

PERSONNEL SELECTION: Looking Toward the Future—Remembering the Past

Leaetta M. Hough and Frederick L. Oswald

The Dunnette Group, St. Paul, Minnesota 55102; e-mail: Leaetta@msn.com

Department of Psychological Sciences, Purdue University, West Lafayette, Indiana 47907; e-mail: oswald@psych.purdue.edu

Key Words job performance, assessment, validity, adverse impact, personality

■ **Abstract** This chapter reviews personnel selection research from 1995 through 1999. Areas covered are job analysis; performance criteria; cognitive ability and personality predictors; interview, assessment center, and biodata assessment methods; measurement issues; meta-analysis and validity generalization; evaluation of selection systems in terms of differential prediction, adverse impact, utility, and applicant reactions; emerging topics on team selection and cross-cultural issues; and finally professional, legal, and ethical standards. Three major themes are revealed: (a) Better taxonomies produce better selection decisions; (b) The nature and analyses of work behavior are changing, influencing personnel selection practices; (c) The field of personality research is healthy, as new measurement methods, personality constructs, and compound constructs of well-known traits are being researched and applied to personnel selection.

CONTENTS

Introduction	632
Job and Work Analysis	632
Criteria.....	633
<i>Taxonomic Issues</i>	633
<i>Measurement Issues</i>	634
<i>Dynamic Criteria</i>	634
Predictors	634
<i>Cognitive Abilities and Job Knowledge</i>	634
<i>Personality</i>	636
<i>Multiple Predictor Domains</i>	640
Assessment Methods	641
<i>Interview</i>	641
<i>Assessment Centers</i>	642
<i>Biodata</i>	643
Measurement Issues and Validation Strategies	644
<i>Measurement Issues</i>	644

<i>Meta-Analysis and Validity Generalization</i>	645
Evaluation of Selection Systems	645
<i>Differential Prediction</i>	645
<i>Adverse Impact</i>	646
<i>Utility</i>	647
<i>Applicant Reactions</i>	647
Emerging Topics	648
<i>Team Member Selection</i>	648
<i>Cross-Cultural Selection Issues</i>	649
Professional, Legal, and Ethical Standards	650
<i>Professional</i>	650
<i>Legal</i>	650
<i>Ethical</i>	651
Parting Remarks	651

INTRODUCTION

Global, widespread, and diverse forces impact today's economies and marketplaces, with important implications for personnel selection (Dunnette 1997, Howard 1995, Ilgen & Pulakos 1999, Kraut & Korman 1999a, Pearlman & Barney 1999, Schmitt & Chan 1998). Indeed, Herriot & Anderson (1997) call for new selection methods. Our personnel selection review identifies emerging topics, covers traditional ones, and suggests new avenues for research and practice.

JOB AND WORK ANALYSIS

Recognizing the changing nature of work, many researchers and practitioners conduct "work" analysis, focusing on tasks and cross-functional skills of workers, rather than "job" analysis with its focus on static jobs (Cascio 1995, Nelson 1997, Pearlman 1997, Sanchez & Levine 1999). Perhaps the O*NET, the computerized delivery system for the *Dictionary of Occupational Titles (DOT)*, has made the greatest operational strides in addressing this shift (Peterson et al 1999). The O*NET is a flexible database containing occupational information structured around a "content model" linking work behaviors to worker attributes, much in line with Dunnette's call to bridge these "two worlds of behavioral taxonomies" (1976, p. 477). As information on work (e.g. jobs, organizational contexts, and work characteristics) and the worker (e.g. knowledge, skills, interests, and motivation) changes, the computerized nature of the O*NET allows an equally responsive change in its database. Collecting and using such information to select individuals for jobs, tasks, or roles is becoming more critical than ever (Campbell 1999).

Work/job analysis now includes personality variables alongside traditional cognitive, behavioral, and situational variables. The National Skill Standards Board (K Pearlman, unpublished manuscript) and O*NET (Peterson et al 1999) both incorporate personality-based work requirements. One personality-based job analysis instrument generates job profiles along seven personality scales (Hogan & Rybicki 1998). Another personality-based job analysis evaluated 260 jobs, meaningfully distinguishing 12 *DOT*-based job clusters (Raymark et al 1997).

Sixteen potential cognitive and social sources of rating inaccuracy in job analysis may influence different dimensions and psychometric properties of job analysis (Morgeson & Campion 1997). Type of job data and method for clustering jobs affect similarities and distinctions between jobs (Colihan & Burger 1995), and Q-factor analysis of rated task importance for two job titles shows meaningful within-title variation (Sanchez et al 1998). Future job analysis studies might identify substantive sources of variance attributable to types of raters, workers, or both.

CRITERIA

Taxonomic Issues

Job performance constitutes all measurable work behaviors relevant to organizational goals and within the individual's control (Campbell et al 1996). Job performance is complex, dynamic, and multidimensional, and consequently personnel selection systems might predict individual differences for several types of job performance (e.g. task proficiency and leadership behaviors). Models incorporating multiple predictors and multiple criteria first apply rational weights to performance criteria and then derive least-squares optimal weights for the predictors. Meta-analytic correlations between ability, conscientiousness (predictors), individual task performance, and organizational citizenship (criteria) illustrate how validity can vary greatly depending on criterion weights (Murphy & Shiarella 1997). Absent criterion data for establishing regression weights, rational weighting of a selection battery increases appropriateness and legal defensibility. Weights could multiply job analysis ratings of importance, time spent, consequences of errors, and time-to-proficiency (Arthur et al 1996).

Contextual performance, or organizational citizenship behavior, is a relatively new and multifaceted job performance construct (Borman & Motowidlo 1997). Coleman & Borman (1999) classified organizational citizenship behaviors into three broad categories: interpersonal citizenship behavior (benefiting employees), organizational citizenship performance (benefiting organizations), and job/task conscientiousness (benefiting work itself). Hierarchical regression analyses suggest that interpersonal facilitation is a part of contextual performance, but job dedication (similar to job/task conscientiousness within the organizational citizenship framework) is a part of task performance (Van Scotter & Motowidlo

1996). Organizations clearly require both task and contextual performance (Kiker & Motowidlo 1999).

Measurement Issues

Regarding criterion reliability, interrater reliability coefficients are more appropriate to use than intrarater (coefficient alpha) or test-retest reliabilities, and interrater reliabilities of supervisory ratings of overall and dimensional job performance are higher than peer ratings (Viswesvaran et al 1996). A greater understanding of “unreliable” criterion variance is needed to address biases in measures of constructs (Schmitt et al 1995). Similarly, criterion range restriction (or enhancement) may be a legitimate organizational phenomenon affecting criterion-related validities and not merely a statistical artifact (James et al 1992). Organizational climate, for instance, does not always attenuate correlations between procedural fairness and customer-perceived performance relationships (Burke et al 1996). Further research should investigate how individual differences, job types, and their interactions influence the mean and variance of criterion measures (see Hattrup & Jackson 1996).

Dynamic Criteria

Job performance and the relative contributions of its determinants (job knowledge, skill, and motivation) change, calling for longitudinal models of reliability and validity (Tisak & Tisak 1996). The nature of performance change (e.g. systematic vs random change or reversible vs irreversible change) and how constructs relate between individual and group levels are critical theoretical and methodological issues (Chan 1998a,b). Nonlinear mixed-effects models simultaneously estimate individual and group levels of change (Cudeck 1996), accommodating missing data, prespecified error structures (see DeShon et al 1998a), and individuals not measured at the same time points. For a review and tutorial of quadratic and linear models of longitudinal change, see Chan (1998a).

Personality items predicted change in eight consecutive quarters of securities sales performance, with a curvilinear group mean increase over time (i.e. greater increase initially) and different rates of increase for each individual (Ployhart & Hakel 1998). Psychomotor ability predicted initial piece-rate performance of sewing machine operators, and cognitive ability predicted performance change. Individuals with less experience and lower levels of initial performance changed more (Deadrick et al 1997).

PREDICTORS

Cognitive Abilities and Job Knowledge

Self-selection on cognitive ability may precede personnel selection. Job seekers may select into or “gravitate” toward jobs with ability requirements commensurate with the seekers’ own general cognitive ability (Wilk & Sackett 1996).

Ability self-evaluations might lead to seeking coaching and practice on ability tests. Firefighter applicants with lower cognitive ability scores are likely to attend a free test preparation program, although the program's effects on raising ability test scores (reading, listening, and spelling) are minimal (Ryan et al 1998b).

General Cognitive Ability Various cognitive ability tests (e.g. verbal, numerical, and spatial tests) intercorrelate positively, and the common variance often operationalizes *g*, a single general cognitive ability factor. For many jobs and practical work outcomes (job knowledge acquisition, training performance, and job performance), *g* predicts well (e.g. Levine et al 1996). For prediction in less complex jobs or in later stages of complex learning, *g* is less useful but rarely useless (Gottfredson 1997). Greater understanding of *g* is needed (Campbell 1996), as its determinants and theoretical meaning are debated (see Lubinski 2000). Individuals higher in *g* show lower intercorrelations between specific abilities (Legree et al 1996). Matching individuals' specific abilities (or ability profiles) to particular jobs may therefore be especially important for individuals with higher *g* (Lubinski & Benbow 1999). Item response theory and computerized adaptive testing have clarified relationships between *g* and specific abilities contributing to it (Sands et al 1997, Segall 1999). Developers of good cognitive ability selection tests cannot rely on knowledge of *g* alone.

Ability and Job Knowledge Multiple-ability test battery data from 3000 Air Force enlistees supported a hierarchical ability structure (Carretta & Ree 1996) that fit the data well, considerably better than two nonhierarchical models or a *g*-only model (see also Carretta et al 1998). Men and women show similar hierarchical ability structures (Carretta & Ree 1997).

Carretta & Doub (1998) tested the mediating effect of prior mechanical and electrical knowledge on the relationship between *g* and subsequent job knowledge of Air Force trainees. Comparing racial groups (White, Black, and Hispanic) resulted in little moderating effect. Comparing gender groups prior job knowledge mediated subsequent job knowledge for males but not for females. The effect of *g* was weaker in individuals with more prior job knowledge; conversely, *g* was stronger in individuals with less prior job knowledge. It seems that *g* predicts the rate of job knowledge acquisition, which in turn has a larger direct influence on performance ratings than the indirect effect of *g* (Ree et al 1995).

Academic Achievement and Language Proficiency Meta-analysis found that undergraduate college grade point average (GPA) predicted job performance across many types of organizations, especially for job performance measured closer in time to the GPA (Roth et al 1996). Another meta-analysis reported substantial criterion-related validities for aptitude tests predicting GPA in graduate school. Subject-specific tests (possibly a job knowledge analog) had higher validities than verbal and mathematical tests, but both were of useful magnitude (Kuncel et al 1999).

