fundamentally about acquiring human capital through rigorous assessment methods. Human capital is defined as the aggregate of individual knowledge, skills, abilities, and other characteristics (KSAOs). If selection does not enhance organizational effectiveness in a sustainable manner, then why should we do it? Yet, as innocent as this vision appears, there is still

a long way to go before we make it a reality. Research on personnel selection has been productive over the last 100 years, and as a profession we know a great deal about how to best identify KSAO requirements for jobs, develop methods to assess those KSAOs, and combine the scores from the assessments into appropriate hiring decisions. But this by itself does not ensure that selection creates *sustainable organizational effectiveness*. Rather, fulfilling this vision will require a broader perspective than the individual-level research that has so dominated the selection profession since its inception.

This chapter has three purposes. First, it will briefly review the historical antecedents of personnel selection, giving a particular emphasis to contrasting changes in private industry with personnel selection research and practice. Second, it will provide a critical review of current selection research and practice. Included in this review is a process and framework for thinking about predictor constructs and methods. Finally, the chapter concludes with a look toward the future and identifies several directions for future research.

## Scope of the Chapter

The scope of this chapter is on personnel selection, that is, the processes, methods, and strategies used to make appropriate hiring decisions. This chapter is not about recruitment (i.e., attracting people to become and remain applicants), nor is it about criteria and job performance. However, please recognize that selection is dependent on both recruitment and criteria. Selection is most important when there are many good candidates from which to choose. There are a variety of recruiting practices and policies that need to be considered, which are discussed in Yu and Cable (chapter 7 of this handbook). The choice of selection KSAOs and methods is dependent on the nature of performance criteria (Wallace, 1965). This chapter will discuss criteria, but for a full treatment of criterion issues please see Campbell (chapter 6 of this handbook).

Various inferences of validity are used to support a relationship between KSAOs and criteria, but this chapter is not going to discuss the concept of validity in great detail. Modern interpretations of validity (e.g., Messick, 1995) treat all inferences of validity as supporting *construct validity* (whether the assessment measures the intended KSAO; whether scores are appropriate for their intended purpose). This chapter's use of validity is consistent with this interpretation, but it will frequently be necessary to discuss different types of validity evidence. These types include *criterion-related* (correlations between measures of KSAOs and criteria) and *content* (judgment of the overlap between a KSAO or predictor method and a criterion). Key features of construct validity also include *convergent* validity (strong correlations between measures of similar KSAOs) and *discriminant* validity (weak correlations between measures of dissimilar KSAOs).

Finally, this chapter will generally avoid discussing legal issues and implications. This is intentional, even though the legal and political environments obviously have a substantial impact on what is considered acceptable selection practice. Laws and regulations vary dramatically across countries and cultures (Myors et al., 2008). Most published selection research has been conducted in the United States, but discussing these laws makes the present chapter unnecessarily country specific. Therefore, this chapter will generally avoid consideration of legal issues, but it will consider factors that lead to legal issues (e.g., demographic subgroup test score differences).

## The Past: Steadfast Selection in an Ever-Changing World

Personnel selection is considered by many to be one of the great success stories of industrial and organizational (I/O) psychology. Schmidt and Hunter (1998), summarizing 85 years of research, show some pretty impressive relationships between selection predictors (e.g., cognitive ability, conscientiousness, interviews) and job performance. They also show some pretty impressive *estimates* regarding the financial impact of using these predictors (vs. not using them). One should recognize, however, that this research only partially speaks to the vision of selection proposed above because it ignores the consideration of sustainable organizational effectiveness. Understanding why requires a brief discussion of the changing world of work and the history of selection.

The practice of personnel selection has existed at least since 700 A.D. in China (see Ployhart, Schneider, & Schmitt, 2006). They implemented what were essentially civil service exams to determine who would gain access to desirable government positions. Surprisingly, these exams lasted until the early 1900s. By some reports, the Chinese civil service exams were adapted for use throughout Asia. Around the same time that the Chinese civil service exams were being eliminated, the scientific study of selection in Western countries evolved from the educational testing of Binet and Simon. Their purpose was to use tests of mental ability to identify children who would not perform well in the public school system; these tests evolved into tests of general mental ability. Binet and Simon's functional approach to testing became widely used during World War I by many countries. After the war, these principles and tests were then applied to industry in numerous forms. World War II further solidified the functional value of selection via the application of tests to assign recruits to positions (a process known as *classification*). After World War II, testing and selection were firmly ingrained in civil and private sectors.

Major accomplishments in the period after World War II included advancements in the conceptualization and measurement of KSAOs to include a variety of cognitive and non-cognitive constructs. For example, much of the research on the latent structure of cognitive ability and personality was conducted in the two decades after World War II. In the mid 1960s, there was a call by McNemar (1964) to start showing how individual differences such as intelligence contributed to real world outcomes (a term he

called "social useful" by Schmidt, Hunter, generalization and of this contribution it has had a profound research and practical research that will be based on meta-analysis, it was believed was situation-specific study would need to (even if the job was tion ended the situation that prof SIOP Principles (2010) the role of meta-analysis in edited book by Mui lent review of validit

The major contributions to advance the concept of job performance, an important selection issue; comparing predictions. In the 1990s, selection "reemergence" of personal validity-diversity dilemmas from many of the more racial-ethnic subgroups that dilemma continues, as does research (e.g., counterproductive types of predictor methods). much of the work in tradition.

In contrast, the remained constant in selection research. New economies in many shift from an agricultural economy. The hallmark was the "interchangeable manual, repetitive that the assembly line, he, in fact, got the idea in Chicago. The purpose ensure consistent production manner possible. This performed in terms of the role of personnel selection people working the line. These KSAOs were but not change very much, primarily one of "accuracy."

called "social usefulness"). A decade later, research by Schmidt, Hunter, and colleagues led to validity generalization and meta-analysis. The importance of this contribution cannot be overstated because it has had a profound effect on personnel selection research and practice. For example, much of the research that will be reviewed in the next section is based on meta-analyses. Prior to validity generalization, it was believed that criterion-related validity was situation-specific, and hence a new validation study would need to be conducted in every context (even if the job was the same). Validity generalization ended the situational-specificity hypothesis to the point that professional guidelines such as the SIOP Principles (2003) now explicitly recognize the role of meta-analysis as a validation strategy. An edited book by Murphy (2002) provides an excellent review of validity generalization.

The major contribution of the 1980s was to advance the conceptualization and measurement of job performance, and address a number of fundamental selection issues (e.g., estimating cross-validation; comparing predictive versus concurrent designs). In the 1990s, selection researchers welcomed the "reemergence" of personality, yet grappled with the validity-diversity dilemma, a dilemma that results from many of the most valid predictors having large racial-ethnic subgroup differences. Research on that dilemma continues into the twenty-first century, as does research on the performance domain (e.g., counterproductive work behaviors) and new types of predictor methods and constructs. Thus far, much of the work in the 2000s has followed in this tradition.

In contrast, the world of work has hardly remained constant in the 100 or so years of scientific selection research. Near the end of the 1800s, the economies in many developed countries began to shift from an agricultural economy to an industrial economy. The hallmark of the industrial economy was the "interchangeability" of employees performing manual, repetitive tasks. While many believe that the assembly line was started by Henry Ford, he, in fact, got the idea from the slaughterhouses in Chicago. The purpose of the assembly line was to ensure consistent production in the most efficient manner possible. The job rarely changed and was performed in terms of relatively isolated tasks. The role of personnel selection was to ensure that the people working the line had the necessary KSAOs; these KSAOs were based on discrete tasks that did not change very much. Further, since the job was primarily one of "additive inputs," the economic

value of selection was based on the idea that individual performance could be decomposed into more discrete elements and summed to produce a unit-level estimate of value.

Let's fast forward to modern times, to what many call the "knowledge economy," which is largely an extension of the information economy. In the knowledge economy, the key "products" are ideas and information; hence the critical resource that firms use to compete with each other is human capital (although the overall importance of human capital relative to other forms of capital is variable, as illustrated in the current economic situation). The term *human capital* has its origins in economics (Becker, 1964) but has come to be known as the aggregate KSAOs of a firm's workforce. The term *knowledge worker* has its roots in Drucker (1966) and refers to employees whose primary contribution is based on ideas and information. The shift toward knowledge work is profound because, unlike other forms of capital, human capital cannot be owned by a firm. Employees can choose to withhold effort, switch jobs, or steal ideas and sell them to competitors. Jobs are more fluid because work is increasingly being performed in project teams, which form to complete a project and then disband when the project is finished. Finally, globalization and technological changes such as the Internet have made the world "flat" so that human capital need not be in the same geographic area. Obvious examples include the offshoring of customer service, engineering, and accounting jobs to India. A flat world contributes to greater exposure to people from diverse cultures and languages.

Just as the developed economies have come to recognize the importance of human capital, incredible societal changes have evolved to frequently make human capital in scarce supply. These include an aging workforce, a more diverse workforce, and a global economy. Popularized by the "War for Talent" studies conducted by McKinsey, many firms now realize that talent (a term to reference desired human capital) is critical for their strategic success. For example, a broad sample of almost 7,000 managers found that approximately 90% of them believed recruiting and selection were increasingly difficult challenges affecting their business (Axelrod, Handfield-Jones, & Welsh, 2001). Although the McKinsey research is not without some sizable limitations, the fact remains that many organizational leaders recognize that talent acquisition is critical to their survival. For example, a special report in *The Economist* (2006) highlighted the global challenges

of finding qualified talent to meet the needs of the new economy.

Thus, the modern world is characterized by work that is knowledge-based, dynamic, often uses project teams, and is frequently global in scope. Of course, there are still plenty of jobs that are similar to those in the older industrial age, and many jobs are still primarily hands-on and physical in nature. But the jobs that drive the developed and developing economies are increasingly knowledge-based. Given such incredible change, one would think personnel selection practices have changed as well. Interestingly, this is not so. True, personnel selection practices and processes have evolved through programmatic research, and technology has affected the way in which personnel selection is implemented (e.g., web-based testing), but by and large the basic model for selection is still one of the industrial age, which matches individuals to jobs devoid of context. Although it is doubtful that the intention of validity generalization research is to eliminate a concern for context, this may be what has happened. Selection is still a profession of treating tasks as though they are relatively static and unrelated to other employees, linking individual KSAOs to those tasks isolated from the broader context, and estimating value based on an additive model (see Cascio & Aguinis, 2008a; Ployhart & Schneider, 2002).

Some evidence of this is found in a recent review of research published in the last 45 years at *Journal of Applied Psychology* and *Personnel Psychology*. Cascio and Aguinis (2008b) found that topics on personnel selection were consistently within the top five in both journals. However, the topics studied within personnel selection tended to be somewhat peripheral to those reflecting important human capital trends. Indeed, as they so eloquently noted:

On the basis of our review, if we extrapolate past emphases in published research to the next 10 years, we are confronted with one compelling conclusion, namely, that I-O psychology will not be out front influencing the debate on issues that are (or will be) of broad organizational and societal appeal. (p. 1074)

The field of personnel selection has held steadfast despite a whirlwind of change. This is not as bad as it sounds because, at least in terms of practices designed to maximize individual job performance, the field of personnel selection is at a high level of sophistication. The next section will review the research on these practices. However, in terms of achieving the vision set out at the start of this chapter, personnel selection research must take a broader

perspective, and this perspective will be discussed in the third and final section.

### The Present: We Shall Be Known by Our Models and Methods

This section provides a focused yet critical review of contemporary selection research and practice. The field of personnel selection combines diverse theory and research from a variety of disciplines, including individual differences, differential psychology, psychometrics, and statistics. As such, there is no "theory" of personnel selection, but rather various frameworks that are used to organize the way we research and practice selection (e.g., Aguinis & Smith, 2007; Binning & Barrett, 1989; Heneman & Judge, 2006; chapter 7 in Ployhart et al., 2006; Schmitt, Cortina, Ingerick, & Wiechmann, 2003). Given such diversity of perspectives, it is important to develop an understanding of the basic process of selection and to compare and contrast various predictor methods. Therefore, a general framework of the personnel selection process is first presented, followed by a framework to compare and contrast measures of KSAOs.

#### The Personnel Selection Model

Of Robert Guion's many contributions to the field of selection, one of the most profound is the recognition that the process of personnel selection is really one of hypothesis testing. That is, we make a hypothesis that a given predictor is related to job performance, and then collect the data to test that hypothesis. A less sophisticated way to think about selection is one of making a bet. That is, we bet on a *future* event (i.e., effective job performance) based on limited information we have *today* (i.e., scores on predictors). Such predictions are a tough business, but we make better predictions if we carefully collect and thoroughly attend to the relevant pieces of information. This process is facilitated by using a model of personnel selection specifying how we should collect the relevant information.

Figure 8.1 shows the basic personnel selection model and is adapted from Binning and Barrett (1989). The boxes represent indicators that we can measure; they are manifest scores that can be observed (e.g., test performance, interview scores). Circles represent unobservable latent constructs that cannot be directly observed or measured but are instead inferred through scores on manifest measures. The distinction between manifest variables and latent constructs is important because the boxes represent measures of KSAOs that are affected



Figure 8.1 The line represents the

to some degree of deficiency. The imperfection KSAO predictor represent. Functionality based because it has that make it predictive ability. The or hypothesis one-headed direction, and two recursive

The most make in selection are individual (inference 8) hiring decision expectation or latent job bet, but the direction, the more able to support inference 8 decision inferences that judgment, a

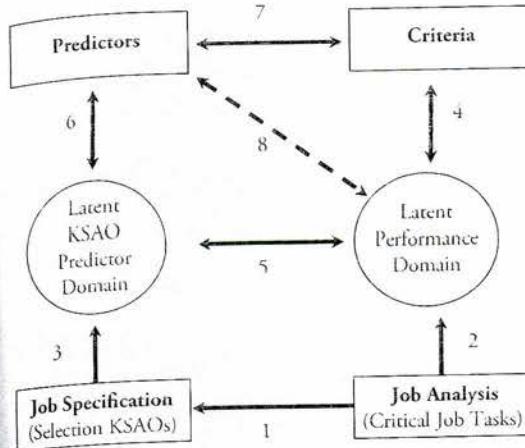
Notice in direction to 8. This process broadly, job specific cat

critical review practice. The reverse theory argues, including psychology, there is no other way we

Aguinis & I.; Heneman et al., 2006; Ann, 2003). An important process of various pre- amework of t presented, and contrast

tions to the found is the nel selection is, we make elated to job i to test that think about is, we bet on nance) based (i.e., scores tough busi we carefully levant pieces ted by using ing how we

nel selection and Barrett ors that we ; that can be view scores). it constructs measured but on manifest manifast variat because the at are affected



**Figure 8.1** The Personnel Selection Framework. The dashed line represents the key interest in selection.

to some degree by unreliability, contamination, and deficiency. The latent constructs are “free” from such imperfections and represent our theory of what the KSAO predictor and performance domains should represent. For example, we may measure cognitive ability based on a written test, but the test is fallible because it has various sources of measurement error that make it an imperfect indicator of “true” cognitive ability. The arrows represent the many *inferences* or hypotheses that occur in the selection process; one-headed arrows represent a specific causal direction, and two-headed arrows represent a covariance or recursive relationship.

The most important hypothesis or bet that we make in selection is that scores on manifest predictors are indicators of latent performance on the job (inference 8 in Figure 8.1). This is because we base hiring decisions on fallible predictor scores, with the expectation that doing so will be related to “true” or latent job performance. This is a far from perfect bet, but the better we collect the relevant information, the more accurate our bet. The better we are able to support inferences 1–7, the more accurate inference 8 becomes. Thus, one can see that hiring decisions are ultimately derived from multiple inferences that should be based on sound reasoning, judgment, and, if at all possible, empirical support.

Notice in Figure 8.1 that there is a specific causal direction to developing and supporting inference 8. This process starts with a job analysis. Speaking broadly, *job analysis* is a systematic, purposeful process designed to comprehensively identify the important tasks and KSAOs required for effective performance on the job. Job analysis follows a specific causal sequence, such that we first identify

the critical tasks and then identify the essential KSAOs necessary to perform these tasks (inference 1). Identification of essential KSAOs is frequently called *job specification* (Harvey & Wilson, 2000); the distinction highlights the fact that the choice of KSAOs and predictors is inherently determined by the job. The essential KSAOs are often called *selection KSAOs* to recognize that they will be used as a basis for hiring.

There are many ways to collect job analysis and job specification information, such as meetings with subject matter experts (SMEs), surveys, observation, and reviews of manuals and standard operating procedures. Historically, there were many different job analysis techniques, such as functional job analysis (Fine & Wiley, 1974), the Position Analysis Questionnaire (PAQ), and critical incident techniques (Flanagan, 1954). The common feature across these alternatives is to identify the critical tasks and KSAOs. Of the many tasks performed on a job, not all are essential. Hence, after compiling a list of all job tasks, job analysis methods often ask SMEs to rate how frequent, difficult, or critical the task is for the job. These and other approaches are comprehensively detailed in Gatewood and Feild (1998) and Brannick, Levine, and Morgeson (2007). However, in today’s world it is more common to use a combination of approaches as relevant, such as SME interviews, observation, and surveys. But private firms are often resistant to devote the considerable money and resources needed to conduct a comprehensive job analysis. In such situations, firms may instead choose to use the U.S. Federal Government’s Occupational Information Network (referred to as O\*NET). The O\*NET system covers nearly every occupation in the United States. It provides a holistic perspective on work and provides information on six major content areas for each job: worker characteristics, worker requirements, experience requirements, occupational requirements, occupation-specific requirements (tasks more specific to the job), and occupation characteristics (labor and economic factors affecting the occupation). This information is also linked to a variety of state databases. O\*NET is freely available on the internet (<http://online.onetcenter.org/>). Regardless of which approach is used, what is most important is that the job analysis is comprehensive and has identified the most essential tasks and KSAOs needed to perform the job.

After identifying the critical tasks and selection KSAOs, the next steps involve defining the latent performance domain (inference 2) and the latent

predictor domain (inference 3). Because these domains are latent, inference 5 cannot be tested, but must be inferred through inference 1. Defining the latent performance domain determines the types of criteria that will be measured (inference 4). These criteria may include job performance behaviors, but also many other types of criteria such as turnover, absenteeism, and safety (see Campbell, chapter 6 of this handbook, for a comprehensive treatment of job performance). Likewise, defining the latent predictor domain determines the types of KSAO constructs and predictor measures that will be used for the basis of making selection decisions (inference 6; see Ryan and Sackett, chapter 5 of this handbook). Inference 7 represents the empirical correlation (i.e., criterion-related validity) of the predictor scores with the criterion scores. Thus, as one can see, appropriate support for inferences 1–7 is necessary to adequately support inference 8 (Binning & Barrett, 1989).

Before leaving Figure 8.1, it is important to recognize that each inference is often based on considerable human judgment. In recent years, it has become appreciated that a host of perceptual, cognitive, and social factors can influence this judgment and ultimately the results and accuracy of job analyses. Morgeson and Campion (1997) present a comprehensive review of factors likely to influence job analysis results (e.g., conformity pressures, impression management). Empirical research has found that job analysis information is frequently affected by these factors (e.g., Morgeson, Delaney-Klinger, Mayfield, Ferrara, & Campion, 2004), as well as organizational context (Dierdorff & Morgeson, 2007). However, the major source of variance in job analysis ratings appears to be due to idiosyncratic differences among the raters (Van Iddekinge, Putka, Raymark, & Eidson, 2005). These findings do not negate the importance of job analysis, but rather raise awareness of the fact that jobs are perceived and in many ways socially constructed (Ilgen & Hollenbeck, 1991).

To provide empirical support for the inferences (hypotheses) shown in Figure 8.1, there are a variety of methodological issues that must be considered. These methodological issues represent the intersection between statistics and psychometrics with personnel selection processes and practices. Importantly, these methodological and statistical issues directly impact the types and nature of support for the inferences shown in Figure 8.1. Hence, questions about the appropriateness of these issues directly relate to questions about the appropriateness

of the conclusions and generalizations we make from selection research. Summaries of methodological and statistical issues are presented in chapter 3 (by Hanges and Wang), and chapter 4 (by DeShon) of this handbook, as well as Johnson and Oswald (2010) and Van Iddekinge and Ployhart (2008). Here only a concise summary is provided.

While the key inference in personnel selection is that of inference 8 in Figure 8.1, one of the best ways we can substantiate this inference is through empirically supporting inference 7 via a correlation coefficient (i.e., criterion-related validity). There are a variety of factors (sometimes called statistical or measurement artifacts) that can distort (usually downward) the size of the correlation coefficient and hence support for inference 7 (and by extension, inference 8).

*Measurement error* represents one of the most pervasive artifacts because every manifest variable contains multiple sources of variance. The extent to which these various sources of variance are unwanted contributes to the error variance of the measure. As alluded to in Figure 8.1 by inferences 5 and 6, measures of predictors and criteria are imperfect representations of the latent underlying “true score.” Measurement error is one of the causes of these imperfections. *Measurement error* contributes to the unreliability of the measure, and the lower the reliability, the lower the criterion-related validity (i.e., the size of the correlation). For example, error variance may stem from the wording of test questions or transient error. As a result, a considerable amount of research has sought to examine what sources of error variance may be present in predictors and criteria, and how to best correct the observed correlations to obtain estimates of the construct relationships. That is, research has sought to take the observed (also known as an *uncorrected*) correlation from inference 7 and, through various corrections, obtain an *estimate* (via a *corrected* correlation) of inference 5. Schmidt and Hunter (1996) provide an excellent overview of different sources of error in psychological measurement. By realizing there are multiple sources of error, the difficulty then comes in trying to understand whether to correct all forms of error variance, and the sequence by which to correct them. This is a complex issue beyond the scope of the present chapter, but interested readers are referred to DeShon (2002), Murphy and DeShon (2000), Schmidt, Le, and Ilies (2003), and Schmidt, Viswesvaran, and Ones (2000). What is important to realize is that different assumptions about what comprises error variance will influence

we make methodological inferences about the validity of selection procedures. Chapter 3 by DeShon and Oswald (2008) provides a detailed discussion of the various types of validity evidence that can be used to support the validity of selection procedures. In this chapter, we focus on the validity of selection procedures based on correlations between predictors and criterion variables.

*Range restriction* represents a second artifact affecting the size of the correlation coefficient used to support inference 7. Correlations are attenuated to the extent that one or both variables lack variance. Range restriction occurs when the variance in the predictor, criterion, or both, is reduced relative to what it would be in the larger population. For example, suppose that we look at the correlation between interview ratings and supervisory ratings of job performance. Unfortunately, the supervisors refuse to make distinctions between employees and rate *all* employees a 10 on a 10-point scale. There will be zero variance in the criterion, and hence the correlation will be zero. However, in reality, if supervisors had been willing to make appropriate distinctions, the correlation coefficient may have been .20. Thus, as with measurement error, range restriction lowers the magnitude of the correlation coefficient, and as with measurement error, there are a variety of corrections that can be used to provide better estimates of inference 5 via extension of inference 7. Sackett and Yang (2000) identified 11 different scenarios that affect the magnitude of indirect and direct range restriction, and the appropriate formulae needed to correct correlations in these situations. Further work on using range restriction corrections in practice is offered by Hunter, Schmidt, and Le (2006), Sackett, Laczo, and Arvey (2002), Sackett, Lievens, Berry, and Landers (2007), and Schmidt, Oh, and Le (2006).

*Should statistical artifact corrections be used and, if so, which corrections should be used?* These simple questions evoke some dramatic responses from I/O psychologists. At the risk of oversimplification, there are two camps; one camp advocates the use of corrections to enhance the size of the correlation coefficient, and the other camp either espouses a conservative approach to the use of corrections or ignores them entirely. This "divide" can be found in numerous places, for example, from the debate over whether to correct rating criteria for interrater unreliability (measurement error discussed above), to the debate over whether personality criterion-related validities are large (discussed below). The crux of the debate is that there is no empirical way to support inference 5, hence the issue is whether the observed (uncorrected) correlation supporting inference 7 provides a reasonable estimate of inference 5. Therefore, any support for inference 5 by

extension of inference 7 must be evaluated in terms of whether the correction procedures are appropriate. This is where opinions differ, and an excellent exchange on this issue is provided by Le, Oh, Shaffer, and Schmidt (2007), Schmidt, Le, Oh, and Shaffer (2007), and Schmitt (2007).

### CREATING PREDICTOR COMPOSITES

Most selection systems will use multiple predictor constructs and methods, such as resumes, interviews, and possibly some form of cognitive, knowledge, or personality tests. When multiple predictors comprise a selection battery, it becomes necessary to combine the scores in some fashion to produce a final hiring decision. This task becomes more complicated when the system is one involving multiple hurdles or cutoffs, such that only applicants who score above a threshold at one stage move on to subsequent stages of the process. Gatewood and Feild (1998) offer an excellent description of different selection system combination strategies. There are multiple ways to estimate predictor importance, including relative importance (i.e., the contribution of each predictor to the overall regression model), incremental validity, and relative weight/dominance analysis. A variety of issues that should be considered when evaluating predictor importance are reviewed by Johnson and Oswald (2010) and Van Iddekinge and Ployhart (2008). One should also recognize that different ways of combining predictors can influence the overall subgroup difference and adverse impact of the selection system (Sackett & Ellingson, 1997; Sackett & Roth, 1996).

### VALIDATION METHODOLOGICAL ISSUES

The two basic criterion-related validation designs include predictive (predictor scores from *applicants* are correlated with their *later* performance scores collected as incumbents on the job) and concurrent (current *incumbent* employees complete the predictors, and their scores are correlated with their performance on the job). A variety of KSAOs may differ between these two contexts, including test-taking motivation, knowledge of the job and organization, and experience. As such, the validation design can influence the magnitude of the correlation observed to support inference 7 in Figure 8.1. The size of the criterion-related validity for non-cognitive constructs and predictors is usually larger in concurrent contexts (Van Iddekinge & Ployhart, 2008). Likewise, a predictor administered to job incumbents may be responded to differently by applicants and hence offer clouded support for

construct validity (inference 6 in Figure 8.1). A good example is personality testing; as discussed below, administering a self-report personality inventory to applicants produces very high mean scores and reduced variance, hence support for inference 6 and thus 7 becomes diminished in predictive validation studies. Similarly, range restriction can occur in different ways and by different amounts in concurrent and predictive studies (Sackett & Yang, 2000).

A new issue that has become important in recent years is the retesting of applicants. Many civil service organizations, and even some private firms, allow rejected applicants to reapply for the job after a specified period of time has elapsed. In general, cognitive test scores tend to increase with repeated testing (Hausknecht, Halpert, Di Paolo, & Moriarty Gerrard, 2007) whereas personality test scores do not (Ellingson, Sackett, & Connolly, 2007; Hogan, Barrett, & Hogan, 2007). Criterion-related validity tends to be a bit higher for the initial test score for cognitive tests (Hausknecht, Trevor, & Farr, 2002).