Increasing immigration and concomitant workforce diversity suggest measuring English language proficiency when selecting for certain jobs. For entry-level meat trimmers ($N = 87$) whose native language was not English, a written and spoken English proficiency test was internally consistent and clearly linked to job analysis information. The test predicted supervisory ratings of overall job performance (Chan et al 1999). Spoken English proficiency assessment is now possible via real-time computer processing and analysis of human speech (Bernstein 1999).

Adverse Impact Cognitive ability measures tend to show nontrivial racial-group mean differences. Helms' (1992) hypothesis states that White-Black mean differences might be reduced by couching ability test content within a social context. An expert panel modified abstract ability items to reflect everyday organizational, social, and life situations. Contrary to the hypothesis, marked White-Black differences remained under the new test format, even under large-sample replication and parallel test forms (DeShon et al 1998b). Research needs to expand the number and types of items, explore different administration formats, and examine other specific abilities.

Short-term memory tests (digit span and digit-symbol substitution) show promise as an alternative or supplement to traditional ability tests, with lower adverse impact and good validities with job performance. Meta-analysis estimated White-Black mean differences on short-term memory tests at 0.48 SD, about half the 1.0 SD difference typically found in general cognitive ability tests. Short-term memory tests are reliable and correlate with training performance and job performance [$r \approx 0.45$ (Verive & McDaniel 1996)]. Lower adverse impact combined with respectable overall criterion-related validity encourages future research in this area.

The Armed Services Vocational Aptitude Battery (ASVAB) and an experimental ability test battery predicted training performance criteria in 17 military jobs (Sager et al 1997), showing some necessary tradeoffs. One cannot completely (a) minimize adverse impact for all subgroups compared (see Hoffman & Thornton 1997 for this issue in utility context), (b) maximize both criterion-related validity within jobs and classification efficiency across jobs, or (c) satisfy both (a) and (b). A selection strategy aimed at minimizing adverse impact may differ somewhat from a selection strategy aimed at maximizing mean predicted performance (Sackett & Roth 1996).

Personality

Personality Taxonomies and Constructs The Five Factor Model (FFM), consisting of Extraversion, Agreeableness, Conscientiousness, Neuroticism (Adjustment), and Openness to Experience (see Wiggins & Trapnell 1997), enjoys considerable support. Factor analysis supports the robustness and generalizability of the FFM across different theoretical frameworks, assessments, rating sources,

and cultures (see Hogan & Ones 1997b, Saucier & Goldberg 1998, Wiggins & Trapnell 1997). The model is useful for summarizing information and guiding theory and research (e.g. Mount & Barrick 1995, Tokar et al 1998).

The FFM yields information about the higher-order factor structure of personality; however, it ignores, confounds, or otherwise obscures understanding of variables combined into five broad factors (Hough 1997, 1998b; Hough & Schneider 1996). FFM factors contain facets with high and low criterion-related validities, diluting the criterion-related validity of the factors. A review of meta-analyses concluded the FFM factors do not correlate highly with job performance (Matthews 1997). As alternatives, researchers are turning to nonhierarchical models such as the circumplex (Plutchik & Conte 1997) and other hierarchies. Hough (1997, 1998b) argues for a more refined taxonomy, distinguishing achievement from conscientiousness and extraversion and affiliation from extraversion. Meta-analyses demonstrate the importance of these distinctions for predicting managerial performance (Hough et al 1998) and sales performance (Vinchur et al 1998). Ghiselli's (1966) personality framework was compared with the FFM framework in the Barrick & Mount (1991) meta-analysis: the median uncorrected validity in the Ghiselli meta-analysis was 0.24; the highest mean uncorrected validity for an FFM variable in the Barrick & Mount (1991) study was 0.15 (Hough 1997). Several important personality constructs not within the FFM have been used for predicting work behavior:

Emotionality Emotionality, or affectivity, consists of two bipolar dimensions at the most general level: negative-positive and aroused-unaroused (Averill 1997, Russell & Carroll 1999). A state measure of emotionality did not correlate with job performance, but a dispositional measure did (Wright & Staw 1999). In social welfare workers, negative affectivity correlated positively with emotional exhaustion or "burnout," and emotional exhaustion correlated negatively with job performance (Wright & Cropanzano 1998).

Social Competence Social competence is a compound variable consisting of social insight, social maladjustment, social appropriateness, social openness, social influence, warmth, and extraversion (Schneider et al 1996). Reliable self-report measures of social insight (e.g. Gough 1968) and empathy (e.g. Hogan 1968) have a long history, as do situational judgment measures of social intelligence (e.g. Moss et al 1955). Variables subsumed under social competence might increment predictive validity for criteria emphasizing interpersonal effectiveness.

Conscientiousness Many claim that conscientiousness, a FFM factor, is a valid predictor across organizations, jobs, and situations (Hogan & Ones 1997b, Mount & Barrick 1995, Salgado 1997a, 1998). Others question this wholesale conclusion (Hough 1997, 1998b; Robertson & Callinan 1998). Whether or in what direction conscientiousness predicts performance obviously depends on the criterion construct and how conscientiousness is defined and operationalized. Based on the

Hogan & Ones' (1997) definition of conscientiousness as conformity and socially prescribed impulse control, conscientiousness would likely not predict performance across organizations, jobs, or situations in which creativity or innovation is important (Hough 1997, 1998b; Hough et al 1998).

Integrity Tests Meta-analyses of relations between integrity tests and FFM variables indicate that integrity tests are compound variables consisting primarily of conscientiousness, agreeableness, and adjustment (Ones & Viswesvaran 1998b). Four themes account for most of the variance in overt and personality-based integrity tests: punitive attitudes, admissions of illegal drug use, reliability, and admissions of theft (Hogan & Brinkmeyer 1997). Importantly, integrity tests differ from tests of deception (Murphy & Luther 1997). Sackett & Wanek (1996) provided an insightful and thorough review of integrity testing that dealt with construct- and criterion-related validity evidence; moderator variables; social desirability and applicant reactions; and legal, professional, and governmental evaluations. Meta-analysis indicates that integrity and conscientiousness tests usefully supplement general cognitive ability tests when predicting overall job performance (Schmidt & Hunter 1998). Converging evidence exists for the construct, criterion-related, and incremental validity of integrity tests (Miner & Capps 1996, Ones & Viswesvaran 1998b), but considerable variability may accompany the overall findings.

For example, criteria often used in integrity-testing research are problematic. Self-report or admission of counterproductive behavior confounds reliability with validity for overt integrity tests and underestimates the extent of counterproductive work behaviors, as do more direct measures (e.g. detected theft). Moreover, counterproductivity is not a unitary construct (Ashton 1998, Sackett & Wanek 1996). Two recent meta-analyses summarized correlations between integrity test scores and two facets of counterproductivity: reported drug abuse [$r = 0.21$ (Schmidt et al 1997)] and number of job-related accidents [$r = 0.52$ (Ones & Viswesvaran 1998b)]. Workplace violence, a facet of counterproductivity, tends to be better predicted by narrow measures such as aggression and violence scales than by broad honesty tests. Meta-analysis indicates that the validity for predicting workplace violence is higher for violence scales than for integrity tests [$r = 0.48$ vs 0.26 (Ones et al 1994)]. A physical-aggression measure predicted aggressive penalty minutes ($r = 0.33$) in high school hockey games, but not nonaggressive penalty minutes ($r = 0.04$) (Bushman & Wells 1998).

Customer Service Orientation Meta-analysis finds that customer service orientation is a compound variable consisting of agreeableness ($r = 0.70$), adjustment ($r = 0.58$), and conscientiousness ($r = 0.43$) (Ones & Viswesvaran 1996), and customer service scales correlate with performance in customer service jobs ($r = 0.31$) (Frei & McDaniel 1998).

Core Self-Evaluation Core self-evaluation is a compound variable consisting of self-esteem, generalized self-efficacy, locus of control, and emotional stability (Judge et al 1998). Meta-analysis estimated the validity of core self-evaluation for predicting job performance at 0.30 (TA Judge & JE Bono, submitted for publication). A much larger meta-analysis suggests that self-efficacy, a facet of core self-evaluation, correlates higher with job performance than does core self-evaluation as a whole [$r = 0.38$ (Stajkovic & Luthans 1998)].

Other Meta-Analyses Meta-analysis of the Five Factor Model validities in U.S. studies from 1992 through 1997 produced results similar to past U.S. meta-analyses (Anderson & Viswesvaran 1998), but meta-analyses involving only European samples produced somewhat different results (Salgado 1997a, 1998). Both conscientiousness and emotional stability correlated positively with job performance across occupational groups, and both contributed incremental variance beyond general mental ability in predicting overall job performance. Other meta-analyses of validities of FFM factors indicate that agreeableness, as well as conscientiousness and emotional stability, predicts performance in jobs involving interpersonal interaction (Mount et al 1998). Managerial potential scales predict overall managerial job performance [$r \approx 0.40$ (Ones et al 1998)], although meta-analytic validities tend to be low for FFM factors predicting overall managerial performance (Hough et al 1998). Some FFM facets had much higher validities, shedding light on how facet-level variables might combine to form managerial potential scales with high criterion-related validity. Many of these meta-analytic researchers corrected study correlations for predictor range restriction by using national norm SDs from personality test manuals. This practice of using norm SDs appears warranted, because SDs of job applicants on personality measures are about 2%–9% less than those based on national norms (Ones & Viswesvaran 1999).

Conditional Reasoning Pioneered by James (1998, 1999), conditional reasoning assumes that individuals' personalities are differentiated by the type of logical reasoning used to justify their actions. For example, people who score high on achievement motivation tend to attribute success to internal rather than external sources and consider demanding tasks challenging rather than frustrating. An achievement conditional reasoning scale correlated positively with scholastic criteria, in-basket performance, and other achievement scales (James 1998, Migetz et al 1999a, Smith et al 1995). An aggression conditional reasoning scale correlated negatively with overall job performance (Hornick et al 1999, James 1998) and positively with counterproductive work behavior (Burroughs et al 1999, Migetz et al 1999b, Patton et al 1999). This approach appears to overcome many problems related to intentional distortion.

Intentional Distortion Not surprisingly, meta-analysis shows large mean-score differences between honest and directed-faking conditions (Viswesvaran & Ones

1999). The amount of distortion in naturally occurring applicant settings is uncertain however. Rosse et al (1998) found that applicants ($N = 197$) scored on average 0.69 SD higher than incumbents ($N = 73$) on FFM facet-level scales. In contrast, three separate samples involving over 40,500 applicants and over 1700 incumbents found significantly less distortion on similar scales (Hough 1998a). Ability and motivation to fake may be key determinants in the amount of distortion found in applicant settings (Snell et al 1999). Meta-analysis indicates that explicit warnings not to distort do reduce distortion [0.23 SD (Dwight & Donovan 1998)].

A slew of recent studies has investigated intentional distortion effects on criterion-related validities. Many assert that distortion does not tend to moderate, mediate, suppress, or attenuate the criterion-related validities of personality scales (Barrick & Mount 1996; Hogan 1998; Hough 1997, 1998a,b; Ones & Viswesvaran 1998b,c; Ones et al 1996). Others, such as Douglas et al (1996), Snell & McDaniel (1998), and Zickar & Drasgow (1996), contend that distortion seriously reduces criterion-related validity. Hough (1998a) resolved the apparent conflict by stratifying results by employment setting. In directed-faking settings, self-report scale scores have dramatically lower criterion-related validities than those obtained in applicant or incumbent settings; in applicant settings, self-report scale scores have the same or slightly lower criterion-related validities than those obtained from job incumbents in research-only settings. Similarly, construct validity may be negatively affected in directed-faking studies (Ellingson et al 1999a), but the effect does not seem to be as serious in applicant settings (Collins & Gleaves 1998, Ellingson et al 1999b, Ones & Viswesvaran 1998c).

Coaching individuals on personality tests potentially threatens the effectiveness of traditional social desirability scales. Subtle items can be more resistant to coaching and distortion than obvious items (Alliger et al 1996). Theory-driven approaches to scale development and validity data to refine items produce subtle items resistant to distortion and with excellent validity (Gough 1994).

Race and Ethnic Background Similar personality factor structures for Blacks and Whites are found (Collins & Gleaves 1998). Meta-analyses of White, Black, Hispanic, and Native American groups indicate minimal group mean differences for three overt integrity tests (Ones & Viswesvaran 1998a) and for FFM factors, although Hispanics scored 0.60 SD higher than Whites on social desirability scales (Hough 1998b). Personality variables have little adverse impact against minorities, if any. Score correction strategies using social desirability scales to correct distortion in content scale scores might affect Hispanics more than others.

Multiple Predictor Domains

Applied psychology has long postulated that ability and motivation interact in predicting job performance: High performers must have the requisite ability and effort to do the job; neither ability nor effort alone suffices. However, regression

analyses of data across job samples, performance criteria, and different ability and motivation measures have yielded nonexistent or very slight incremental ability-motivation interaction effects (Sackett et al 1998). A combined meta-analysis of ability-personality correlations and review of the empirical findings on ability-interest and personality-interest relationships have produced an integrated model identifying four categories or “trait complexes”: science/math, clerical/conventional, social, and intellectual/cultural (Ackerman & Heggstad 1997). A vocational interest structure was empirically linked to Gottfredson’s (1986) ability-based job classification framework, yielding four similar categories (Oswald & Ferstl 1999). Linking different predictor domains to a common job classification framework may clarify the constellations of predictors that are useful for selection into various types of jobs (see Arthur & Bennett 1995, Hatrup et al 1998, Johnson et al 1997, Mael et al 1996b, Vinchur et al 1998) and may improve synthetic validity efforts.

ASSESSMENT METHODS

Interview

Interview Structure Compared with unstructured interviews, structured employment interviews define content more explicitly. Their successes are therefore more likely to replicate, and they are better analyzed and meta-analyzed to determine their transportability to other jobs and work settings. Nonetheless, organizations still prefer unstructured interviews by a wide margin (Graves & Karren 1996). A comprehensive review indicates that various components of structured interviews influence the interview’s psychometric properties, legal defensibility, and applicant/interviewer reactions (Campion et al 1997). Recent research offers at least three other compelling reasons for structuring the core of the interview (see also the Dipboye 1997 review).

Reliability Structured interviews tend to have higher interrater reliability than unstructured interviews. Meta-analysis reports average interrater reliabilities of 0.67 for high structure vs 0.34 for low structure (Conway et al 1995).

Standardization Standardized interviews place more burden on the instrument than any particular interviewer’s interviewing and assessment skills. A highly standardized situational interview, in which applicants respond to hypothetical critical work incidents, can be less susceptible to rating biases (Kataoka et al 1997). Computerized phone interviews have efficiently obtained standardized applicant information, with validity coefficients similar to those for traditional interviews (Schmidt & Rader 1999). Interviewer experience and training further standardize the interview (Conway et al 1995, Campion et al 1997, Huffcutt & Woehr 1999). Training in note taking improves attention, encoding, recall, and

evaluation of interview information focusing on work behaviors (Burnett et al 1998).

Fairness Structured interviews treat applicants in a consistent manner. Mean differences by race are more likely reduced in highly structured interviews containing content related to noncognitive constructs, especially for high-complexity jobs (Huffcutt & Roth 1998). Court outcomes on disparate impact and disparate treatment have favored organizations high on three interview characteristics: standardized administration, high job relatedness, and multiple raters (Williamson et al 1997).

Interpersonal and Nonverbal Behavior Interviewee characteristics (e.g. gaze, hand movement, and physical attractiveness) can predict several dimensions of managerial effectiveness (leadership, teamwork, and planning/organization), even when characteristics are coded independently of the content (Burnett & Motowidlo 1998; see also Motowidlo & Burnett 1995). The convergent and discriminant validities between interpersonal behavior in the interview and different job performance criteria should be considered.

Assessment Centers

Assessment centers (ACs) have long been haunted by evidence of content- and criterion-valid ratings lacking construct validity (Arthur et al 1999, Spsychalski et al 1997, Woehr & Arthur 1999). Confusion about the constructs being measured, rating errors, type and form of rating procedures, and participant inconsistencies in behavior across exercises are possible explanations (Arthur & Tubre 1999, Guion 1998). Features improving AC ratings include having (a) only a few conceptually distinct constructs, (b) concrete, job-related construct definitions, (c) frame-of-reference assessor training with evaluative standards, (d) cross-exercise assessment, and (e) several psychology-trained assessors (Lievens 1998, Woehr & Arthur 1999). An AC designed and implemented on the basis of research and professional-practice guidelines (see Task Force on Assessment Center Guidelines 1989) produced construct-valid AC ratings. Generalizability theory facets associated with individuals and constructs accounted for 60% of the total variance, and facets associated with assessors and exercises accounted for 11% of the total variance (Arthur et al 1999).

ACs are expensive and prone to cost-benefit comparisons with other predictors. AC ratings have significant incremental validity over personality variables, and vice versa, when predicting managerial performance (Goffin et al 1996, with $N = 68$). AC ratings also have incremental validity over cognitive ability, although some AC exercise validities are founded primarily on their cognitive component (Goldstein et al 1998). By using policy capturing and meta-analysis, overall AC ratings have been predicted from both cognitive ability and personality variables [$R = 0.77$ (Collins et al 1999)]. The key question is whether policy-

captured predictors predict job performance better than AC ratings (Howard 1999). Black-White mean exercise score differences have ranged from 0.03 to 0.40 SDs, with Blacks scoring lower (Goldstein et al 1999). Exercises emphasizing interpersonal skills more than cognitive ability have resulted in less or no adverse impact for Blacks (Goldstein et al 1999, Bobrow & Leonards 1997).

Biodata

Several researchers have focused on much-needed construct-oriented approaches to biodata (biographical information). CN MacLane (submitted for publication) refined the federal government's Individual Achievement Record biodata scales, which measure social and cognitive abilities (see Gandy et al 1994). Personality-based biodata scales predict leadership (Stricker & Rock 1998) and life insurance sales (McManus & Kelly 1999). A review of eleven studies examined the validity of biodata scales based on Mumford & Stokes' (1992) rigorous construct-oriented item-generation procedures. Scales were content and construct valid, with criterion-related validities similar to those for traditional empirical keying (Mumford et al 1996).

Biodata theory relies heavily on the principle that past behavior is the best predictor of future job performance (i.e. the "consistency" principle). Failure or negative life experiences also explain why biodata predict performance (Russell 1999). Moxie (i.e. courage or "ego-resiliency") may moderate how negative life experiences influence development and subsequent job performance (Dean et al 1999, Muchinsky 1999). "Negative" is often in the eye of the beholder, however. Both positive and negative responses to elements of a broader life-events taxonomy may be needed.

Conclusions about the effectiveness of rational, empirical-keying, and factor-analytic biodata scale development strategies are inconsistent. Rational scales have predicted sales performance at least as well as empirically keyed and factor-analytic scales (Stokes & Searcy 1999). Factor-analytic and rational scales have predicted several customer service criteria much better than empirical keying (Schoenfeldt 1999). A meta-analysis found similar levels of criterion-related cross-validities across all three scale construction strategies (Hough & Paullin 1994). Rational, empirical-keying, and factor-analytic scale strategies need not be executed and compared separately; the strategies may iteratively inform one another.

Regarding the generalizability of biodata, the reliability, factor structure, and validity of biodata keys appear stable across two English-speaking countries (Dal-essio et al 1996). Validity for a biodata inventory predicting managerial progress generalized across organizations and educational levels (Carlson et al 1999). "Contemporary" items tend to be more valid than "future/hypothetical" or "historical" items, and items that ask respondents about others' opinions of them are more valid than direct self-report items (Lefkowitz et al 1999). Meta-analysis found that the amount and task-level specificity of work experience correlated

most highly with job performance (Quiñones et al 1995). Rational biodata scales may produce inadequate levels of validity for separate racial/ethnic groups, but empirical item analysis can be used to produce a scale valid across groups (Schmitt & Pulakos 1998, Schmitt et al 1999). Whitney & Schmitt (1997) discovered differential item functioning between racial subgroups in about one quarter of the biodata items they examined.

MEASUREMENT ISSUES AND VALIDATION STRATEGIES

Measurement Issues

Many published studies in personnel selection continue to suffer from low statistical power due to small sample sizes (Mone et al 1996, Salgado 1998). Confidence intervals directly convey the impact of sample size on the accuracy of statistics (Hunter 1997). A computer program calculates confidence intervals on correlation coefficients corrected for measurement unreliability and range restriction (Salgado 1997b). Formulas are accurate in the large-sample case; for the small-sample case, the bootstrap method (i.e. calculating a distribution of correlations by resampling data with replacement) has generated accurate confidence intervals for ability-training performance validities (Russell et al 1998). Instead of confidence intervals, Murphy & Myors (1999) provide noncentral *F* tables and real-world examples that test minimum-effect null hypotheses for *t*-tests, correlations, and ANOVAs.

Measurement error variance can distort patterns of research results and mislead conclusions. Circumstances in 26 applied-research situations show when to correct validities for such error by using the appropriate reliability coefficient (Schmidt & Hunter 1996). Structural equation modeling can test the statistical significance of corrected correlation coefficients and the difference between two such correlations (Hancock 1997). Recent studies considered maximizing the reliability of a linear composite by weighting the constituent variables as a function of their reliabilities (Cliff & Caruso 1998, Li et al 1996, Raykov 1997, Wang 1998).