### A Framework for Comparing Predictors

Ployhart et al. (2006) define a predictor as follows:

Any manifest measure whose scores represent individual differences on latent knowledge, skills, abilities, or other characteristics linked to effective criterion performance, and which is used (at least in part) as the basis for selection or promotion. (p. 366)

Predictors are the “indicators” that we use to gauge whether an applicant is likely to perform well on the job (i.e., inference 8 in Figure 8.1). As a result, selection scholars spend enormous energy seeking to better understand and improve predictors. This research has produced many different types of predictor constructs and KSAOs. *Predictor constructs* are measures of homogenous KSAOs and include cognitive, knowledge, personality, values/needs/interests, and physical domains. *Predictor methods* assess multiple constructs simultaneously (i.e., are multidimensional) and include experience and biographical data (biodata), interviews, situational judgment tests, assessment centers, and work samples. Thus, there are many choices in predictors of job performance, even for the same type of performance and job. As a result, selection experts often must make choices about which predictor measures to use for a given KSAO (i.e., inference 6 in Figure 8.1). Following Aguinis and Smith (2007), Heneman and Judge (2006), and Ployhart

et al. (2006), this section uses a framework that consists of multiple dimensions to compare and contrast predictors: criterion-related validity, fakability, subgroup differences, economic return, and user acceptability.

*Criterion-related validity* represents the degree to which scores on the predictor are related to scores on the criterion (inference 7 in Figure 8.1). Criterion-related validity is where “the rubber meets the road;” it is the *empirical* evidence to support inferences that scores on the predictor are related to latent job performance (Binning & Barrett, 1989). While criterion-related validity studies are frequently not feasible in practice (e.g., need for a large sample size), criterion-related validity is traditionally one of the most important indices of a predictor’s value and one of the key drivers of economic utility (at least as estimated in classic utility models). According to Ployhart et al. (2006), uncorrected criterion-related validity estimates of .10, .20, and .30 represent small, medium, and large validities.

*Fakability* concerns whether predictor scores are affected by intentional response distortion. Given that most applicants want to score high enough to receive a job offer, it is possible that applicants will intentionally provide more favorable reports of self-reported past behavior, accomplishments, personality, and work ethic and values, than they might if there were no consequences for their test scores. Hence, in high-stakes testing environments where the outcome is important, one must be critical of whether the scores truly reflect the latent underlying KSAOs or impression management. Indeed, nearly every study comparing high-stakes (e.g., applicants) to low-stakes (e.g., job incumbents) testing situations finds higher scores on personality (and related self-report measures) in high-stakes contexts (Smith & Robie, 2004). It appears the criterion-related validity can be cut in half as a result of response distortion (Hough, 1998). Although there is some disagreement about whether faking affects the validity or practical usefulness of non-cognitive measures, it is a very important concern for practitioners because inflated test scores reduce the distinctions between applicants.

*Subgroup differences* indicate how much of a mean difference in predictor scores exists between subsamples within an overall sample. Normally, the subgroups are those representing protected groups through employment regulations; common examples include race, ethnicity, sex, and age. When there are mean differences between subgroups, and organizations hire the top-scoring applicants first,

the consequence is that members from the higher scoring subgroup will be hired more frequently than those from the lower scoring subgroup. The larger the subgroup difference, the more likely it is that higher scoring subgroup members will be hired. For example, physical ability tests have large differences between men and women, so in a selection process with even a moderate selection ratio (e.g., hire 30% of the applicant pool), men could be hired to the near total exclusion of women. Subgroup differences are usually expressed in terms of standardized mean differences (denoted as  $d$  values), so any score differences are conveniently expressed in standard deviation units. Please note that subgroup differences on predictor scores do not mean that the predictor is biased or discriminatory against the lower scoring group. Two groups may have a sizable mean difference, but if the test predicts equally well for both groups, there is no differential prediction (Cleary, 1968; Schmidt, 1988). Hence, subgroup differences are not (by themselves) proof of discrimination.

*Economic return* indicates whether the value provided by a predictor or selection system outweighs the costs of development and/or implementation. Every predictor has an associated cost that includes development or purchase, administration, and scoring. Among personnel selection scholars, utility analysis is often used as a means to convey economic return. There are numerous models of utility analysis (see Boudreau & Ramstad, 2003), but they share some common features in that they attempt to show how much financial improvement will occur from implementing a selection system with a given amount of validity, selection ratio, and related factors. Reports of some studies show these financial improvements to be quite large (Schmidt, Hunter, McKenzie, & Muldrow, 1979). However, many practitioners are not persuaded by utility estimates, particularly when they do not have HR or I/O backgrounds (e.g., Latham & Whyte, 1994). This is likely due to the fact that utility analysis is complex, based on many assumptions, and ultimately an estimate of a unit-level consequence of individual-level selection (Schneider, Smith, & Sipe, 2000). There are many other selection metrics that practitioners attend to, including job performance, retention, time-to-hire, cost-per-hire, and productivity-per-employee. Different firms have different metrics, but whatever metrics are used, economic return is paramount in practice.

*User acceptability* refers to whether the various stakeholders who are affected by the selection process believe (or at least accept) the results. It is

sometimes called face validity, but face validity is more narrow in that it is defined as whether people that believe the predictor measures what it claims to measure. User acceptability is a broader term because even though a predictor may be face valid, it still may not be acceptable because of other factors (e.g., cost, appearance, subgroup differences). Note that user acceptability is an inclusive term that includes managers, HR personnel, and the applicants (Kehoe, Mol, & Anderson, 2009). Each of these stakeholders defines acceptability differently. Line managers want a short, quick, and accurate system. HR personnel want a cost-effective, legally defensible, and sustainable system. Applicants want a short, quick, and fair system and, although they may often say that they want it to be accurate, in reality they likely want it to be one in which they do well. These are frequently different and competing perspectives, and it is a tension that must be balanced. For example, managers will sometimes choose a less valid system over a more valid selection for many reasons, including cost, unacceptable negative effects on diversity, or political battles between HR and management. At least in the United States, there is no law that requires a private firm to use a valid selection system unless it is discriminatory against a protected subgroup (Pyburn, Ployhart, & Kravitz, 2008), so manager beliefs have a strong influence on what gets implemented.

In the next sections, the criterion-related validity, fakability, subgroup differences, economic return, and user acceptability framework is used to help critically compare and contrast research on predictor KSAO constructs and methods. Table 8.1 provides an overview and summary.

### Predictor KSAO Constructs

There are five major domains of predictor constructs: cognitive, knowledge, personality, values/needs//interests, and physical abilities. These represent the broad spectrum of individual differences of most relevance to personnel selection. Ryan and Sackett in chapter 5 of this handbook discusses these individual differences in more detail, so here the literature is summarized more specifically to personnel selection practices.

#### Cognitive

There has been extensive research that seeks to understand the latent structure of cognitive ability (also known as general mental ability, or  $g$ ), and several different models exist (see Drasgow, 2003). However, most share in common a hierarchical

**Table 8.1. Framework for Comparing Predictor Constructs and Methods (adapted from Ployhart et al., 2006)**

Predictors	Dimension				
	Criterion-Related Validity	Fakability	Subgroup Differences (Race/Gender)	Economic Return	User Acceptability
Effect Sizes	<i>Correlation</i> .10 = Low .20 = Moderate .30 = High	N/A	<i>d</i> value .20 = Small .50 = Moderate .80 = Large	N/A	N/A
			Predictor Constructs		
<i>Cognitive</i>					
Cognitive Ability	High	Low	Large/Small	High	Moderate
Job Knowledge	High	Low	Moderate/Small	High	Favorable
<i>Non-cognitive</i>					
Personality	Low	High	Small/Small	Moderate	Unfavorable
Values/Needs/Interests	Low	Moderate to High	Small/Small	Low	Unfavorable
<i>Physical</i>					
Psychomotor & Physical Ability	High	Low	Small/Large	Moderate	Moderate
			Predictor Methods		
Experience & Biodata	Moderate	Moderate	Small/Small	Low to Moderate	Moderate to Unfavorable
Interview-Structured	High	Moderate	Small/Small	Moderate	Moderate
Interview-Unstructured	Low	Moderate to High	Small/Small	Low	Moderate
Situational Judgment	Moderate	Moderate	Moderate/Small	High	Favorable
Assessment Centers	High	Low to Moderate	Small to Moderate/Small	Moderate	Favorable
Work Samples	High	Low	Small to Large /Unknown	Moderate	Favorable

Note: Estimates for validity and subgroup differences are based on uncorrected values.

structure, such that specific abilities (e.g., verbal, numerical, and reasoning ability) are subsumed within general cognitive ability (operationalized as the first factor when conducting a factor analysis on the more specific abilities). Carroll's (1993) hierarchical model is perhaps the most comprehensive, which posits that general cognitive ability subsumes eight more specific abilities (e.g., fluid intelligence,

crystallized intelligence, general memory and learning), which in turn subsume even more specific abilities (e.g., speed of reasoning, reading comprehension, memory span). Several studies have demonstrated that the specific abilities do not predict broad criteria better than general cognitive ability (Ree, Earles, & Teachout, 1994). Hence, because general cognitive ability is more predictive and

more efficient than administering a longer battery of specific abilities, most personnel selection practitioners use general cognitive ability.

Cognitive ability represents one of the most important individual differences in employment contexts because it is one of the strongest predictors of performance across most jobs (Schmidt & Hunter, 1998). Further, as the cognitive demands of the job increase, generally so does the criterion-related validity of cognitive ability (Hunter & Hunter, 1984). Given the increased prevalence of knowledge work, the importance of cognitive ability may become even greater as the modern economy evolves. Cognitive ability tests have no concerns about faking because the answers are either correct or incorrect. Cognitive ability also tends to produce at least moderate acceptance from users and high economic return because measures of general ability are relatively inexpensive and easy to administer.

However, general cognitive ability also manifests some of the largest racio-ethnic subgroup differences of any predictor construct or method. Relative to whites, blacks score up to a full standard deviation lower and Hispanics score from half to over three-quarters of a standard deviation lower (Asians score about one-fifth of a standard deviation higher than whites; see Roth, Bevier, Bobko, Switzer, & Tyler, 2001). These differences are large enough that, under typical selection ratios, the hiring rate of blacks and oftentimes Hispanics will be considerably lower than that of whites (Sackett & Wilk, 1994). Sex differences tend to be small for general cognitive ability. Racio-ethnic and sex differences are more variable for specific abilities, and in some cases sex differences can be quite a bit larger (Hough, Oswald, & Ployhart, 2001). For example, men score higher than women on quantitative abilities, while women score higher than men on verbal abilities.

The fact that cognitive ability produces high criterion-related validity but also large racio-ethnic subgroup differences puts practitioners in what has been called the validity-diversity dilemma (Pyburn et al., 2008). Because of the negative effect that cognitive ability will have on the hiring rates of many racio-ethnic minority candidates, there has been enormous effort devoted to trying to understand and reduce racio-ethnic subgroup differences (see Becker, Schmitt, Ellingson, & Kabin, 2001). This is critical because organizational decision makers frequently shy away from using (or fully weighting) cognitive ability tests due to their negative effects on diversity (Pyburn et al., 2008); this also explains

why the user acceptance of cognitive ability is only moderate. The size and pervasiveness of racio-ethnic subgroup differences make them extremely difficult to eliminate, so the goal has been to find strategies and approaches for reducing them. Ployhart and Holtz (2008) identified 16 different strategies that have been offered to reduce racio-ethnic and sex subgroup differences and found that the only strategy that would meaningfully reduce subgroup differences and not substantially reduce criterion-related validity was to supplement a cognitive predictor with non-cognitive predictors that have smaller subgroup differences (see Bobko, Roth, & Pototsky, 1999; Sackett & Ellingson, 1997; Sackett & Roth, 1996). One thing that is important to realize is that even though cognitive ability tests produce large racio-ethnic subgroup differences, they are not biased toward any minority candidates because the criterion-related validity is not moderated by racio-ethnic group membership (Schmidt, 1988). Excellent reviews and debates of the cognitive ability-testing literature can be found in Drasgow (2003) and a special issue of *Human Performance* (2002).

### **Knowledge**

Whereas cognitive ability is generic to different work situations, knowledge is domain specific because it is acquired through experience or education. Examples include knowledge of accounting, finance, human resources, and selection practices. Knowledge may be broadly construed into two main types: *declarative* is knowledge of facts and principles, while *procedural* is knowledge of processes and how to apply facts and principles to a given problem. Knowledge is not fixed or constant but accumulates over time (Kanfer & Ackerman, 2004). Those with more experience and greater cognitive ability will be more likely to acquire knowledge; thus knowledge partially mediates the relationship between cognitive ability and performance (Hunter, 1983). Knowledge may be generic to a specific occupation, or specific to a particular firm, although in selection we are most typically interested in occupation or job-specific knowledge (as opposed to firm-specific knowledge).

Job knowledge predictors tend to have high criterion-related validity (Schmidt & Hunter, 1998), rivaling that of general cognitive ability (but of course being job or domain specific). A benefit of knowledge tests being domain specific is that they are clearly face valid and have favorable user acceptability. Because they are easy to administer and

relatively simple to develop, they also offer high economic return. On the other hand, knowledge tests have moderate racio-ethnic subgroup differences (whites score about half a standard deviation higher; Roth, Huffcutt, & Bobko, 2003). Sex differences are small, although they may be larger if the knowledge overlaps with abilities that manifest sex subgroup differences (e.g., verbal ability and a knowledge test of English). It is also possible that knowledge tests may suffer from faking, but this is not much of an issue because knowledge tests have right/wrong answers.

### **Personality**

Most selection research conducted over the last 20 years has been based on minor variations of the Five Factor Model (FFM). The FFM is composed of the following traits: emotional stability (sometimes known as neuroticism), extraversion, openness to experience, agreeableness, and conscientiousness. As with cognitive ability, the FFM is hierarchical in that these five traits are the broadest and subsume various more specific traits within them (e.g., conscientiousness subsumes dutifulness, achievement, and dependability). Although other structures of personality exist, the FFM has dominated selection research and practice.