Estimation formulas for the population validity (expected prediction in the entire population) and cross-validity (expected prediction in other independent samples) were thoroughly reviewed (Raju et al 1997), and formulas were compared in a Monte Carlo study with data from a large sample of Air Force enlistees (Raju et al 1999). Generally, the Ezekiel, Smith, and Wherry procedures all provided good squared population validity estimates, and Burket's formula best estimated the squared population cross-validity. In stepwise regression, results showed that the sample size-to-predictor ratio had to be relatively large (10:1 at least) to yield good cross-validity estimates. All cross-validity estimation formulas performed similarly well; none was clearly superior (Schmitt & Ployhart

1999). Instead of stepwise regression, researchers often judge the relative size of regression weights to decide on the important variables within a particular model. A new type of relative importance weights enhances the interpretability of regression results when predictors are highly intercorrelated (Johnson 1999).

Finally, personnel selection research must often deal with missing data. Generally, pairwise deletion is better than listwise deletion, and estimating missing scores via regression is better than substituting missing scores with their unconditional mean (Roth et al 1996). Personnel selection would profit from understanding substantive processes underlying “missingness.” Job promotions or transfers, emotional exhaustion from work, and organizational redesign all may reflect different types of longitudinal and cross-sectional attrition processes.

Meta-Analysis and Validity Generalization

Meta-analysis has had a far-reaching impact on policymaking, real-world application, and academic research (Hunter & Schmidt 1996) and is a useful quantitative tool for summarizing large bodies of personnel selection research (Murphy 1997). The meta-analytic mean effect size across studies (e.g. mean correlation coefficient) tends to be fairly accurate. In contrast, estimates of the variance of effect sizes (after correcting for statistical artifacts) can deviate from their actual population values by practically significant amounts (Oswald & Johnson 1998), which may affect meta-analytic conclusions about selection research. Variance estimates can be downwardly biased if one ignores the fact that individuals were selected on a variable correlated with the variables in the meta-analysis [i.e. incidental range restriction (Aguinis & Whitehead 1997)]. Statistical homogeneity tests of the variance have low statistical power and tend to discourage the search for moderator effects (Sánchez-Meca & Marín-Martínez 1997).

Meta-analysis is one of many lines of evidence supporting the use of a selection test. In addition to meta-analytically averaging correlations, researchers might consider what predicts group or organizational mean differences on predictors and criteria (Ostroff & Harrison 1999). Hierarchical linear modeling (HLM) has been advocated for this purpose; Hofmann (1997) presented some practical organizational examples. Synthetic validity evidence may provide additional validity information (Hoffman & McPhail 1998).

EVALUATION OF SELECTION SYSTEMS

Differential Prediction

A selection test with equal regression slopes across subgroups (e.g. race, gender, and age groups) does not necessarily measure a latent construct the same way across subgroups. Only data conforming to a special set of mathematical constraints will show a lack of differential prediction and latent construct equivalence simultaneously (Millsap 1995, 1997). To the extent that these constraints

are violated, this finding may challenge previous selection research suggesting a lack of differential prediction (e.g. cognitive ability research).

Aguinis & Stone-Romero (1997) examined range restriction effects on the statistical power of moderated multiple regression for detecting differential prediction. Range restriction often mistakenly led to concluding no differential prediction when validity differences were moderate (0.4 correlation units). Larger validity differences were detected, and smaller differences were not, regardless of range restriction. Computer software for estimating the statistical power of differential prediction in moderated multiple regression is available (Aguinis & Pierce 1998b). Future software could incorporate violations of the assumption of equal error variances between subgroup regression models. Assumption violations occur when larger subgroup sample sizes are paired with the smaller subgroup validity coefficient (e.g. Aguinis & Pierce 1998a, DeShon & Alexander 1996), but for some organizational data, violations are not severe enough to affect statistical inferences from moderated multiple regression (Oswald et al 1999). Error variance in the independent variables affects differential prediction (Terris 1997), and errors-in-variables regression addresses this problem. Errors-in-variables regression detected differential prediction more accurately than moderated multiple regression when reliability coefficients were >0.65 and sample sizes were >250 (Anderson et al 1996).

Adverse Impact

Common sense might assert that combining low-adverse-impact predictors with a high-adverse-impact predictor improves adverse impact over using the high-adverse-impact predictor alone. However, Sackett & Ellingson (1997) presented tables illustrating reduced subgroup differences for some composites but increased differences for others. Composite measures invariably reduce subgroup differences less than expected. In particular, the “four-fifths rule” set out in the Uniform Guidelines (U.S. Equal Employment Opportunity Commission et al, 1978) is usually met only under very high selection ratios (≥ 0.90) or very slight composite group-mean differences ($d \leq 0.20$).

Similar conclusions came from meta-analytic estimates of criterion-related validities for a cognitive ability measure, a noncognitive composite (interview, biodata, and conscientiousness), and a composite of both, all independently predicting overall job performance. Selection batteries excluding cognitive ability almost always satisfied the four-fifths rule; batteries that included cognitive ability alone or in a composite almost never satisfied the rule (Schmitt et al 1997). Findings are echoed in two large-sample studies of firefighter and police officer job applicants selected on cognitive ability and personality measures (Ryan et al 1998a). Bobko et al (1999) updated the Schmitt et al (1997) meta-analytic matrix, discovering that even a noncognitive composite could violate the four-fifths rule when selection ratios were $\leq 50\%$. Group-mean differences on noncognitive composites should be determined before assuming they reduce adverse impact. Also,

adverse impact is partly a function of the criteria chosen and how they are weighted (Hattrup et al 1997).

Alternative forms of administration potentially reduce adverse impact and increase overall validity. Videotaped versions of situational judgment test material can reduce adverse impact and have greater face validity than paper-and-pencil versions. Presumably, validity improves because the content in the video format is preserved, and irrelevant variance related to reading comprehension is removed (Chan & Schmitt 1997). Similar conclusions come from comparing a video-based ability test to a traditional ability test (Pulakos & Schmitt 1996).

Banding Given subgroup differences on selection tests (most notably cognitive ability tests), statistical banding can fulfill important goals such as maintaining workforce diversity and improving perceptions of process and outcome fairness in a selection procedure (Truxillo & Bauer 1999). Linearly transforming bands on predicted criterion scores into bands on predictor scores (Aguinis et al 1998) may improve banding because criterion differences tend to matter more to organizations than predictor differences (see Campbell 1996, Gottfredson 1999). Further advances in banding might consider that reliable predictors tend to produce smaller statistical bandwidths, in which differences larger than the band may still not be practical differences or translate into practical criterion differences.

Utility

A few recent studies focus on how utility information communicates the effectiveness of a selection system to organizational stakeholders. Whyte & Latham (1997) replicated the counterintuitive results of Latham & Whyte (1994), discovering that communicating positive utility information can actually decrease managers' intentions to use a selection system. Utility information may prove beneficial as supplementary information, not delivered face-to-face as a "hard sell" (Cronshaw 1997). Communicating utility in terms of multiple outcomes (e.g. dollars, job performance, and organizational effectiveness) may lead to greater acceptance of the utility message by different stakeholders (Roth & Bobko 1997). Considering and balancing all particular stakeholder positions is difficult but beneficial to organizations (Austin et al 1996).

Applicant Reactions

Positive applicant reactions increase the chances of hiring the best applicants, facilitate the ability to recruit effectively, avoid the possibility of costly litigation, and contribute to the organization's reputation (Gilliland & Steiner 1999, Ryan & Greguras 1998, Schmitt & Chan 1999). Selection systems can be viewed as socialization mechanisms imparting job information to applicants and affecting their work-related thoughts, attitudes, and behaviors (Anderson & Ostroff 1997). Managing applicant reactions does not imply making the organization attractive to all individuals; accurate perceptions can lead to applicant withdrawal. Black-

White test-taking attitude differences may not affect group differences in applicant withdrawal (Schmit & Ryan 1997), but they might, which would add to adverse-impact concerns (Chan 1997).

Chan et al (1998) discovered that (a) pretest reactions affected cognitive ability test performance, and test performance affected posttest reactions, and (b) personality test performance was unrelated to either pretest or posttest reactions. Integrity test results paralleled the personality test conclusions, and the overt integrity test had greater perceived job relatedness than the personality-based test (Whitney et al 1999).

Test outcome (passing or failing) contributes strongly to subsequent applicant reactions. Applicants passing a test for a clerical position rated organizational attractiveness, intentions to work for the organization, and test fairness higher than their initial reactions (Bauer et al 1998). Test outcome affects the perceived fairness of the hiring decision much more than the selection ratio (Thorsteinson & Ryan 1997). Ployhart et al (1999) determined that fairness perceptions of cognitive and job knowledge tests increased with a positive selection outcome and with sensitively conveyed personal and procedural information regardless of outcome. Selected applicants' self-perception improved with personal or procedural information; rejected applicants' self-perception declined. Sensitive explanations amplified this result, implying that providing information about the selection procedure would increase fairness and organizational perceptions but be counterbalanced by lower self-perceptions for rejected applicants.

Invasiveness of personnel selection measures has been investigated. Verifiable, impersonal, and face-valid biodata items tend to be perceived as less invasive, especially for individuals who understand the general purpose of biodata (Mael et al 1996a). Ways to obtain potentially invasive information without violating applicants' needs for privacy have been offered [e.g. explaining the job relevance of the item (Mael 1998)].

EMERGING TOPICS

Team Member Selection

Organizations increasingly use team-based structures for organizing, motivating, and performing work (see Guzzo & Salas 1995, Howard 1995, Kehoe 1999, Klimoski & Zukin 1999, Kraut & Korman 1999b, O'Neil 1997, Sundstrom 1999). Much has been learned about factors affecting team performance and effectiveness (see reviews by Cohen & Bailey 1997, and West & Allen 1997), but more work remains for personnel selection. Selection systems need to consider differences between team selection and traditional selection methods, particular work team circumstances (task type, role differentiation, and resources), and selection into new versus preexisting work teams.

Individual characteristics and types of tasks interact within a team to influence team performance and effectiveness. On a creative problem-solving task, a mid-range of extraverts appears best; too many or too few depress performance ratings slightly. Conscientiousness did not predict team performance on the creative task (Barry & Stewart 1997), as Hough (1992) found at the individual level of analysis. In other team studies (LePine et al 1997, Barrick et al 1998), conscientiousness did predict task performance. Given a designated leader and group members each with unique expertise, team decision-making accuracy over time tends to be best when all members are high in conscientiousness and general cognitive ability (LePine et al 1997). Barrick et al (1998) found that conscientiousness, general cognitive ability, and extraversion all predicted overall team performance ratings in manufacturing work teams, where team members contributed independently to the outcome. Team agreeableness predicted teamwork for those tasks in which intergroup conflict was possible. One disagreeable team member was often enough to disrupt team performance, indicating the importance of interpersonal skills when selecting for some teams. Negotiation, an interpersonal skill, was validly measured with a simulation exercise (O'Neil et al 1997a,b). Significant advances in team selection research await good taxonomies of "team difference" variables (the analog of individual-difference variables for individuals) and situational variables relevant to teams.

Cross-Cultural Selection Issues

With their expanding global markets, culturally diverse work teams, and expatriate work assignments, international and multinational organizations place new demands on selection processes and measurement tools. Validities of domestic selection instruments may not generalize to international sites, because different predictor and criterion constructs may be relevant, or, if the constructs are the same, the behavioral indicators may differ. Interpersonal skill, open-mindedness, and adaptability are important factors for expatriate success, and family situation is the most commonly cited reason for failure (Arthur & Bennett 1995, Nyfield & Baron 1999). The vast majority of companies base their expatriate selection decisions on technical competence alone (Aryee 1997), so finding a very high failure rate among expatriates is unsurprising [between 15% and 40% (Shackleton & Newell 1997)]. A clear need for improving expatriate selection exists.

Several personality inventories originally developed in English have demonstrated similar psychometric properties across languages and cultures (see Katigbak et al 1996, McCrae & Costa 1997, Nyfield & Baron 1999). The International Committee on Test Standards produced a set of stringent standards for translating tests into another language (see Hambleton 1999). Psychologists from many different cultures might be involved in all phases of inventory development and validation, a strategy used to develop the Global Personality Inventory (Schmit et al 1999). Cultural variables likely moderate the validity of selection procedures. The House et al (1997) review concluded that Hofstede's (1980) four constructs

(power distance, uncertainty avoidance, individualism vs collectivism, and masculinity vs femininity) described and differentiated cultures most usefully.

PROFESSIONAL, LEGAL, AND ETHICAL STANDARDS

Professional

Three initiatives sponsored by the American Psychological Association provide guidelines and policies regarding test-taker rights and responsibilities, test standards, and test-user qualifications. First, the American Psychological Association is in the final stages of approving 10 test-taker rights and corresponding responsibilities (Joint Committee on Testing Practices 1999). Second, the “Standards for Educational and Psychological Testing” (unpublished manuscript) revises and updates standards for psychological tests and for psychological measurement in general. All parties involved in employment decisions involving psychological assessment should become familiar with these new standards. Third, the “Test User Qualifications” document is currently under review (Fox 1999). These three documents will impact personnel selection practices and have significant legal and ethical implications.

Legal

Over the past 20 years or so, more employment litigation has been brought under common-law torts than under federal or state equal employment opportunity statutes (Highberger 1996). Nevertheless, legal challenges to personnel selection decisions are often based on the Civil Rights Acts, Americans with Disabilities Act (ADA), and Age Discrimination in Employment Act. In the vast majority of these cases, the Equal Employment Opportunity Commission is typically not involved (Sharf & Jones 1999), although commission guidelines provide important compliance information for the public and the courts. Unstructured interviews account for the majority of federal court cases involving selection tools, followed by cognitive ability tests, and physical ability tests; together, they were judged to be discriminatory in about 40% of the cases, with cognitive tests faring somewhat better (Terpstra et al 1999). For important practical guidance and discussions of the many issues, risks, and myths regarding fair employment as well as trends in employment litigation, see Barrett (1996, 1998), Jeanneret (1998), and Sharf & Jones (1999).

The U.S. Supreme Court significantly limited the scope of ADA, ruling that impairments should be evaluated in their corrected or “mitigated” state (*Sutton vs United Airlines* 1999, *Murphy vs United Parcel Service* 1999). The U.S. Equal Employment Opportunity Commission (1999) answered 46 frequently asked questions pertaining to employers’ legal obligations and the rights of the disabled in both the application and employment settings. Tippins (1999) provided practical ADA guidance for several testing scenarios, and Bruyère (1999) outlined

the psychologist's role in upholding ADA provisions in all phases of the employment process.

Affirmative Action and Reverse Discrimination Although individuals differ greatly in their perceptions of affirmative action programs (Kravitz et al 1997), negative consequences consistently occur when employees or applicants believe hiring is based on group membership rather than merit (Heilman 1996, Heilman et al 1998, Kravitz et al 1997, Stanush et al 1998). Affirmative action programs are associated with slight improvement in employment conditions for women and racial minorities and appear to have virtually no effect on organizational effectiveness (Kravitz et al 1997).

No general agreement exists on how to prevent discrimination or remedy past discrimination (Campbell 1996). Reverse-discrimination court cases have clarified that race or other job-irrelevant class membership can not be used when making employment-related decisions. Therein lies a conflict: The Uniform Guidelines indicate that organizations should seek out valid non- or less-discriminating predictors, yet developing a selection system with such measures requires attention to class membership. In *Hayden vs County of Nassau* (1999), the claim that an entrance exam designed to minimize discriminatory impact on minority job candidates necessarily discriminated against nonminority job candidates was ruled to be without merit. This case sets a precedent in affirming the reasonableness of designing selection systems to minimize adverse impact against protected groups.

Ethical

Ethical issues in personnel selection are complex, context-specific, and relative to each concerned party. We refer the reader to two important new sources in the field. Lowman (1998) authored an updated ethics casebook for human resource professionals practicing within organizations, and Jeanneret (1998) discussed ethical issues involved in individual assessment, detailing the responsibilities for both assessors and organizations.

PARTING REMARKS

New areas in personnel selection are unfolding, and traditional areas continue to improve. (a) Greater conceptual and methodological attention has been devoted to understanding and predicting how organizationally relevant criteria might change over time. Given the present and future state of rapid change in the world of work, this line of research is critically important for improving personnel selection and overall organizational effectiveness. (b) Personality theory and measurement within a personnel selection context have burgeoned. New personality constructs and compound constructs of well-known traits are being brought into

the fold. (c) Applicant reactions to personnel selection procedures have been energetically studied. (d) Team member and cross-cultural selection issues have drawn greater research attention. (e) Refined taxonomic structures are being developed across many different domains in personnel selection, both from the worker and work perspectives. We predict that selection systems will become more complex as a consequence of all this work; selection systems will mirror today's realities and prove to be more effective and rewarding for individuals and organizations alike. We also predict that Guion's (1998) book on personnel selection will fast become a classic. It is readable, practical, thoughtful, and thorough, and it assures its readers that much has been learned in personnel selection. Although the turn of the millennium marks a distinct ending and beginning, personnel selection theory and practice remain in constant process, looking toward the future while remembering the past.

Visit the Annual Reviews home page at www.AnnualReviews.org.

LITERATURE CITED

- Ackerman PL, Heggstad ED. 1997. Intelligence, personality, and interests: evidence for overlapping traits. *Psychol. Bull.* 121:219–45
- Aguinis H, Cortina J, Goldberg E. 1998. A new procedure for computing equivalence bands in personnel selection. *Hum. Perform.* 11:351–65
- Aguinis H, Pierce CA. 1998a. Heterogeneity of error variance and the assessment of moderating effects of categorical variables: a conceptual review. *Organ. Res. Methods* 1:296–314
- Aguinis H, Pierce CA. 1998b. Statistical power computations for detecting dichotomous moderator variables with moderated multiple regression. *Educ. Psychol. Meas.* 58:668–76
- Aguinis H, Stone-Romero EF. 1997. Methodological artifacts in moderated multiple regression and their effects on statistical power. *J. Appl. Psychol.* 82:192–206
- Aguinis H, Whitehead R. 1997. Sampling variance in the correlation coefficient under indirect range restriction: implications for validity generalization. *J. Appl. Psychol.* 82:528–38
- Alliger GM, Lilienfeld SO, Mitchell KE. 1996. The susceptibility of overt and covert integrity tests to coaching and faking. *Psychol. Sci.* 7:32–39
- Anderson GD, Viswesvaran C. 1998. *An update of the validity of personality scales in personnel selection: a meta-analysis of studies published between 1992–1997*. Presented at Annu. Meet. Soc. Ind. Organ. Psychol., 13th, Dallas
- Anderson LE, Stone-Romero EF, Tisak J. 1996. A comparison of bias and mean squared error in parameter estimates of interaction effects: moderated multiple regression versus errors-in-variables regression. *Multivariate Behav. Res.* 31:69–94
- Anderson N, Herriot P, eds. 1997. *International Handbook of Selection and Assessment*, Vol. 13. Chichester, UK: Wiley. 652 pp.
- Anderson N, Ostroff C. 1997. Selection as socialization. See Anderson & Herriot 1997, pp. 413–40
- Arthur W, Bennett W. 1995. The international assignee: the relative importance of factors perceived to contribute to success. *Pers. Psychol.* 48:99–114
- Arthur W, Doverspike D, Barrett GV. 1996. Development of a job analysis-based procedure for weighting and combining con-