Personality testing within selection contexts has traveled a bumpy road (for a historical review, see Schneider, 2007). Early summaries of criterion-related validity reached rather pessimistic conclusions about their usefulness (Ghiselli, 1966; Guion & Gottier, 1965). Meta-analyses conducted 25 years later found highly similar uncorrected criterion-related validities, but reached the opposite conclusion and led to a "rebirth" of personality testing (Barrick & Mount, 1991). Schmitt (2004) noted that for the most part, uncorrected personality criterion-related validities have not changed much over time and tend to be moderate to low. For example, conscientiousness has the largest criterion-related validity of the FFM constructs, but this uncorrected estimate is only .13 (Barrick & Mount, 1991). It is apparent that personality testing now sees widespread use in practice because even though the validities are low, they are not zero and hence contribute to the prediction of performance in modest ways. But many have come to question whether, given such validities, personality tests should be used. An exchange of opinions on this topic may be found in Morgeson et al. (2007a, 2007b), Ones, Dilchert, Viswesvaran, and Judge (2007), and Tett and Christiansen (2007).

Further compounding the frustrations with personality testing is the likelihood of applicants faking their responses to increase the probability of scoring well. Unlike cognitive ability and job knowledge, personality tests do not have correct answers, and it is relatively easy to misrepresent oneself on most self-report measures. As with personality test validity, there is not strong consensus over whether faking is a problem. Some argue that faking will not negate the criterion-related validity of personality tests (Hough & Furnham, 2003; Ones & Viswesvaran, 1998; Ones, Viswesvaran, & Reiss, 1996), even though the validity may be reduced by up to half (Hough, 1998). It is nearly ubiquitous that personality tests administered to applicants will have higher mean scores than when the same tests are completed by incumbents (Smith & Robie, 2004). This by itself may not affect validity too much, but it will often affect one's ability to distinguish between candidates because of tie scores due to ceiling effects. Hence, user acceptability of personality tests tends to be mixed because applicants and organizational decisions makers often believe that the scores are rendered incomprehensible due to faking. Faking and low validities also lead to personality tests having low economic return.

Given such a discouraging review, one might wonder why personality testing has had a rebirth in practice. The reason likely has to do with the fact that personality constructs have relatively small racio-ethnic and sex subgroup differences (although the differences are larger for facets relative to FFM constructs; Foldes, Duehr, & Ones, 2008). Hence, including personality along with a cognitive ability test will generally enhance validity and (to a degree) reduce racio-ethnic subgroup differences (Sackett & Roth, 1996). It is also important to realize that there are strategies that can be implemented to increase the criterion-related validity of personality tests. In particular, the use of personality composites developed to maximally predict performance for certain occupations (e.g., managerial jobs, customer service; Hough & Schneider, 1996) or tailoring the personality items to specific contexts (Lievens, DeCorte, & Schollaert, 2008) can manifest validities larger than the FFM constructs. Similarly, linking specific traits to narrow criteria can increase validity (Dudley, Orvis, Lebiecki, & Cortina, 2006). New approaches to personality measurement have also been proposed to help reduce faking and enhance validity (e.g., James, 1998). In commonsense terms, it is hard to believe that personality does not affect job performance; the challenge for I/O psychologists

is how to better measure it. For more information on personality at work, please see edited books by Barrick and Ryan (2003) and Schneider and Smith (2004), as well as reviews by Hough and Furnham (2003) and Hough and Oswald (2008).

### **Values, Needs, and Interests**

A person's values or needs represent his or her latent styles, preferences, or desires (Dawis & Lofquist, 1984). As such, they direct attention and motivation without the individual often being aware of them. Interests represent a person's preferences for certain types of work. Holland's RIASEC model (1997) has been the most influential and is part of the O\*NET system. The RIASEC model suggests that people choose occupations that match their values, needs, and interests. The six dimensions in the RIASEC model include *realistic* (conforming), *investigative* (analytic), *artistic* (open), *social* (gregarious), *enterprising* (ambitious), and *conventional* (obedient). Together, values, needs, and interests influence the environments that individuals will prefer to enter, and the satisfaction that they will derive from different environments. For example, a person with a high need for achievement might be more attracted to firms that require and reward achievement than firms that emphasize equality among employees.

Values, needs, and interests have rather relatively low criterion-related validity for job performance criteria. Rather, values, needs, and interests can have a strong influence on the perceived fit that one has with a job and organization. *Perceived fit* is a person's perception of whether his/her KSAOs, values, needs, or interests match the work environment. There are different kinds of fit, including a perceived match between the person's values and those of the job (person-job fit) and organization (person-organization), as well as whether the person's KSAOs match those required and rewarded by the job (needs-supplies; Cable & DeRue, 2002). Perceived fit, in turn, is predictive of job/firm attraction, the satisfaction one derives from a job, and retention (Cable & Parsons, 2001; Kristof-Brown, Zimmerman, & Johnson, 2005). However, a recent meta-analysis found that fit was a rather weak predictor of job performance (Arthur, Bell, Villado, & Doverspike, 2006). Like personality tests, measures of values, needs, and interests are moderately to highly fakable and show low economic return. Values, needs, and interests are often perceived negatively by applicants, but value measures are often desired by employers who wish to determine whether an applicant will

fit within the organization (particularly for entry-level managerial jobs in which the candidate will be groomed for higher level positions). These measures have small ratio-ethnic and sex subgroup differences. Although this summary does not offer much empirical enthusiasm for using measures of values, needs, and interests as a basis for hiring decisions, the relationships that these constructs have with job choice and retention make them important parts of the broader staffing context. For a broad review of the fit literature, see Edwards (2008) and an edited book by Ostroff and Judge (2007).

### **Psychomotor and Physical Abilities**

Whereas most jobs in the older economy were based on physical ability, most jobs in the modern economy are based on mental ability. This might explain why there has not been much new research on physical ability testing in I/O psychology over the last 20 years (the other explanation is that this research has shifted to other disciplines, such as human factors). In the I/O literature, most research has focused on two main types. The first type is psychomotor abilities that include sensory abilities (e.g., vision) and dexterity. The second type is physical abilities that include muscular strength, cardiovascular endurance, and movement quality (Hogan, 1991).

Physical abilities have high criterion-related validity and low fakability (no wonder; it would be impossible to fake a good bench press!). Economic return is moderate (because of the cost of testing; it must be hands-on). Both applicants and organizational decision makers tend to find these tests at least moderately acceptable. On the other hand, the sex subgroup differences can be large, ranging from one standard deviation (for muscular endurance) to over two standard deviations (for muscular power) favoring males (Ployhart & Holtz, 2008).

### **Predictor Methods**

There are five major types of predictor methods: experience and biographical data (biodata), interviews, situational judgment tests, assessment centers, and work samples. As a caveat to the review that follows, one should realize that, by definition, predictor methods can assess multiple constructs. Therefore, even though the criterion-related validity, fakability, subgroup differences, economic return, and user acceptability of predictor methods are discussed, one should realize that these findings average across different KSAO constructs. The conclusions presented below may differ depending

on the specific construct being assessed (Arthur & Villado, 2008). In general, predictor methods that assess or are more strongly related to cognitive ability will show higher criterion-related validity and larger racio-ethnic subgroup differences. This latter finding has been recognized as far back as Spearman (1927), who noted that the stronger the cognitive loading of an assessment, the larger the black-white subgroup difference. Please keep this caveat in mind when reviewing the following summary.

### ***Experience and Biographical Data (Biodata)***

Predictors based on experience rely on the adage that “the best predictor of future behavior is past behavior” (Wernimont & Campbell, 1968). Experience is a broad, multidimensional construct that often serves as a proxy for knowledge. There are many different kinds of experience, such as experience with an occupation, a job within a firm, or employment within a given firm. For example, a person might have 15 years of experience at the same firm, but 10 years of experience as a software designer and five years as a manager. Experience may range from generic to firm-specific, although this distinction has had much more attention in economics (e.g., Becker, 1964) than it has had in I/O psychology. Experience may further be assessed by amount, time, or type (Quiñones, Ford, & Teachout, 1995; Tesluk & Jacobs, 1998). Biographical data (biodata) is a more structured way to assess experience. With biodata, one usually starts by identifying the kinds of developmental experiences and/or constructs that are necessary for success on the job. Items are then written to measure these experiences and/or constructs. The end result is a measurement instrument that is structured to assess specific types of experience relevant to a specific job. Mael (1991) and Russell (1994) are useful sources for how to generate biodata items.

Experience and biodata-based predictors have moderate validity, with biodata having better uncorrected criterion-related validity than simpler measures of experience. Both tend to have small racio-ethnic and sex subgroup differences, although the more the items represent cognitively oriented variables, such as educational degrees or academic achievements, the more likely it is that there may be racio-ethnic differences (Berry, Gruys, & Sackett, 2006). Experience and biodata forms may be fakable; however, faking can be reduced through the use of warnings, using verifiable items, and asking participants to elaborate on their answers (e.g., if the

question asks how many languages a person speaks, require the applicant to elaborate and list the languages; Schmitt, Oswald, Kim, Gillespie, Ramsay, & Yoo, 2003). Experience and biodata vary in terms of their acceptability; most people believe that experience is related to job performance but the nature of biodata forms sometimes leads to less acceptance. Experience and biodata offer low to moderate economic return, depending on how strongly related they are to job performance. An edited book on biodata by Stokes, Mumford, and Owens (1994) provides considerably more detail.

### ***Interviews***

Interviews are one of the most frequently used selection methods (Ulrich & Trumbo, 1965) and have been the focus of research for decades (for a recent review, see Posthuma, Morgeson, & Campion, 2002; see also an edited book by Eder & Harris, 1999). Interviews have different purposes; for example, there are screening interviews used to make hiring decisions, and recruiting or informational interviews, which serve as recruiting tools (for present purposes, we will focus primarily on screening interviews). Interviews are considered a predictor method because there are multiple ways to structure interviews and to ask the questions, and thus there are multiple constructs that can be assessed. *Interview structure* refers to how the interview is developed, administered, and scored in order to enhance construct validity and efficiency. Campion, Palmer, and Campion (1997) identified 15 different elements of structure, which include basing the interview on a job analysis, asking the same questions of all applicants, minimizing the use of extraneous information, making detailed notes, and training the interviewers. A related aspect of structure is whether the interview is conducted by a single person or a panel composed of multiple interviewers. In terms of *interview questions*, there are three broad types, which include behavior (asking about past behavior), situational (asking about future behavior or hypothetical situations), or job knowledge (asking about how to perform specific job tasks). Finally, as one might imagine, different questions may be asked to measure different *constructs*. A meta-analysis by Huffcutt, Conway, Roth, and Stone (2001) found that the most common constructs assessed by interviews were personality and social skills. Other commonly assessed constructs included cognitive ability, knowledge and skill, interests and preferences, fit, and physical abilities and attributes.

In general, the more structured the interview, the higher the criterion-related validity and reliability (Cortina, Goldstein, Payne, Davison, & Gilliland, 2000). There is some evidence that behavioral interviews have higher validity than situational interviews (Huffcutt, Weekley, Wiesner, Degroot, & Jones, 2001). It appears that one of the main reasons that unstructured interviews have lower validity is because they have less reliability; but using at least four independent interviewers for each applicant can help offset this limitation (Schmidt & Zimmerman, 2004). Interviews tend to receive moderate acceptance from applicants and managers, and they have moderate economic return. They also show relatively small raccio-ethnic and sex subgroup differences, with the differences being smaller for structured interviews. The content and structure of an interview can influence the prevalence of applicant impression management (McFarland, Ryan, & Kriska, 2003), and unstructured interviews are more fakable than unstructured interviews.

### **Situational Judgment Tests (SJT)**

Situational judgment tests (SJT) are sometimes considered "low-fidelity simulations" because they present applicants with a brief work-related situation, and then ask how they should or would respond to the situation (Motowidlo, Dunnette, & Carter, 1990). The situations are usually derived from the critical incident technique and represent real situations that are fairly common, but difficult and challenging to address. The situations will not have a clearly correct answer, so they create variability in test-taker responses. In terms of responses, SJTs usually present four to six behavioral options that indicate what a person should or would do. Like interviews, SJTs are considered predictor methods because they can be used to measure a variety of constructs. However, SJTs do not measure constructs in a homogenous manner, and it is unclear whether an SJT can be developed to do so (Schmitt & Chan, 2006). SJTs tend to most commonly measure some composite of cognitive ability, personality, experience, and knowledge (McDaniel, Morgeson, Finnegan, Braverman, & Campion, 2001; McDaniel & Nguyen, 2001). The construct validity of SJTs is in part affected by the wording of instructions. When the instructions ask what you "should do," the SJT is more strongly related to cognitive ability; when the instructions ask what you "would do," the SJT is more strongly related to personality (McDaniel, Hartman, Whetzel, & Grubb, 2007).

The criterion-related validity of SJTs is moderate and, although related to cognitive ability and personality, still shows incremental validity over these other predictors (e.g., Clevenger, Pereira, Wiechmann, Schmitt, & Harvey, 2001). SJTs tend to be viewed favorably by applicants and managers because they are face valid and are tied to specific parts of the job (see Bauer & Truxillo, 2006). They offer reasonably high economic return because they can be administered in an efficient manner and have moderate validity. Sex subgroup differences are small, but raccio-ethnic subgroup differences tend to be small to moderate (see Ployhart & Holtz, 2008). Raccio-ethnic differences are largely due to the reading requirements and cognitive loading of an SJT; the more the SJT requires reading ability, the larger the subgroup difference. Some research has found that faking may reduce the criterion-related validity of SJTs (Peeters & Lievens, 2005), although considerably more needs to be learned to understand faking effects with SJTs. An edited book by Weekley and Ployhart (2006) offers more detailed information about SJTs.

### **Assessment Centers**

Assessment centers are predictor methods that present candidates with different *exercises* designed to assess a variety of KSAOs. Assessment centers are not known by any particular exercise, but rather a collection of exercises. The most commonly used exercises include role plays, leaderless group discussions, in-basket exercises that stress setting priorities and delegation, business games and simulations, and mock presentations (Spychalski, Quiñones, Gaugler, & Pohley, 1997). Most assessment centers use an average of four to five exercises (Woehr & Arthur, 2003). These exercises measure of variety of different KSAO *dimensions*, with the average being 10 to 11. The most common dimensions include social skills, communication, motivation, persuasion and influence, organization and planning, and problem solving (Arthur, Day, McNelly, & Edens, 2003). In most assessment centers, trained raters will observe or grade the applicant's performance on the dimensions.

Assessment centers have high criterion-related validity because they usually require one to actually perform the behavior (instead of a self-report of the behavior), and they measure a variety of different KSAO dimensions linked to performance on the job. Likewise, they are difficult to fake because they involve the actual performance of the target behaviors. Sex subgroup differences are small, but

ratio-ethnic subgroup differences are small to moderate, depending on the construct and exercise. As the relationship between the exercise and cognitive ability increases, so does the ratio-ethnic subgroup difference (Goldstein, Yusko, Braverman, Smith, & Chung, 1998; Goldstein, Yusko, & Nicolopoulos, 2001). Economic return is moderate because of the cost to develop and administer the assessment center. Both applicants and managers tend to perceive assessment centers favorably, although the administrative costs can sometimes make them less favorable to managers.

Most of the empirical data provide good support for the use of assessment centers. However, their biggest limitation is that they often show little construct validity (an issue shared with SJTs). Specifically, it is nearly always the case that assessment centers manifest exercise factors and not dimension (construct) factors (Klimoski & Brickner, 1987; Lance, Lambert, Gewin, Lievens, & Conway, 2004). The reason for this finding appears to be due to the nature of applicants themselves: construct validity is only likely to be found when candidates manifest high consistency across exercises *and* high differentiation across dimensions (Lievens, 2002). This seems rather unlikely in practice. Research continues to explore and debate the construct validity of assessment centers (see Lance, 2008).

### **Work Samples**

As their name implies, work samples are assessments that are based on actual samples of behavior and tasks performed on the job. Common examples include flight simulators for pilots, welding tests for welders, and, in many ways, a doctoral dissertation for Ph.D. students. Like SJTs and assessment centers, work samples are predictor methods that can measure a variety of KSAO constructs. Much of their popularity is owed to an article by Wernimont and Campbell (1968) that argued for prediction using "samples" of behavior. That is, they suggested that, in contrast to self-report measures that offer "signs" of an applicant's potential, it is more productive to evaluate actual samples of behavior. In general, the higher the psychological and physical fidelity of a predictor, the greater its validity.

Because there is such a high physical and psychological overlap between the job and the assessment, work samples have high criterion-related validity. Further, because they represent actual job behaviors and tasks, they have high acceptance from applicants and organizational decision makers. It is nearly impossible to fake work samples because applicants

need to actually perform the relevant behaviors. Economic return on work samples is only moderate because of the costs to develop and administer them. The size of subgroup differences is larger than most prior research had indicated, however. Roth, Bobko, McFarland, and Buster (2008) found that ratio-ethnic subgroup differences range from small to large, depending on the nature of the constructs assessed. These latter findings raise some concern over whether work samples are an effective means to address the diversity-validity dilemma. It is difficult to report the sex subgroup difference; some studies report small differences and others report larger differences, so at this point it is not appropriate to make even a general speculation.

### **Measurement Methods**

Thus far we have considered predictor constructs and methods, but it is important to realize that even predictor constructs are always assessed by a particular measurement method (Arthur & Villado, 2008). For example, we speak of cognitive ability tests as though they are always measured using paper-and-pencil formats, but in practice they may be assessed by a variety of methods including other paper formats, the Internet, and even interviews. Hence, it bears repeating that every predictor is associated with a measurement method. If one collapses across constructs, measurement methods will differ in their reliability, cost, customization, user acceptability, and sources of contamination and deficiency. Ployhart et al. (2006) review these issues in some detail, but to summarize, the major types of assessment methods include paper-and-pencil (written) assessments, Internet/computerized assessments, oral/aural assessments (e.g., phone interviews or video-based testing), and rating/judgment assessments. There is little research that focuses on comparing methods of assessment, but that which has been conducted finds (not surprisingly) that methods are important. For example, Chan and Schmitt (1997) found that video-based assessments of SJTs had smaller black-white subgroup differences than written SJTs (due to lower reading requirements). Currently unknown is an understanding of the *ideal* measurement method for different selection goals, purposes, and constructs.

### **View From the Top of the Mountain: Does Anybody Know We Are Here?**

This concise review illustrates that personnel selection is a vibrant area of research. More direct evidence is provided by Cascio and Aguinis (2008b),

who find that topics in personnel selection rank in the top five most-published articles in the *Journal of Applied Psychology* and *Personnel Psychology* over the last several decades. Personnel selection practices have become highly refined, and the technology underlying the personnel selection process is sophisticated. As a profession, selection scholars sit atop a mountain of achievement and understanding. Yet it is lonely at the top. There have been concerns raised over whether personnel selection as a profession of scientists and practitioners is having an impact on, quite bluntly, the real world. These concerns come from several perspectives and sources. Rynes, Brown, and Colbert (2002) surveyed actual managers and found that the HR area of selection showed the largest discrepancies between research findings versus manager beliefs. Cascio and Aguinis (2008a, 2008b) argued that continued emphasis on the usual personnel selection topics is somewhat disconnected from broader industry trends and emphasis on human capital, and that personnel selection must become more contextualized if it is to become more relevant. Ployhart (2006) noted that personnel selection is missing a golden opportunity to demonstrate its strategic value, and that continued focus on research disconnected from practical concerns in private industry limits the relevance of our science. Highhouse (2008) suggested that we need to understand the beliefs, judgments, and decision making processes of managers who must approve and live with selection systems. Finally, Anderson (2005) discussed the very different worlds in which academics and practitioners live and work.

What is at issue here is not whether we have selection processes and practices that improve hiring decisions. On this point, there is an abundance of evidence to suggest that we do (Schmidt & Hunter, 1998, and the research reviewed above). Instead, what is at issue is whether these practices are implemented to the extent that they should be. Highhouse (2008) is almost certainly correct that manager beliefs have a huge influence on the types of selection practices that are implemented. But one could also legitimately argue that personnel selection has not done much to show its value to firm-level outcomes—that is, show its value to *the business* (Ployhart, 2006; Schneider et al., 2000). Utility analysis represents an attempt to show the value of personnel selection practices, and while there are many sophisticated utility analysis methods and models (see Boudreau & Ramstad, 2003; Cascio 2000), it has not had much *widespread* impact on conveying the importance of our research

to those outside academia. Indeed, if it had, why are useful selection practices not more frequently implemented?

There is no point in quibbling over the value of utility analysis; it has a particular purpose and is valuable as a tool for understanding the costs and economic return of HR practices. But continued focus on utility analysis, ways to increase validity through statistical corrections, or the search for the “ultimate predictor” are unlikely to increase the use of valid selection practices in private industry. Addressing this issue will require a fairly radical extension of how we define personnel selection research and practice. It will involve broadening Figure 8.1 to include other criteria and processes that extend beyond the individual level, and it will require becoming familiar with the strategic management literature. We cannot achieve the vision of “Ensuring Sustainable Organizational Effectiveness Through the Acquisition of Human Capital” if we continue our current trajectory.

### The Future: Research to Ensure Sustainable Organizational Effectiveness Through the Acquisition of Human Capital

If one adopts the vision of personnel selection proposed above, then it is clear that selection scholars and practitioners have a lot of work to do because we do not know much about the “ensuring sustainable organizational effectiveness” part of the vision—the very part which is likely of most interest to organizational decision makers. In the subsections that follow, I propose five missions through which selection research should be conducted to achieve this vision.

#### *Mission #1: Show the Business-Unit Impact of Personnel Selection*

There is only limited theory and empirical data suggesting that personnel selection actually improves business-unit effectiveness. Utility analysis is an attempt to estimate a cross-level relationship between individual level staffing and unit effectiveness, but it is only an estimate and is often based on questionable assumptions (Schneider et al., 2000). The search for more direct empirical evidence is discouraging. One of the most highly cited studies that is used to support the organizational-level impact of selection, Terpstra and Rozell (1993), surveyed managers and had them self-report *both* the type of selection practices they used and their opinion of their firm's effectiveness. Such common-source, common-method research would never be

published today, but it remains an article highly cited as “proof” that selection matters. More compelling evidence is offered by scholars conducting research in the area of strategic human resource management (SHRM). These studies have found that business units using valid selection systems performed better than those that did not (Hatch & Dyer, 2004; Huselid, 1995; Koch & McGrath, 1996; MacDuffie, 1995). Manager self-reports are usually the way in which selection system variables are measured. These studies are helpful in that they show that use of selection practices contribute to unit effectiveness, but they are limited in that they rely on manager self-reports, and they do not actually show how selection relates to unit effectiveness (see Becker & Huselid, 2006, who discuss this issue more generally with HR).

Schneider et al. (2000) made the point that selection researchers need to start demonstrating how the selection of individuals contributes to better functioning organizations. Subsequent multilevel theory has provided direction for how to conceptualize and link individual- and business-unit level theory and research on human capital and selection (Ployhart & Schneider, 2002, 2005). Ployhart (2004, 2006) proposed a multilevel model of selection that ties individual KSAOs to unit-level human capital and unit effectiveness. Figure 8.2 is adapted from Ployhart and Weekley (2010) and provides a simple overview of this model. First, notice that the business unit’s strategy influences the types of selection practices used (there is more on strategy in the next section; see also Snow and Snell, chapter 30 of this handbook). The types of selection practices used (e.g., cognitive ability tests, interviews) will impact the types of KSAOs acquired and retained. Systematic selection on job-related KSAOs contributes to the formation of unit-level human capital constructs; this process is known as *human capital emergence*. When unit-level human capital is related to unit effectiveness (because it is consistent with the firm’s strategic direction), human capital contributes to sustained competitive advantage and hence offers *human capital advantage*. Notice that this approach is quite different from the usual selection approach that is contained entirely within the individual level of analysis (see Ployhart, 2004; 2006).

There is some empirical research that supports key relationships in Figure 8.2. First, research finds that KSAOs are in fact nested within business units and hence emerge to form unit-level human capital (Ployhart, Weekley, & Baughman, 2006; Schaubroeck, Ganster, & Jones, 1998; Schneider,

Smith, Taylor, & Fleenor, 1998). Second, research has found evidence for human capital advantage. Van Iddekinge et al. (2009) found that units that had a greater percentage of employees who passed a validated selection test had greater customer service performance, retention, and financial success. Ployhart, Weekley, and Ramsey (2009) found that individual differences in service orientation not only created unit-level service orientation (human capital), but that this human capital also had a positive effect on the unit’s financial effectiveness. While this unit-level relationship helps to support the *unit-level financial impact* of personnel selection, it also highlighted some important differences between micro and macro human capital. In particular, the relationship at the unit level was nonlinear over time and heterogeneous across units, suggesting that the stable relationships found at the individual level may not be so stable at the unit level. DeNisi, Hitt, and Jackson (2003) asked, “...if hiring ‘better’ people results in higher productivity, how exactly does the selection of individuals translate into improved organizational performance?” (p. 12). One answer is through human capital emergence and human capital advantage (see Figure 8.2).

Future research needs to expand our understanding of how selection on individual differences contributes to better functioning organizations. We need to move beyond the individual level and utility analysis to understand how selection impacts higher level criteria. The field of SHRM has grown rapidly and arguably had an important impact because its key dependent variables are firm or business unit-level outcomes that organizational decision makers care about (Gerhart, 2005). As an applied science, selection researchers and practitioners should care about them too.

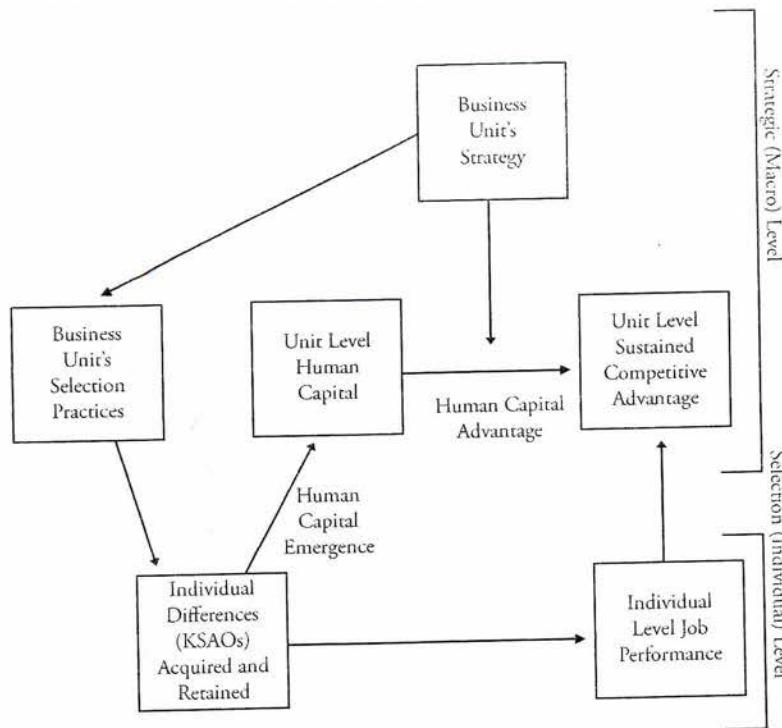
### **Mission #2: Match Selection to the Firm’s Strategy**

The field of strategic management focuses on explaining between-firm differences in effectiveness (see Snow and Snell, chapter 30 of this handbook, for considerably more detail about strategic human resource management). It addresses the very lofty goal of how to compete against other firms. Given that personnel selection should contribute to this lofty goal, one would hope that there is some consideration of selection as a part of a firm’s strategy. Yet there is almost no reference to the strategic management literature in selection scholarship, with the possible exception of SHRM scholarship. The firm’s strategy should drive most operational decisions,

Figu

incl  
& S  
valu  
like  
fore  
nec

J  
in s  
bet  
As  
det  
use  
uni  
Th  
ind  
at  
hui  
cifi  
tio  
cor  
(e:  
wh  
(e:  
ser  
ilar  
ual



**Figure 8.2** A Multilevel Framework for Personnel Selection (adapted from Ployhart & Weekley, 2010).

including choice of HR practices (Wright, Dunford, & Snell, 2001). To the extent that HR practices offer value to the firm's strategy, the practices are more likely to be implemented and supported. It therefore seems that researchers might seek to better connect the ties between firm strategy and selection.