- tent-related tests into a single test battery score. *Pers. Psychol.* 49:971-85
- Arthur W, Tubre T. 1999. The assessment center construct-related validity paradox: a case of construct misspecification? See Quiñones 1999
- Arthur W, Woehr DJ, Maldegen R. 1999. Convergent and discriminant validity of assessment center dimensions: a conceptual and empirical re-examination of the assessment center construct-related validity paradox. *J. Manage.* In press
- Aryee S. 1997. Selection and training of expatriate employees. See Anderson & Herriot 1997, pp. 147-60
- Ashton MC. 1998. Personality and job performance: the importance of narrow traits. *J. Organ. Behav.* 19:289-303
- Austin JT, Klimoski RJ, Hunt ST. 1996. Dilemmas in public sector assessment: a framework for developing and evaluating selection systems. *Hum. Perform.* 9:177-98
- Averill JR. 1997. The emotions: an integrative approach. See Hogan et al 1997, pp. 513-41
- Barrett RS, ed. 1996. *Fair Employment Strategies in Human Resource Management.* Westport, CT: Quorum Books. 319 pp.
- Barrett RS. 1998. *Challenging the Myths of Fair Employment Practices.* Westport, CT: Quorum Books. 190 pp.
- Barrick MR, Mount MK. 1991. The Big Five personality dimensions and job performance: a meta-analysis. *Pers. Psychol.* 44:1-26
- Barrick MR, Mount MK. 1996. Effects of impression management and self-deception on the predictive validity of personality constructs. *J. Appl. Psychol.* 81:261-72
- Barrick MR, Stewart GL, Neubert MJ, Mount MK. 1998. Relating member ability and personality to work-team processes and team effectiveness. *J. Appl. Psychol.* 83:377-91
- Barry B, Stewart GL. 1997. Composition, process, and performance in self-managed groups: the role of personality. *J. Appl. Psychol.* 82:62-78
- Bauer TN, Maertz CP Jr, Dolen MR, Campion MA. 1998. Longitudinal assessment of applicant reactions to employment testing and test outcome feedback. *J. Appl. Psychol.* 83:892-903
- Bernstein J. 1999. *Computer-based assessment of spoken language.* Presented at Annu. Meet. Soc. Ind. Organ. Psychol., 14th, Atlanta
- Bobko P, Roth PL, Potosky D. 1999. Derivation and implications of a meta-analytic matrix incorporating cognitive ability, alternative predictors, and job performance. *Pers. Psychol.* 52:1-31
- Bobrow W, Leonards JS. 1997. Development and validation of an assessment center during organizational change. *J. Soc. Behav. Pers.* 12(5):217-36
- Borman WC, Motowidlo SJ, eds. 1997. Organizational citizenship behavior and contextual performance. *Hum. Perform.* 10:67-192
- Bruyère S. 1999. Disability nondiscrimination in the employment process: the role for psychologists. See Ekstrom & Smith 1999. In press
- Burke MJ, Rupinski MT, Dunlap WP, Davison HK. 1996. Do situational variables act as substantive causes of relationships between individual difference variables? Two large-scale tests of "common cause" models. *Pers. Psychol.* 49:573-98
- Burnett JR, Fan C, Motowidlo SJ, Degroot T. 1998. Interview notes and validity. *Pers. Psychol.* 51:375-96
- Burnett JR, Motowidlo SJ. 1998. Relations between different sources of information in the structured selection interview. *Pers. Psychol.* 51:963-83
- Burroughs SM, Bing MN, James LR. 1999. Reconsidering how to measure employee reliability: an empirical comparison and integration of self-report and conditional reasoning methodologies. See Williams & Burroughs 1999
- Bushman BJ, Wells GL. 1998. Trait aggressiveness and hockey penalties: predicting

- hot tempers on the ice. *J. Appl. Psychol.* 83:969-74
- Campbell JP. 1996. Group differences and personnel decisions: validity, fairness, and affirmative action. *J. Vocat. Behav.* 49:122-58
- Campbell JP. 1999. The definition and measurement of performance in the new age. See Ilgen & Pulakos 1999, pp. 399-429
- Campbell JP, Gasser MB, Oswald FL. 1996. The substantive nature of job performance variability. See Murphy 1996, pp. 258-99
- Campion MA, Palmer DK, Campion JE. 1997. A review of structure in the selection interview. *Pers. Psychol.* 50:655-702
- Carlson KD, Scullen SE, Schmidt FL, Rothstein H, Erwin F. 1999. Generalizable biographical data validity can be achieved without multi-organizational development and keying. *Pers. Psychol.* 52:731-55
- Carretta TR, Doub TW. 1998. Group differences in the role of *g* and prior job knowledge in the acquisition of subsequent job knowledge. *Pers. Individ. Differ.* 23:585-93
- Carretta TR, Ree MJ. 1996. Factor structure of the Air Force officer qualifying test: analysis and comparison. *Mil. Psychol.* 8:29-42
- Carretta TR, Ree MJ. 1997. Negligible sex differences in the relation of cognitive and psychomotor abilities. *Pers. Individ. Differ.* 22:165-72
- Carretta TR, Retzlaff PD, Callister JD, King RE. 1998. A comparison of two U.S. Air Force pilot aptitude tests. *Aviation, Space, Environ. Med.* 69:931-35
- Cascio WF. 1995. Whither industrial and organizational psychology in a changing world of work? *Am. Psychol.* 50:928-39
- Chan D. 1997. Racial subgroup differences in predictive validity perceptions on personality and cognitive ability tests. *J. Appl. Psychol.* 82:311-20
- Chan D. 1998a. The conceptualization and analysis of change over time: an integrative approach incorporating longitudinal mean and covariance structures analysis (LMACS) and multiple indicator latent growth modeling (MLGM). *Organ. Res. Methods* 1:421-83
- Chan D. 1998b. Functional relations among constructs in the same content domain at different levels of analysis: a typology of composition models. *J. Appl. Psychol.* 83:234-46
- 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. *J. Appl. Psychol.* 82:143-59
- Chan D, Schmitt N, Jennings D, Sheppard L. 1999. Developing measures of basic job-relevant English proficiency for the prediction of job performance and promotability. *J. Bus. Psychol.* In press
- Chan D, Schmitt N, Sacco JM, DeShon RP. 1998. Understanding pretest and posttest reactions to cognitive ability and personality tests. *J. Appl. Psychol.* 83:471-85
- Cliff N, Caruso JC. 1998. Reliable component analysis through maximizing composite reliability. *Psychol. Methods* 3:291-308
- Cohen SG, Bailey DE. 1997. What makes teams work: group effectiveness research from the shop floor to the executive suite. *J. Manage.* 23(3):239-90
- Coleman V, Borman W. 1999. Investigating the underlying structure of the citizenship performance domain. *Hum. Resour. Res. Rev.* In press
- Colihan J, Burger GK. 1995. Constructing job families: an analysis of quantitative techniques used for grouping jobs. *Pers. Psychol.* 48:563-86
- Collins JM, Gleaves DH. 1998. Race, job applicants, and the five-factor model of personality: implications for black psychology, industrial/organizational psychology, and the five-factor theory. *J. Appl. Psychol.* 83:531-44
- Collins JM, Schmidt FL, Sanchez-Ku M, Thomas LE, McDaniel M. 1999. Predicting assessment center ratings from cognitive ability and personality. See Quiñones 1999
- Conway JM, Jako RA, Goodman DF. 1995. A meta-analysis of interrater and internal con-

- sistency reliability of selection interviews. *J. Appl. Psychol.* 80:565-79
- Cronshaw SF. 1997. Lo! The stimulus speaks: the insider's view on Whyte and Latham's "The futility of utility analysis." *Pers. Psychol.* 50:611-15
- Cudeck R. 1996. Mixed-effects models in the study of individual differences with repeated measures data. *Multivariate Behav. Res.* 31:371-403
- Dalessio AT, Crosby MM, McManus MA. 1996. Stability of biodata keys and dimensions across English-speaking countries: a test of the cross-situational hypothesis. *J. Bus. Psychol.* 10:289-96
- Deadrick DL, Bennett N, Russell CJ. 1997. Using hierarchical linear modeling to examine dynamic performance criteria over time. *J. Manage.* 23:745-57
- Dean MA, Russell CJ, Muchinsky PM. 1999. Life experiences and performance prediction: toward a theory of biodata. In *Research in Personnel and Human Resources Management*, Vol. 17, ed. G Ferris. Greenwich, CT: JAI. In press
- DeShon RP, Alexander RA. 1996. Alternative procedures for testing regression slope homogeneity when group error variances are unequal. *Psychol. Methods* 1:261-77
- DeShon RP, Ployhart RE, Sacco JM. 1998a. The estimation of reliability in longitudinal models. *Int. J. Behav. Dev.* 22:493-515
- DeShon RP, Smith MR, Chan D, Schmitt N. 1998b. Can racial differences in cognitive test performance be reduced by presenting problems in a social context. *J. Appl. Psychol.* 83:438-51
- Dipboye R. 1997. Structured selection interviews: Why do they work? Why are they underutilized? See Anderson & Herriot 1997, pp. 455-73
- Douglas EF, McDaniel MA, Snell AF. 1996. The validity of non-cognitive measures decays when applicants fake. In *Proc. Acad. Manage.*, ed. JB Keyes, LN Dosier, pp. 127-31. Madison, WI: Omnipress. 594 pp.
- Dunnette M. 1976. Aptitudes, abilities, and skills. In *Handbook of Industrial and Organizational Psychology*, ed. M Dunnette, pp. 473-520. Chicago: Rand McNally. 1740 pp.
- Dunnette MD. 1997. Emerging trends and vexing issues in industrial and organizational psychology. *Appl. Psychol. Int. Rev.* 47(2):129-53
- Dwight SA, Donovan JJ. 1998. *Warning: Proceed with caution when warning applicants not to dissimulate (revised)*. Presented at Annu. Conf. Soc. Ind. Organ. Psychol., 13th, Dallas
- Ekstrom R, Smith D, eds. 1999. *Assessing Individuals with Disabilities: a Sourcebook*. Washington, DC: Am. Psychol. Assoc. In press
- Ellingson JE, Sackett PR, Hough LM. 1999a. Social desirability corrections in personality measurement: issues of applicant comparison and construct validity. *J. Appl. Psychol.* 84:155-66
- Ellingson JE, Smith DB, Sackett PR. 1999b. Investigating the influence of social desirability on personality factor structure. *J. Appl. Psychol.* In press
- Fox HR. 1999. Task force describes test user qualifications. *Score Newsl.* 21(3):3-4
- Frei RL, McDaniel MA. 1998. Validity of customer service measures in personnel selection: a review of criterion and construct evidence. *Hum. Perform.* 11:1-27
- Gandy JA, Dye DA, MacLane CN. 1994. See Stokes et al 1994, pp. 275-309
- Ghiselli EE. 1966. *The Validity of Occupational Aptitude Tests*. New York: Wiley & Sons. 155 pp.
- Gilliland SW, Steiner DD. 1999. Causes and consequences of applicant perceptions of unfairness. In *Justice in the Workplace*, ed. R Cropanzano, Vol. 2 Mahwah, NJ: Lawrence Erlbaum. In press
- Goffin RD, Rothstein MG, Johnston NG. 1996. Personality testing and the assessment center: incremental validity for managerial selection. *J. Appl. Psychol.* 81:746-56
- Goldstein H, Riley Y, Yusko KP. 1999. Exploration of Black-White subgroup differences on interpersonal constructs. *Subgroup dif-*

- ferences in employment testing.* Symp. Annu. Meet. Soc. Ind. Organ. Psychol., 14th, Atlanta
- Goldstein HW, Yusko KP, Braverman EP, Smith DB, Chung B. 1998. The role of cognitive ability in the subgroup differences and incremental validity of assessment center exercises. *Pers. Psychol.* 51:357-74
- Gottfredson LS. 1986. Occupational Aptitude Patterns map: development and implications for a theory of job aptitude requirements. *J. Vocat. Behav.* 29:254-91
- Gottfredson LS. 1997. Why g matters: the complexity of everyday life. *Intelligence* 24:79-132
- Gottfredson LS. 1999. Skills gaps, not tests, make racial proportionality impossible. *Psychol. Public Policy Law.* In press
- Gough HG. 1968. *The Chapin Social Insight Test Manual.* Palo Alto, CA: Consult. Psychol. Press. 14 pp.
- Gough HG. 1994. Theory, development, and interpretation of the CPI Socialization scale. *Psychol. Rep.* 75:651-700 (Suppl.)
- Graves LM, Karren RJ. 1996. The employee selection interview: a fresh look at an old problem. *Hum. Res. Manage.* 35:163-80
- Guion RM. 1998. *Assessment, Measurement, and Prediction for Pers. Decisions.* Mahwah, NJ: Lawrence Erlbaum Assoc. 690 pp.
- Guzzo AR, Salas E, eds. 1995. *Team Effectiveness and Decision Making in Organizations.* San Francisco: Jossey-Bass. 413 pp.
- Hakel M, ed. 1998. *Beyond Multiple Choice: Evaluating Alternatives and Traditional Testing for Selection.* Hillsdale, NJ: Erlbaum. 221 pp.
- Hambleton RK. 1999. Guidelines for adapting educational and psychological tests. *Bull. Int. Test Commis.*
- Hancock GR. 1997. Correlation/validity coefficients disattenuated for score reliability: a structural equation modeling approach. *Educ. Psychol. Meas.* 57:598-606
- Hattrup K, Jackson SE. 1996. Learning about individual differences by taking situations seriously. See Murphy 1996, pp. 507-41
- Hattrup K, O'Connell MS, Wingate PH. 1998. Prediction of multidimensional criteria: distinguishing task and contextual performance. *Hum. Perform.* 11:305-19
- Hattrup K, Rock J, Scalia C. 1997. The effects of varying conceptualizations of job performance on adverse impact, minority hiring, and predicted performance. *J. Appl. Psychol.* 82:656-64
- Hayden vs County of Nassau.* No. 98-6113, 1999. WL 373636 (2nd Cir June 9, 1999)
- Heilman ME. 1996. Affirmative action's contradictory consequences. *J. Soc. Issues* 52(4):105-9
- Heilman ME, Battle WS, Keller CE, Lee RA. 1998. Type of affirmative action policy: a determinant of reactions to sex-based preferential selection? *J. Appl. Psychol.* 83:190-205
- Helms JE. 1992. Why is there no study of cultural equivalence in standardized cognitive ability testing? *Am. Psychol.* 47:1083-101
- Herriot P, Anderson N. 1997. Selecting for change: How will personnel and selection psychology survive? See Anderson & Herriot 1997, pp. 1-34
- Highberger W. 1996. Current evidentiary issues in employment litigation. *Empl. Relat. Law J.* 22(1):31-56
- Hoffman CC, McPhail SM. 1998. Exploring options for supporting test use in situations precluding local validation. *Pers. Psychol.* 51:987-1003
- Hoffman CC, Thornton GC III. 1997. Examining selection utility where competing predictors differ in adverse impact. *Pers. Psychol.* 50:455-70
- Hofmann DA. 1997. An overview of the logic and rationale of hierarchical linear models. *J. Manage.* 23:723-44
- Hofstede G. 1980. *Culture's Consequences: International Differences in Work-Related Values.* Thousand Oaks, CA: Sage
- Hogan J, Brinkmeyer K. 1997. Bridging the gap between overt and personality-based integrity tests. *Pers. Psychol.* 50:587-99

- Hogan J, Ones DS. 1997. Conscientiousness and integrity at work. See Hogan et al 1997, pp. 513–41
- Hogan J, Rybicki SL. 1998. *Performance Improvement Characteristics Job Analysis*. Tulsa, OK: Hogan Assessment Systems
- Hogan R. 1968. Development of an empathy scale. *J. Consult. Clin. Psychol.* 33:307–16
- Hogan R. 1998. Reinventing personality. *J. Soc. Clin. Psychol.* 17:1–10
- Hogan R, Johnson J, Briggs S, eds. 1997. *Handbook of Personality Psychology*. San Diego: Academic. 987 pp.
- Hornick CW, Fox KA, Axton TR, Wyatt BS. 1999. The relative contribution of conditional reasoning and multiple intelligence measures in predicting firefighter and law enforcement officer job performance. See Williams & Burroughs 1999
- Hough LM. 1992. The 'Big Five' personality variables—construct confusion: description versus prediction. *Hum. Perform.* 5(1&2): 139–55
- Hough LM. 1997. The millennium for personality psychology: new horizons or good old daze. *Appl. Psychol. Int. Rev.* 47(2):233–61
- Hough LM. 1998a. Effects of intentional distortion in personality measurement and evaluation of suggested palliatives. *Hum. Perform.* 11:209–44
- Hough LM. 1998b. Personality at work: issues and evidence. See Hakel 1998, pp. 131–59
- Hough LM, Ones DS, Viswesvaran C. 1998. Personality correlates of managerial performance constructs. See Page 1998
- Hough LM, Paullin C. 1994. Construct-oriented scale construction: the rational approach. See Stokes et al 1994, pp. 109–45
- Hough LM, Schneider RJ. 1996. Personality traits, taxonomies, and applications in organizations. See Murphy 1996, pp. 31–88
- House RJ, Wright NS, Aditya RN. 1997. Cross-cultural research on organizational leadership: a critical analysis and a proposed theory. In *New Perspectives on International Industrial/Organizational Psychology*, ed. PC Earley, M Erez, pp. 535–625. San Francisco: New Lexington Press. 790 pp.
- Howard A, ed. 1995. *The Changing Nature of Work*. San Francisco: Jossey-Bass. 590 pp.
- Howard A. 1999. Discussant comments. See Quiñones 1999
- Huffcutt AI, Roth PL. 1998. Racial group differences in employment interview evaluations. *J. Appl. Psychol.* 83:179–89
- Huffcutt AI, Woehr DJ. 1999. Further analysis of employment interview validity: a quantitative evaluation of interviewer-related structuring methods. *J. Organ. Behav.* 20(4):549–60
- Hunter JE. 1997. Needed: a ban on the significance test. *Psychol. Sci.* 8:3–7
- Hunter JE, Schmidt FL. 1996. Cumulative research knowledge and social policy formulation: the critical role of meta-analysis. *Psychol. Public Policy Law* 2:324–47
- Ilgel DR, Pulakos ED. 1999. *The Changing Nature of Performance: Implications for Staffing, Motivation, and Development*. San Francisco: Jossey-Bass. 452 pp.
- James LR. 1998. Measurement of personality via conditional reasoning. *Organ. Res. Methods* 1(2):131–63
- James LR. 1999. *Use of a conditional reasoning measure for aggression to predict employee reliability*. Presented at Annu. Meet. Soc. Ind. Organ. Psychol., 14th, Atlanta
- James LR, Demaree RG, Mulaik SA, Ladd RT. 1992. Validity generalization in the context of situational models. *J. Appl. Psychol.* 77:3–14
- Jeanneret R. 1998. Ethical, legal, and professional issues for individual assessment. In *Individual Psychological Assessment: Predicting Behavior in Organizational Settings*, ed. R Jeanneret, R Silzer, pp. 88–131. San Francisco: Jossey-Bass. 495 pp.
- Johnson JW. 1999. A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behav. Res.* In press
- Johnson JW, Schneider RJ, Oswald FL. 1997. Toward a taxonomy of managerial performance profiles. *Hum. Perform.* 10:227–50

- Joint Committee on Testing Practices. 1999. *Test Taker Rights and Responsibilities: Guidelines and Expectations*. Washington, DC: Am. Psychol. Assoc. In press
- Judge TA, Erez A, Bono JE. 1998. The power of being positive: the relation between positive self-concept and job performance. *Hum. Perform.* 11:167-87
- Kataoka HC, Latham GP, Whyte G. 1997. The relative resistance of the situational, patterned behavior, and conventional structured interviews to anchoring effects. *Hum. Perform.* 10:47-63
- Katigbak MS, Church AT, Akamine TX. 1996. Cross-cultural generalizability of personality dimensions: relating indigenous and imported dimensions in two cultures. *J. Person. Soc. Psychol.* 70:99-114
- Kehoe JF, ed. 1999. *Managing Selection Strategies in Changing Organizations*. San Francisco: Jossey-Bass. In press
- Kiker S, Motowidlo S. 1999. Main and interaction effects of task and contextual performance on supervisory reward decisions. *J. Appl. Psychol.* 84:602-09
- Klimoski R, Zukin LB. 1999. Selection and staffing for team effectiveness. See Sundstrom 1999, pp. 63-91
- Kraut AI, Korman AK. 1999a. The "DELTA Forces" causing change in human resource management. See Kraut & Korman 1999b, pp. 3-22
- Kraut AI, Korman AK, eds. 1999b. *Evolving Practices in Human Resource Management*. San Francisco: Jossey-Bass. 376 pp.
- Kravitz DA, Harrison DA, Turner ME, Levine EL, Chaves W, et al. 1997. *Affirmative Action: a Review of Psychological and Behavioral Research*. Bowling Green, OH: Soc. Ind. Organ. Psychol. 50 pp.
- Kuncel NR, Hezlett SA, Ones DS. 1999. Comprehensive meta-analysis of the predictive validity of the graduate record examinations: implications for graduate student selection and performance. *Psychol. Bull.* In press
- Latham GP, Whyte G. 1994. The futility of utility analysis. *Pers. Psychol.* 47:31-46
- Lefkowitz J, Gebbia MI, Balsam T, Dunn L. 1999. Dimensions of biodata items and their relationships to item validity. *J. Occup. Organ. Psychol.* In press
- Legree PJ, Pifer ME, Grafton FC. 1996. Correlations among cognitive abilities are lower for higher ability groups. *Intelligence* 23:45-57
- LePine JA, Hollenbeck JR, Ilgen DR, Hedlund J. 1997. Effects of individual differences on the performance of hierarchical decision-making teams: much more than *g*. *J. Appl. Psychol.* 85:803-11
- Levine EL, Spector PE, Menon S, Narayanan L, Cannon-Bowers J. 1996. Validity generalization for cognitive psychomotor, and perceptual tests for craft jobs in the utility industry. *Hum. Perform.* 9:1-22
- Li H, Rosenthal R, Rubin DB. 1996. Reliability of measurement in psychology: from Spearman-Brown to maximal reliability. *Psychol. Methods* 1:98-107
- Lievens F. 1998. Factors which improve the construct validity of assessment centers: a review. *Int. J. Select. Assess.* 6:141-52
- Lowman RL, ed. 1998. *The Ethical Practice of Psychology in Organizations*. Washington, DC: Am. Psychol. Assoc. 299 pp.
- Lubinski D. 2000. Scientific and social significance of assessing individual differences: "Sinking shafts at a few critical points." *Annu. Rev. Psychol.* 51:405-44
- Lubinski D, Benbow C. 1999. States of excellence: a psychological interpretation of their emergence. *Am. Psychol.* In press
- Mael FA. 1998. Privacy and personnel selection: reciprocal rights and responsibilities. *Empl. Responsib. Rights J.* 11:187-214
- Mael FA, Connelly M, Morath RA. 1996a. None of your business: parameters of biodata invasiveness. *Pers. Psychol.* 49:613-50
- Mael F, Kilcullen R, White L. 1996b. Soldier attributes for peacekeeping and peacemaking. In *Reserve Component Soldiers as Peacekeepers*, ed. R Phelps, B Farr, pp. 29-