Ployhart and Weekley (2010) discuss this issue in some detail, and Figure 8.2 highlights the link between strategy and selection on a superficial level. As the figure shows, the firm's strategy not only determines the nature of HR practices that will be used but also moderates the relationship between unit-level human capital and unit effectiveness. That is a critically important point because, unlike individual-level validity generalization research, at the unit level one would expect the effect of human capital on effectiveness to be context specific (Barney, 1991). For example, service orientation human capital may be critical for a firm whose competitive advantage is based on customer service (e.g., Neiman Marcus), but much less so for a firm whose competitive advantage is based on efficiency (e.g., Wal-Mart). Note that in both companies, a service orientation predictor would likely have similar levels of criterion-related validity for individual job performance. Thus, because organizational

decision makers seek to support the firm's strategy, research should examine how personnel selection can support strategy and reinforce strategy.

### ***Mission #3: Demonstrate the Sustainability of Personnel Selection Processes and Practices***

There are two ways in which one may conceptualize selection system sustainability. On the one hand, Kehoe et al. (2009) refer to sustainability as a selection system that is consistent with the firm's governance, strategy, culture, and stakeholder preferences. They note that, even though a selection system may have great sophistication and technical rigor, it may not be sustainable because it is not supported or is inconsistent with the firm's vision and mission. Consistent with this thinking, recent approaches have started to look at selection processes in a more integrated manner (Aguinis & Smith, 2007).

The other major way to view sustainability is from a competitive/economic perspective. Sustainable competitive advantage refers to a situation in which a firm generates above-average returns relative to competitors over a reasonable period of time (Armstrong & Shimizu, 2007). Sustainable

competitive advantage clearly implies a longitudinal perspective, and that the resources that create the competitive advantage are not depleted to the point of being useless. Yet selection research is dominated by an emphasis on practices usually conducted cross-sectionally at the individual level. For example, Cascio and Aguinis (2008a) remind us that determinations of whether criterion-related validity is "significant" are often made to a comparison standard of zero validity (i.e., the correlation coefficient is zero). The standard is only that the selection system works better than random hiring; it says nothing about sustainability. As another example, a selection system that attracts and hires the most qualified candidates, who then leave the firm after a few weeks, is not a sustainable system. There may be validity for the predictors, but this by itself does not guarantee sustainability of the system.

A useful means for understanding sustainability is provided by the resource-based view of the firm (RBV; Barney, 1991; Peteraf, 1993). Simply described, the RBV posits that between-firm differences occur because of differences in the firm's resource endowments. Resources may include such forms of capital as physical, financial, and human. Firms differ in their effectiveness because there are between-firm differences in their resources, just as individuals perform differently because there are individual differences in their KSAOs. And as at the individual level, not all firm-level resources contribute to sustained competitive advantage. Resources must be valuable, rare, inimitable, and non-substitutable if they are to provide a basis for sustained competitive advantage. Valuable and rare resources may contribute to competitive advantage in the short term or parity with other competitors (Barney & Wright, 1998). However, to provide sustained competitive advantage, the resource cannot be easily copied by competitors (inimitable), and it cannot be substituted with other resources.

When viewed from the RBV, it becomes apparent that valid personnel selection will not always contribute to *sustained* competitive advantage (see Ployhart & Weekley, 2010). First, if the selection process is not consistently applied, human capital emergence cannot occur. Second, if the human capital is not consistent with the firm's strategy, human capital advantage will be lessened or nonexistent. Third, the value of human capital is context- and firm-specific (Barney, 1991). Here is, of course, where the importance of training is recognized. By converting the "raw materials" of cognitive ability and personality into specific forms of knowledge and

skill, the context-specific KSAOs become important drivers of competitive advantage. Unfortunately, there is only limited research that links selection and training (for an exemplary article that conducts such an integration, see Van Iddekinge et al., 2009). Fourth, only human capital can provide the basis for sustainable competitive advantage (Barney & Wright, 1998). Selection practices like interviews or personality tests can be easily copied by competitors and hence cannot provide sustained competitive advantage. To become sustainable, a critical mass of human capital must be developed because it cannot be easily or quickly duplicated. Obviously, empirical research is needed to test these claims, and doing so will require selection researchers to expand their perspective to take a much more explicit consideration of context (Cascio & Aguinis, 2008a; Schneider et al., 2000). One critical change will be a shift from focusing on selection predictors to an understanding of the human capital that is created from these practices.

#### ***Mission #4: Show the Value of Selection on High-Impact Jobs to High-Impact Managers***

Most empirical personnel selection research has focused on large-scale hiring of entry-level (or more generally, lower level) employees. These jobs are viewed as important, but not necessarily valuable or unique (Lepak & Snell, 1999). Instead, jobs that are valuable and unique are those that "drive the business." Key managerial roles, top management teams, and obviously line directors, presidents, and CEOs are illustrative of jobs that are both valuable and unique. These are considered "A" jobs and should require a different set of selection practices than lower level "B" or "C" jobs. For example, Lepak and Snell (1999) predict that those in jobs with high value and uniqueness should be developed internally, so HR practices should emphasize commitment to the firm and selection processes should focus on aptitude (e.g., basic ability and personality) instead of current knowledge and skill. Differences in HR configurations across employees differing in their value and uniqueness (Lepak & Snell, 2002) help support these predictions.

Future selection research should expand beyond the large-scale studies to also consider situations where there are fewer numbers of jobs, but the jobs are more important. For example, how does one best staff a top management team? Recent research has started to examine psychological variables and processes that may impact top management team effectiveness (Barrick, Bradley, Kristof-Brown, &

Colbert, 2007; Colbert, Kristof-Brown, Bradley, & Barrick, 2008). This work follows in the tradition of an "upper echelons" perspective (Hambrick & Mason, 1984) and provides a compelling means to unite micro and macro perspectives. For many key roles, there will likely be highly intensive selection processes based on relatively few candidates. In this situation, the literature on individual assessment is relevant (see an edited book by Jeanneret & Silzer, 1998).

Switching gears slightly, it is also important that selection research take a broader view of its stakeholders and understand the managers who make the final decisions on selection system processes and own their implementation (Highhouse, 2008). Much of the personnel selection literature reads as though the end "customer" of the research is either an I/O practitioner, HR specialist, or an applicant. In many, if not most, organizational contexts, I/O psychologists will be interacting directly with HR or line managers who may have little to no background in selection. Hence, we need to understand their beliefs, preferences, and constraints. Ployhart (2006) stated: "If organizational decision makers are perhaps the ultimate consumers of our science, how is it that we have little understanding of what our customers want, need, or are willing to use?" (p. 891). It is rare that validity would be the primary consideration for deciding whether to implement a selection system (at least in private industry); so validity must also be balanced against cost, potential for adverse impact, the firm's strategy, internal politics, and acceptability among those who must implement the system. Kehoe et al. (2009) provide a comprehensive treatment of these complex issues.

### **Mission #5: Connect to Private Industry**

Selection as a profession is very applied in its orientation. That does not mean that selection research published in top journals is tightly connected to the current realities facing firms in private industry (Cascio & Aguinis, 2008b). First, there are many pressing issues facing private firms that have received scant attention from the scientific community. A good example is technology; firms struggle with understanding how to best implement web-based testing and selection, and yet there are few empirical studies and lots of fundamental questions that need answers (see Reynolds & Dickter, 2010; Tippins et al., 2006). Second, many research-based findings and best practices do not frequently consider whether they can be applied in private industry. For example, many selection best practices are so

complex, detailed, time-consuming, or burdensome that they are likely to never be used as intended.

Note the emphasis on private firms. Much of the knowledge of selection practices has focused on, and been acquired from, civil service organizations (e.g., state police and firefighter departments). There is nothing wrong with this data, but it must be realized that civil service organizations and private firms differ from each other in many profound ways. For example, ensuring demographic representation may be part of a civil service organization's charter, but there is no such charter among private firms (Pyburn et al., 2008). Consequently, research and debates surrounding the appropriateness of banding have occurred (see Aguinis, 2004), but I have never seen or heard of an application of banding in private industry. This is not to say that we should abandon civil service firms or become slaves to the fickle fashion and frequent foolishness of private industry. Rather, it means that selection research must continue to conduct rigorous scientific research with a clear understanding of how the research applies to private firms *and* civil service firms. To illustrate this need, consider that many private firms struggle with talent issues surrounding globalization. There is some research that speaks to how selection practices may be affected by cultural differences (Myors et al., 2008; Ryan, McFarland, Baron, & Page, 1999), but there is little systematic research examining how global firms should attract, select, and retain talent from multiple cultures and countries. These are the issues that keep HR and line managers up at night, and as a science, selection is less relevant if it does not consider them.

### **Selection Is Dead: Long Live Selection!**

The vision of personnel selection should be to ensure sustainable organizational effectiveness through the acquisition of human capital. Achieving this vision will require selection researchers and practitioners to build from nearly 100 years of research on the acquisition of human capital to also ensure sustainable organizational effectiveness. Doing so does nothing to devalue or abandon the rigorous and sophisticated scholarship that has been the tradition for the last century. On the contrary, doing so only increases the chances that professionally developed selection predictors and practices will be employed to achieve sustained competitive advantage.

### **Author Note**

I thank Doug Hawks for his assistance in compiling the references and gathering articles.

## References

- Aguinis, H. (2004). *Test score banding in human resource selection: Legal, technical, and societal issues*. Westport, CT: Quorum Books.
- Aguinis, H., & Smith, M. A. (2007). Understanding the impact of test validity and bias on selection errors and adverse impact in human resource selection. *Personnel Psychology*, 60, 165–199.
- Anderson, N. R. (2005). Relationships between practice and research in personnel selection: Does the left hand know what the right is doing? In A. Evers, N. R. Anderson, & O. Smit-Voskuyl (Eds.), *The Blackwell handbook of personnel selection* (1–24). Oxford: Blackwell.
- Armstrong, C. E., & Shimizu, K. (2007). A review of approaches to empirical research on the resource-based view of the firm. *Journal of Management*, 33, 959–986.
- Arthur, W., Jr., Bell, S. T., Villado, A. J., & Doverspike, D. (2006). The use of person-organization fit in employment decision making: An assessment of its criterion-related validity. *Journal of Applied Psychology*, 91, 786–801.
- Arthur, W., Jr., Day, E. A., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology*, 56, 125–153.
- Arthur, W., & Villado, A. J. (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection research and practice. *Journal of Applied Psychology*, 93, 435–442.
- Axelrod, E. L., Handfield-Jones, H., & Welsh, T. A. (2001). War for talent, Part 2. *The McKinsey Quarterly*, 2, 9–12.
- Barney, J. B. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, 17, 99–120.
- Barney, J. B., & Wright, P. W. (1998). On becoming a strategic partner: The role of human resources in gaining competitive advantage. *Human Resource Management*, 37, 31–46.
- Barrick, M. R., Bradley, B. H., Kristof-Brown, A. L., & Colbert, A. E. (2007). The moderating role of top management team interdependence: Implications for real teams and working groups. *Academy of Management Journal*, 50, 544–557.
- Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, 1–26.
- Barrick, M. R., & Ryan, A. M. (2003). *Personality and work: Reconsidering the role of personality in organizations*. San Francisco: Jossey-Bass.
- Bauer, T. N., & Truxillo, D. M. (2006). Applicant reaction to situational judgment tests: Research and related practical issues. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement and application* (pp. 233–252). Mahwah, NJ: Erlbaum.
- Becker, B. E., & Huselid, M. A. (2006). Strategic human resource management: Where do we go from here? *Journal of Management*, 32, 898–925.
- Becker, G. S. (1964). *Human capital: A theoretical and empirical analysis with special reference to education*. New York: National Bureau of Economic Research.
- Berry, C., Gruys, M., & Sackett, P. R. (2006). Educational attainment as a proxy for cognitive ability in selection: Effects on levels of cognitive ability and adverse impact. *Journal of Applied Psychology*, 91, 696–705.
- Binning, J. F., & Barrett, G. V. (1989). Validity of personnel decisions: A conceptual analysis of the inferential and evidential bases. *Journal of Applied Psychology*, 74, 478–494.
- Bobko, P., Roth, P. L., & Potosky, D. (1999). Derivation and implications of a meta-analysis matrix incorporating cognitive ability, alternative predictors, and job performance. *Personnel Psychology*, 52, 561–589.
- Boudreau, J. W., & Ramstad, P. M. (2003). Strategic industrial and organizational psychology and the role of utility analysis models. In W. C. Borman, D. R. Ilgen, & R. J. Klimoski (Eds.), *Handbook of psychology* (Vol. 12, pp. 193–224). Hoboken, NJ: Wiley.
- Brannick, M. T., Levine, E. L., & Morgeson, F. P. (2007). *Job and work analysis: Methods, research, and applications for human resource management* (2nd ed.). Thousand Oaks, CA: Sage.
- Cable, D. M., & DeRue, D. S. (2002). The convergent and discriminant validity of subjective fit perceptions. *Journal of Applied Psychology*, 87, 875–884.
- Cable, D. M., & Parsons, C. K. (2001). Socialization tactics and person-organization fit. *Personnel Psychology*, 54, 1–24.
- Campion, M. A., Palmer, D. K., & Campion, J. E. (1997). A review of structure in the selection interview. *Personnel Psychology*, 50, 655–702.
- Carroll, J. B. (1993). *Human cognitive abilities*. New York: Cambridge University Press.
- Cascio, W. F. (2000). Costing human resources: The financial impact of behavior in organizations. Cincinnati, OH: Southwestern.
- Cascio, W. F., & Aguinis, H. (2008a). Staffing twenty-first-century organizations. *Academy of Management Annals*, 2, 133–165.
- Cascio, W. F., & Aguinis, H. (2008b). Research in industrial and organizational psychology from 1963 to 2007: Changes, choices, and trends. *Journal of Applied Psychology*, 93, 1062–1081.
- Chan, D., & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology*, 82, 143–159.
- Cleary, T. A. (1968). Test bias: Prediction of grades of Negro and white students in integrated colleges. *Journal of Educational Measurement*, 5, 115–124.
- Clevenger, J., Pereira, G. M., Wiechmann, D., Schmitt, N., & Harvey, V. S. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology*, 86, 410–417.
- Colbert, A. E., Kristof-Brown, A. L., Bradley, B. H., & Barrick, M. R. (2008). CEO transformational leadership: The role of goal importance congruence in top management teams. *Academy of Management Journal*, 51, 51–96.
- Cortina, J. M., Goldstein, N. B., Payne, S. C., Davison, H. K., & Gilliland, S. W. (2000). The incremental validity of interview scores over and above cognitive ability and conscientiousness scores. *Personnel Psychology*, 53, 325–351.
- Davis, R. V., & Lofquist, L. H. (1984). *A psychological theory of work adjustment*. Minneapolis: University of Minnesota Press.
- DeShon, R. P. (2002). A generalizability theory perspective on measurement error corrections in validity generalization. In K. R. Murphy (Ed.), *Validity generalization: A critical review* (pp. 365–402). Mahwah, NJ: Erlbaum.
- DeNisi, A. S., Hitt, M. A., & Jackson, S. E. (2003). The knowledge-based approach to sustainable competitive advantage. In S. E. Jackson, M. A. Hitt, & A. S. DeNisi (Eds.), *Managing knowledge for sustained competitive advantage* (pp. 3–33). San Francisco: Jossey-Bass.
- Dierdorff, E. C., & Morgeson, F. P. (2007). Consensus in work role requirements: The influence of discrete occupational context on role expectations. *Journal of Applied Psychology*, 92, 1228–1241.

- Dudley, N. M., Orvis, K. A., Lebiecki, J. A., & Cortina, J. M. (2006). A meta-analytic investigation of conscientiousness in the prediction of job performance: Examining the intercorrelations and the incremental validity of narrow traits. *Journal of Applied Psychology*, 91, 40–57.
- Drasgow, F. (2003). Intelligence in the workplace. In W. C. Borman, D. R. Ilgen, & R. J. Klimoski (Eds.), *Handbook of psychology: Industrial and organizational psychology* (Vol. 12, pp. 107–130). Hoboken, NJ: Wiley.
- Drucker, P. (1966). *The effective executive*. New York: Harper & Row.
- Eder, R. W., & Harris, M. M. (Eds.). (1999). *The employment interview handbook*. Thousand Oaks, CA: Sage.
- The Search for Talent, *The Economist*. (October 7, 2006). 381 (8498), 11.
- Edwards, J. R. (2008). Person-environment fit in organizations: An assessment of theoretical progress. *The Academy of Management Annals*, 2, 167–260.
- Ellingson, J. E., Sackett, P. R., & Connolly, B. S. (2007). Personality assessment across selection and development contexts: Insights into response distortion. *Journal of Applied Psychology*, 92, 386–395.
- Fine, S. A., & Wiley, W. W. (1974). An introduction to functional job analysis. In F. A. Fleishman & A. R. Bass (Eds.), *Studies in personnel and industrial psychology* (3rd ed., pp. 6–13). Homewood, IL: Irwin.
- Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, 51, 327–355.
- Foldes, H. J., Duehr, E. E., & Ones, D. S. (2008). Group differences in personality: Meta-analyses comparing five U.S. racial groups. *Personnel Psychology*, 61, 579–616.
- Gatewood, R. D., & Feild, H. S. (1998). *Human resource selection*. Fort Worth, TX: The Dryden Press/Harcourt Brace.
- Gerhart, B. (2005). Human resources and business performance: Findings, unanswered questions, and an alternative approach. *Management Review*, 16, 174–185.
- Ghiselli, E. E. (1966). *The validity of occupational aptitude tests*. New York: Wiley.
- Goldstein, H. W., Yusko, K. P., Braverman, E. P., Smith, D. B., & Chung, B. (1998). The role of cognitive ability in the subgroup differences and incremental validity of assessment center exercises. *Personnel Psychology*, 51, 357–374.
- Goldstein, H. W., Yusko, K. P., & Nicolopoulos, V. (2001). Exploring black-white subgroup differences of managerial competencies. *Personnel Psychology*, 54, 783–807.
- Guion, R. M., & Gottier, R. F. (1965). Validity of personality measures in personnel selection. *Personnel Psychology*, 18, 49–65.
- Hambrick, D. C., & Mason, P. A. (1984). Upper echelons: The organization as a reflection of its top managers. *Academy of Management Review*, 9, 193.
- Harvey, R. J., & Wilson, M. A. (2000). Yes Virginia, there is an objective reality in job analysis. *Journal of Organizational Behavior*, 21, 829–854.
- Hatch, N. W., & Dyer, J. H. (2004). Human capital and learning as a source of sustainable competitive advantage. *Strategic Management Journal*, 25, 1155–1178.
- Hausknecht, J. P., Halpert, J. A., Di Paolo, N. T., & Moriarty Gerrard, M. O. (2007). Retesting in selection: A meta-analysis of coaching and practice effects for tests of cognitive ability. *Journal of Applied Psychology*, 92, 373–385.
- Hausknecht, J. P., Trevor, C. O., & Farr, J. L. (2002). Retaking ability tests in a selection setting: Implications for practice effects, training performance, and turnover. *Journal of Applied Psychology*, 87, 243–254.
- Heneman, R. G., & Judge, T. A. (2006). *Staffing organizations* (5th ed.). New York: Irwin/McGraw-Hill.
- Highhouse, S. (2008). Stubborn reliance on intuition and subjectivity in employee selection. *Industrial and Organizational Psychology*, 3, 333–342.
- Hogan, J. (1991). Physical abilities. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 2, pp. 753–831). Palo Alto, CA: Consulting Psychologists Press.
- Hogan, J., Barrett, P., & Hogan, R. (2007). Personality measurement, faking, and employment selection. *Journal of Applied Psychology*, 92, 1270–1285.
- Holland, J. L. (1997). Making vocational choices: A theory of vocational personalities and work environments (3rd ed.). Odessa, FL: PAR.
- Hough, L. M. (1998). Personality at work: Issues and evidence. In M. Hakel (Ed.), *Beyond multiple choice: Evaluating alternatives to traditional testing for selection* (pp. 131–166). Mahwah, NJ: Erlbaum.
- Hough, L. M., & Furnham, A. (2003). Use of personality variables in work settings. In W. C. Borman, D. R. Ilgen, & R. J. Klimoski (Eds.), *Handbook of psychology: Industrial and organizational psychology* (Vol. 12, pp. 131–169). Hoboken, NJ: Wiley.
- Hough, L. M., & Oswald, F. L. (2008). Personality testing and industrial-organizational psychology: Reflections, progress, and prospects. *Industrial and Organizational Psychology*, 3, 272–290.
- Hough, L. M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection, and amelioration of adverse impact in personnel selection procedures: Issues, evidence, and lessons learned. *International Journal of Selection and Assessment*, 9, 152–194.
- Hough, L. M., & Schneider, R. J. (1996). Personality traits, taxonomies, and applications in organizations. In K. R. Murphy (Ed.), *Individual differences and behavior in organizations* (pp. 3–30). San Francisco: Jossey-Bass.
- Huffcutt, A. I., Conway, J. M., Roth, P. L., & Stone, N. J. (2001). Identification and meta-analytic assessment of psychological constructs measured in employment interviews. *Journal of Applied Psychology*, 86, 897–913.
- Huffcutt, A. I., Weekley, J. A., Wiesner, W. H., Degroot, T. G., & Jones, C. (2001). Comparison of situational and behavior description interview questions for higher-level positions. *Personnel Psychology*, 54(3), 619–644.
- Hunter, J. E. (1983). A causal analysis of cognitive ability, job knowledge, job performance, and supervisory ratings. In F. Landy, S. Zedeck, & J. Cleveland (Eds.), *Performance measurement and theory* (pp. 257–266). Hillsdale, NJ: Erlbaum.
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72–95.
- Hunter, J. E., Schmidt, F. L., & Le, H. (2006). Implications for direct and indirect range restriction for meta-analysis methods and findings. *Journal of Applied Psychology*, 91, 594–612.
- Huselid, M. A. (1995). The impact of human resource management practices on turnover, productivity, and corporate financial performance. *Academy of Management Journal*, 38, 635–672.
- Ilgen, D. R., & Hollenbeck, J. R. (1991). The structure of work: Job design and roles. In M. D. Dunnette & L. M. Hough

- (Eds.), *Handbook of industrial and organizational psychology*, (Vol. 2, pp. 165–207). Palo Alto, CA: Consulting Psychologists Press.
- James, L. R. (1998). Measurement of personality via conditional reasoning. *Organizational Research Methods*, 1, 131–163.
- Jeanneret, R. P., & Silzer, R. (1998). *Individual assessment: Predicting behavior in organizational settings*. San Francisco: Jossey-Bass.
- Johnson, J. W., & Oswald, F. L. (2010). Test administration and the use of test scores. In J. L. Farr & N. T. Tippins (Eds.), *Handbook of employee selection* (pp. 151–170). New York: Taylor & Francis.
- Kanfer, R., & Ackerman, P. L. (2004). Aging, adult development, and work motivation. *Academy of Management Review*, 29, 440–458.
- Kehoe, J., Mol S., & Anderson, N. R. (2010). Managing sustainable selection programs. In J. L. Farr & N. T. Tippins (Eds.), *Handbook of employee selection* (pp. 213–234). New York: Taylor & Francis.
- Klimoski, R. J., & Brickner, M. (1987). Why do assessment centers work? The puzzle of assessment center validity. *Personnel Psychology*, 40, 243–260.
- Koch, M. J., & McGrath, R. G. (1996). Improving labor productivity: Human resource management policies do matter. *Strategic Management Journal*, 17, 335–354.
- 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.
- Lance, C. E. (2008). Why assessment centers do not work and the way they are supposed to. *Industrial and Organizational Psychology*, 1, 84–97.
- Lance, C. E., Lambert, T. A., Gewin, A. G., Lievens, F., & Conway, J. M. (2004). Revised estimates of dimension and exercise variance components in assessment center postexercise dimension ratings. *Journal of Applied Psychology*, 89, 377–385.
- Latham, G. P., & Whyte, G. (1994). The futility of utility analysis. *Personnel Psychology*, 47, 31–47.
- Le, H., Oh, I., Shaffer, J., & Schmidt, F. (2007). Implications of methodological advances for the practice of personnel selection: How practitioners benefit from meta-analysis. *Academy of Management Perspectives*, 21, 6–15.
- Lepak, D. P., & Snell, S. A. (1999). The human resource architecture: Toward a theory of human capital allocation and development. *Academy of Management Review*, 24, 31–48.
- Lepak, D. P., & Snell, S. A. (2002). Examining the human resource architecture: The relationships among human capital, employment, and human resource configurations. *Journal of Management*, 28, 517–543.
- Lievens, F. (2002). Trying to understand the different pieces of the construct validity puzzle of assessment centers: An examination of assessor and assessee effects. *Journal of Applied Psychology*, 87, 675–686.
- Lievens, F., DeCorte, W., & Schollaert, E. (2008). A closer look at the frame-of-reference effect in personality scale scores and validity. *Journal of Applied Psychology*, 93, 268–279.
- MacDuffie, J. P. (1995). Human-resource bundles and manufacturing performance: Organizational logic and flexible production systems in the world auto industry. *Industrial & Labor Relations Review*, 48, 197–221.
- Mael, F. A. (1991). A conceptual rationale for the domain and attributes of biodata items. *Personnel Psychology*, 44, 763–792.
- McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. L. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology*, 60(1), 63–91.
- McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology*, 86(4), 730–740.
- McDaniel, M. A., & Nguyen, N. T. (2001). Situational judgment tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment*, 9(1–2), 103–113.
- McFarland, L. A., Ryan, A. M., & Kriska, S. D. (2003). Impression management use and effectiveness across assessment methods. *Journal of Management*, 29, 641–661.
- McNemar, Q. (1964). Lost: Our intelligence. Why? *American Psychologist*, 19, 871–882.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741–749.
- Morgeson, F. P., & Campion, M. A. (1997). Social and cognitive sources of potential inaccuracy in job analysis. *Journal of Applied Psychology*, 82, 627–655.
- Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007a). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology*, 60, 683–729.
- Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007b). Are we getting fooled again? Coming to terms with limitations in the use of personality tests in personnel selection. *Personnel Psychology*, 60, 1029–1049.
- Morgeson, F. P., Delaney-Klinger, K., Mayfield, M. S., Ferrara, P., & Campion, M. A. (2004). Self-presentation processes in job analysis: A field experiment investigating inflation in abilities, tasks, and competencies. *Journal of Applied Psychology*, 89, 674–686.
- Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology*, 75, 649–647.
- Murphy, K. R., & DeShon, R. P. (2000). Interrater correlations do not estimate the reliability of job performance ratings. *Personnel Psychology*, 53, 873–900.
- Murphy, R. R. (2002). *Validity generalization: A critical review*. Mahwah, NJ: Erlbaum.
- Myors, B., Lievens, F., et al. (2008). International perspectives on the legal environment for selection. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 1, 206–246.
- Ones, D. S., & Viswesvaran, C. (1998). The effects of social desirability and faking on personality and integrity assessment for personnel selection. *Human Performance*, 11, 245–269.
- Ones, D. S., Dilchert, S., Viswesvaran, C., & Judge, T. A. (2007). In support of personality assessment in organizational settings. *Personnel Psychology*, 60, 995–1027.
- Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology*, 81, 660–679.
- Ostroff, C., & Judge, T. A. (2007). *Perspectives on organizational fit*. Mahwah, NJ: Erlbaum.
- Peeters, H., & Lievens, F. (2005). Situational judgment tests and their predictiveness of college students' success: The influence

- of faking. *Educational and Psychological Measurement*, 65, 70–89.
- Peteraf, M. A. (1993). The cornerstones of competitive advantage: A resource-based view. *Strategic Management Journal*, 14, 179–191.
- Ployhart, R. E. (2004). Organizational staffing: A multilevel review, synthesis, and model. In J. J. Martocchio (Eds.), *Research in personnel and human resource management* (Vol. 23, pp. 121–176). Oxford: Elsevier.
- Ployhart, R. E. (2006). Staffing in the 21st century: New challenges and strategic opportunities. *Journal of Management*, 32, 868–897.
- Ployhart, R. E., & Holtz, B. C. (2008). The diversity-validity dilemma: Strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology*, 61, 153–172.
- Ployhart, R. E., & Schneider, B. (2002). A multilevel perspective on personnel selection: Implications for selection system design, assessment, and construct validation. In F. J. Dansereau & F. Yammarino (Eds.), *Research in multi-level issues, Volume 1: The many faces of multi-level issues* (pp. 95–140). Oxford: Elsevier.
- Ployhart, R. E., & Schneider, B. (2005). Multilevel selection and prediction: Theories, methods, and models. In A. Evers, O. Smit-Voskuyl, & N. Anderson (Eds.), *Handbook of personnel selection* (pp. 495–516). Oxford: Blackwell.
- Ployhart, R. E., Schneider, B., & Schmitt, N. (2006). *Staffing organizations: Contemporary practice and theory* (3rd ed.). Mahwah, NJ: Erlbaum.
- Ployhart, R. E., & Weekley, J. A. (2010). Strategy, selection, and sustained competitive advantage. In J. Farr & N. Tippins (Eds.) *Handbook of employee selection*. Mahwah, NJ: Erlbaum.
- Ployhart, R. E., Weekley, J. A., & Baughman, K. (2006). The structure and function of human capital emergence: A multi-level examination of the ASA model. *Academy of Management Journal*, 49, 661–677.
- Ployhart, R. E., Weekley, J. A., & Ramsey, J. (2009). The consequences of human resource stocks and flows: A longitudinal examination of unit service orientation and unit effectiveness. *Academy of Management Journal*.
- Posthumus, R. A., Morgeson, F. P., & Campion, M. A. (2002). Beyond employment interview validity: A comprehensive narrative review of recent research and trends over time. *Personnel Psychology*, 55, 1–81.
- Pyburn, K., Ployhart, R. E., & Kravitz, D. A. (2008). The diversity-validity dilemma: Overview and legal context. *Personnel Psychology*, 61, 143–151.
- Quiñones, M. A., Ford, J. K., & Teachout, M. S. (1995). The relationship between work experience and job performance: A conceptual and meta-analytic review. *Personnel Psychology*, 48, 887–910.
- Ree, M. J., Earles, J. A., & Teachout, M. S. (1994). Predicting job performance: Not much more than g. *Journal of Applied Psychology*, 79, 518–524.
- Reynolds, D. H., & Dickter, D. N. (2010). Technology and employee selection. In J. Farr & N. Tippins, (Eds.), *The handbook of employee selection* (2010). Mahwah, NJ: Erlbaum.
- Roth, P. L., Bevier, C. A., Bobko, P., Switzer, F. S., III, & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology*, 54, 297–330.
- Roth, P. L., Bobko, P., McFarland, L. A., & Buster, M. (2008). Work sample tests in personnel selection: A meta-analysis of black-white differences in overall and exercise scores. *Personnel Psychology*, 61, 637–662.
- Roth, P. L., Huffcutt, A. I., & Bobko, P. (2003). Ethnic group differences in measures of job performance: A new meta-analysis. *Journal of Applied Psychology*, 88, 694–706.
- Russell, C. J. (1994). Generation procedures for biodata items: A point of departure. In G. S. Stokes, M. D. Mumford, & W. A. Owens (Eds.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction* (pp. 17–38). Palo Alto, CA: Consulting Psychologists Press.
- Ryan, A. M., McFarland, L. A., Baron, H., & Page, R. (1999). An international look at selection practices: Nation and culture as explanations for variability in practice. *Personnel Psychology*, 52, 359–391.
- Rynes, S. L., Brown, K. G., & Colbert, A. E. (2002). Seven misconceptions about human resource practices: Research findings versus practitioner beliefs. *Academy of Management Executive*, 16, 92–103.
- Sackett, P. R., & Ellingson, J. E. (1997). The effects of forming multi-predictor composites on group differences and adverse impact. *Personnel Psychology*, 50, 707–721.
- Sackett, P. R., Laczo, R. M., & Arvey, R. D. (2002). The effects of range restriction on estimates of criterion interrater reliability: Implications for validation research. *Personnel Psychology*, 55, 807–825.
- Sackett, P. R., Lievens, F., Berry, C. M., & Landers, R. N. (2007). A cautionary note on the effects of range restriction on predictor intercorrelations. *Journal of Applied Psychology*, 92, 538–544.
- Sackett, P. R., & Roth, L. (1996). Multi-stage selection strategies: A Monte Carlo investigation of effects on performance and minority hiring. *Personnel Psychology*, 49, 549–572.
- Sackett, P. R., Schmitt, N., Ellingson, J. E., & Kabin, M. B. (2001). High-stakes testing in employment, credentialing, and higher education: Prospects in a post-affirmative action world. *American Psychologist*, 56, 302–318.
- Sackett, P. R., & Wilk, S. L. (1994). Within-group norming and other forms of score adjustment in preemployment testing. *American Psychologist*, 49, 929–954.
- Sackett, P. R., & Yang, H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Psychology*, 85, 112–118.
- Schaubroeck, J., Ganster, D. C., & Jones, J. R. (1998). Organization and occupation influences in the attraction-selection-attrition process. *Journal of Applied Psychology*, 83, 869–891.
- Schmidt, F. L. (1988). The problem of group differences in ability test scores in employment selection. *Journal of Vocational Behavior*, 33, 272–292.
- Schmidt, F. L., & Hunter, J. E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods*, 1, 199–223.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262–274.
- Schmidt, F. L., Hunter, J. E., McKenzie, R., & Muldrow, T. (1979). Impact of valid selection procedures on workforce productivity. *Journal of Applied Psychology*, 64, 609–626.
- Schmidt, F. L., Le, H., & Ilies, R. (2003). Beyond alpha: An empirical examination of the effects of different sources of measurement error on reliability estimates for measures of

- individual-differences constructs. *Psychological Methods*, 8, 206–224.
- Schmidt, F., Le, H., Oh, I., & Shaffer, J. (2007). General mental ability, job performance, and red herrings: Responses to Osterman, Hauser, and Schmitt. *Academy of Management Perspectives*, 21, 64–76.
- Schmidt, F. L., Oh, I. S., & Le, H. (2006). Increasing the accuracy of corrections for range restriction: Implications for selection procedure validities and other research results. *Personnel Psychology*, 59, 281–305.
- Schmidt, F. L., Viswesvaran, C., & Ones, D. S. (2000). Reliability is not validity and validity is not reliability. *Personnel Psychology*, 53, 901–912.
- Schmidt, F. L., & Zimmerman, R. D. (2004). A counterintuitive hypothesis about employment interview validity and some supporting evidence. *Journal of Applied Psychology*, 89(3), 553–561.
- Schmitt, N. (2004). Beyond the Big Five: Increases in understanding and practical utility. *Human Performance*, 17, 347–357.
- Schmitt, N. (2007). The value of personnel selection: Reflections on some remarkable claims. *Academy of Management Perspectives*, 21, 19–23.
- Schmitt, N., & Chan, D. (2006). Situational judgment tests: Method or construct? In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and application* (pp. 135–155). Mahwah, NJ: Erlbaum.
- Schmitt, N., Cortina, J. M., Ingerick, M. J., & Wiechmann, D. (2003). Personnel selection and employee performance. In W. C. Borman, D. R. Ilgen, & R. J. Klimoski (Eds.), *Handbook of psychology, Volume 12: Industrial and organizational psychology* (pp. 77–105). Hoboken, NJ: Wiley.
- Schmitt, N., Oswald, F. L., Kim, B. H., Gillespie, M. A., Ramsay, L. J., & Yoo, T. Y. (2003). Impact of elaboration on socially desirable responding and the validity of biodata measures. *Journal of Applied Psychology*, 88, 979–988.
- Schneider, B. (2007). Evolution of the study and practice of personality at work. *Human Resource Management*, 46, 583–610.
- Schneider, B., & Smith, D. B. (2004). *Personality and organizations*. Mahwah, NJ: Erlbaum.
- Schneider, B., Smith, D. B., & Sipe, W. P. (2000). Personnel selection psychology: Multilevel considerations. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 91–120). San Francisco: Jossey-Bass.
- Schneider, B., Smith, D. B., Taylor, S., & Fleenor, J. (1998). Personality and organizations: A test of the homogeneity of personality hypothesis. *Journal of Applied Psychology*, 83, 462–470.
- Smith, D. B., & Robie, C. (2004). The implications of impression management for personality research in organizations. In B. Schneider & D. B. Smith (Eds.), *Personality and organizations* (pp. 111–138). Mahwah, NJ: Erlbaum.
- Society for Industrial and Organizational Psychology, I. (2003). *Principles for the validation and use of personnel selection procedures* (4th ed.). Bowling Green, OH: Author.
- Spearman, C. (1927). *The abilities of man*. New York: Macmillan.
- Spsychalski, A. C., Quiñones M. A., Gaugler, B. B., & Pohley, K. (1997). A survey of assessment center practices in organizations in the United States. *Personnel Psychology*, 50, 71–90.
- Stokes, G. S., Mumford, M. D., & Owens, W. A. (Eds.). (1994). *Biodata handbook: Theory, research, and use of biographical information in selection and performance appraisal*. Palo Alto, CA: Consulting Psychologists Press.
- Terpstra, D. E., & Rozell, E. J. (1993). The relationship of staffing practices to organizational level measures of performance. *Personnel Psychology*, 46, 27–48.
- Tesluk, P. E., & Jacobs, R. R. (1998). Toward an integrative model of work experience. *Personnel Psychology*, 51, 321–355.
- Tett, R. P., & Christiansen, N. D. (2007). Personality tests at the crossroads: A response to Morgeson, Campion, Dipboye, Hollenbeck, Murphy, and Schmitt. *Personnel Psychology*, 60, 967–993.
- Tippins, N. T., Beaty, J., Drasgow, F., Gibson, W. M., Pearlman, K., Segall, D. O., & Shepherd, W. (2006). Unproctored internet testing in employment settings. *Personnel Psychology*, 59, 189–225.
- Ulrich, L., & Trumbo, D. (1965). The selection interview since 1949. *Psychological Bulletin*, 63, 100–116.
- Van Iddekinge, C. H., Ferris, G. R., Perrewe, P. L., Perryman, A. Z., Blass, F. R., & Heetderks, T. D. (2009). Effects of selection and training on unit-level performance over time: A latent growth modeling approach. *Journal of Applied Psychology*, 94, 829–843.
- Van Iddekinge, C. H., & Ployhart, R. E. (2008). Developments in the criterion-related validation of selection procedures: A critical review and recommendations for practice. *Personnel Psychology*, 61, 871–925.
- Van Iddekinge, C. H., Putka, D. J., Raymark, P. H., & Eidson, C. E., Jr. (2005). Modeling error variance in job specification ratings: The influence of rater, job, and organization-level factors. *Journal of Applied Psychology*, 90, 323–334.
- Wallace, S. R. (1965). Criteria for what? *American Psychologist*, 20, 411–417.
- Weekley, J. A., & Ployhart, R. E. (Eds.). (2006). *Situational judgment tests: Theory, measurement and application*. Mahwah, NJ: Erlbaum.
- Wernimont, P. R., & Campbell, J. P. (1968). Signs, samples, and criteria. *Journal of Applied Psychology*, 52, 372–376.
- Woehr, D. J., & Arthur, W., Jr. (2003). The construct-related validity of assessment center ratings: A review and meta-analysis of the role of methodological factors. *Journal of Management*, 29, 231–258.
- Wright, P. M., Dunford, B. D., & Snell, S. A. (2001). Human resources and the resource based view of the firm. *Journal of Management*, 27, 701–721.