57. Alexandria, VA: U.S. Army Res. Inst. Behav. Sci. 449 pp.
- Matthews G. 1997. The Big Five as a framework for personality assessment. See Anderson & Herriot 1997, pp. 475–92
- McCrae RR, Costa PT Jr. 1997. Personality trait structure as a human universal. *Am. Psychol.* 52:509–16
- McManus MA, Kelly ML. 1999. Personality measures and biodata: evidence regarding their incremental predictive value in the life insurance industry. *Pers. Psychol.* 52:137–48
- Migetetz DZ, James LR, Ladd RT. 1999a. *A validation of the conditional reasoning measurement*. Presented at Annu. Meet. Soc. Ind. Organ. Psychol., 14th, Atlanta
- Migetetz DZ, McIntyre M, James LR. 1999b. Measuring reliability among contingent workers. See Williams & Burroughs 1999
- Millsap RE. 1995. Measurement invariance, predictive invariance, and the duality paradox. *Multivariate Behav. Res.* 30:577–605
- Millsap RE. 1997. Invariance in measurement and prediction: their relationship in the single-factor case. *Psychol. Methods* 2:248–60
- Miner JB, Capps MH. 1996. *How Honesty Testing Works*. Westport, CT: Quorum Books. 192 pp.
- Mone MA, Mueller GC, Mauland W. 1996. The perceptions and usage of statistical power in applied psychology and management research. *Pers. Psychol.* 49:103–20
- Morgeson FP, Campion MA. 1997. Social and cognitive sources of potential inaccuracy in job analysis. *J. Appl. Psychol.* 82:627–55
- Moss FA, Hunt T, Omwake KT, Woodward LG. 1955. *Social Intelligence Test Manual*. Washington, DC: Cent. Psychol. Serv. 4 pp.
- Motowidlo SJ, Burnett JR. 1995. Aural and visual sources of validity in structured employment interviews. *Organ. Behav. Hum. Decis. Process.* 61(3):239–49
- Mount MK, Barrick MR. 1995. The Big Five personality dimensions: implications for research and practice in human resource management. *Res. Pers. Hum. Resour. Manage.* 13:153–200
- Mount MK, Barrick MR, Stewart GL. 1998. Five-factor model of personality and performance in jobs involving interpersonal interactions. *Hum. Perform.* 11:145–65
- Muchinsky PM. 1999. Biodata: a mirror of moxie. See Stennett et al 1999. In press
- Mumford MD, Costanza DP, Connelly MS, Johnson JF. 1996. Item generation procedures and background data scales: implications for construct and criterion-related validity. *Pers. Psychol.* 49:361–98
- Mumford MD, Stokes GS. 1992. Developmental determinants of individual action: theory and practice in the application of background data measures. In *Handbook of Industrial and Organizational Psychology*, ed. MD Dunnette, LM Hough, 2:61–138. Palo Alto, CA: Consult. Psychol. Press. 957 pp. 2nd ed.
- Murphy vs United Parcel Service, Inc.* No. 97–1992 (119 S Ct 2133, June 22, 1999)
- Murphy KR, ed. 1996. *Individual Differences and Behavior in Organizations*. San Francisco: Jossey-Bass. 606 pp.
- Murphy KR. 1997. Meta-analysis and validity generalization. See Anderson & Herriot 1997, pp. 323–42
- Murphy KR, Luther N. 1997. Assessing honesty, integrity, and deception. See Anderson & Herriot 1997, pp. 369–88
- Murphy KR, Myers B. 1999. Testing the hypothesis that treatments have negligible effects: minimum-effect tests in the general linear model. *J. Appl. Psychol.* 84:234–47
- Murphy KR, Shiarella AH. 1997. Implications of the multidimensional nature of job performance for the validity of selection tests: multivariate frameworks for studying test validity. *Pers. Psychol.* 50:823–54
- Nelson JB. 1997. The boundaryless organization: implications for job analysis, recruitment, and selection. *Hum. Resour. Plan.* 20(4):39–49

- Nyfield G, Baron H. 1999. Cultural context in adapting selection practices across borders. See Kehoe 1999. In press
- O'Neil HF Jr, ed. 1997. *Workforce Readiness: Competencies and Assessment*. Mahwah, NJ: Erlbaum. 467 pp.
- O'Neil HF Jr, Allred K, Dennis RA. 1997a. Use of computer simulation for assessing the interpersonal skill of negotiation. See O'Neil 1997, pp. 205–28
- O'Neil HF Jr, Allred K, Dennis RA. 1997b. Validation of a computer simulation for assessment of interpersonal skills. See O'Neil 1997, pp. 229–54
- Ones DS, Hough LM, Viswesvaran C. 1998. Personality correlates of managerial performance constructs. See Page 1998
- Ones DS, Viswesvaran C. 1996. *What do pre-employment customer service scales measure? Explorations in construct validity and implications for personnel selection*. Presented at Annu. Meet. Soc. Ind. Organ. Psychol., 11th, San Diego
- Ones DS, Viswesvaran C. 1998a. Gender, age, and race differences on overt integrity tests: results across four large-scale job applicant data sets. *J. Appl. Psychol.* 83:35–42
- Ones DS, Viswesvaran C. 1998b. Integrity testing in organizations. In *Dysfunctional Behavior in Organizations: Vol. 2, Nonviolent Behaviors in Organizations*, ed. RW Griffin, A O'Leary-Kelly, JM Collins, pp. 243–76. Greenwich, CT: JAI. 318 pp.
- Ones DS, Viswesvaran C. 1998c. The effects of social desirability and faking on personality and integrity assessment for personnel selection. *Hum. Perform.* 11:245–69
- Ones DS, Viswesvaran C. 1999. Job-specific applicant pools and national norms for personality scales: implications for range restriction corrections in validation research. *J. Appl. Psychol.* In press
- Ones DS, Viswesvaran C, Reiss AD. 1996. Role of social desirability in personality testing for personnel selection: the red herring. *J. Appl. Psychol.* 81:660–79
- Ones DS, Viswesvaran C, Schmidt FL, Reiss AD. 1994. *The validity of honesty and violence scales of integrity tests in predicting violence at work*. Presented at Annu. Meet. Acad. Manage., Dallas
- Ostroff C, Harrison D. 1999. Meta-analysis, level of analysis, and best estimates of population correlations: cautions for interpreting meta-analytic results in organizational behavior. *J. Appl. Psychol.* 84:260–70
- Oswald FL, Ferstl KL. 1999. Linking a structure of vocational interests to Gottfredson's (1986) Occupational Aptitude Patterns map. *J. Vocat. Behav.* 54:214–31
- Oswald FL, Johnson JW. 1998. On the robustness, bias, and stability of statistics from meta-analysis of correlation coefficients: some initial Monte Carlo findings. *J. Appl. Psychol.* 83:164–78
- Oswald FL, Saad SA, Sackett PR. 1999. The homogeneity assumption in differential prediction analysis: Does it really matter? *J. Appl. Psychol.* In press
- Page RC, chair. 1998. In *Personality determinants of managerial potential performance, progression and ascendancy*. Symp. Annu. Meet. Soc. Ind. Organ. Psychol., 13th, Dallas
- Patton TW, Walton W, James LR. 1999. Measuring personal reliability via conditional reasoning: identifying people who will work reliably. See Williams & Burroughs 1999
- Pearlman K. 1997. Twenty-first century measures for twenty-first century work. In *Transitions in Work and Learning: Implications for Assessment*, ed. A Lesgold, MJ Feuer, A Block, pp. 136–79. Washington, DC: Natl. Acad. Press. 283 pp.
- Pearlman K, Barney MF. 1999. Selection for a changing workplace. See Kehoe 1999. In press
- Peterson NG, Mumford MD, Borman WC, Jeanneret PR, Fleishman EA. 1999. *An Occupational Information System for the 21st Century: the Development of O*NET*. Washington, DC: Am. Psychol. Assoc. 319 pp.

- Ployhart RE, Hakel MD. 1998. The substantive nature of performance variability: predicting interindividual differences in intraindividual performance. *Pers. Psychol.* 51:859-901
- Ployhart RE, Ryan AM, Bennett M. 1999. Explanations for selection decisions: applicants' reactions to informational and sensitivity features of explanations. *J. Appl. Psychol.* In press
- Plutchik R, Conte HR, eds. 1997. *Circumplex Models of Personality and Emotions*. Washington, DC: Am. Psychol. Assoc. 484 pp.
- Pulakos ED, Schmitt N. 1996. An evaluation of two strategies for reducing adverse impact and their effects on criterion-related validity. *Hum. Perform.* 9:241-58
- Quiñones MA, chair. 1999. *Assessment centers, 21st century: new issues, and new answers to old problems*. Symp. Annu. Meet. Soc. Ind. Organ. Psychol., 14th, Atlanta
- Quiñones MA, Ford JK, Teachout MS. 1995. The relationship between work experience and job performance: a conceptual and meta-analytic review. *Pers. Psychol.* 48:887-910
- Raju NS, Bilgic R, Edwards JE, Fleer PF. 1997. Methodology review: estimation of population validity and cross-validity, and the use of equal weights in prediction. *Appl. Psychol. Meas.* 21:291-305
- Raju NS, Bilgic R, Edwards JE, Fleer PF. 1999. Accuracy of population validity and cross-validity estimation: an empirical comparison of formula-based, traditional empirical, and equal weights procedures. *Appl. Psychol. Meas.* 23:99-115
- Raykov T. 1997. Estimation of composite reliability for congeneric measures. *Appl. Psychol. Meas.* 21:173-84
- Raymark PH, Schmit MJ, Guion RM. 1997. Identifying potentially useful personality constructs for employee selection. *Pers. Psychol.* 50:723-36
- Ree MJ, Carretta TR, Teachout MS. 1995. Role of ability and prior job knowledge in complex training performance. *J. Appl. Psychol.* 80:721-30
- Robertson I, Callinan M. 1998. Personality and work behaviour. *Eur. J. Work Organ. Psychol.* 7:321-40
- Rosse JG, Stecher MD, Levin RA, Miller JL. 1998. The impact of response distortion on preemployment personality testing and hiring decisions. *J. Appl. Psychol.* 83:634-44
- Roth PL, Bobko P. 1997. A research agenda for multi-attribute utility analysis in human resource management. *Hum. Resource Manage. Rev.* 7:341-68
- Roth PL, Campion JE, Jones SD. 1996. The impact of four missing data techniques on validity estimates in human resource management. *J. Bus. Psychol.* 11:101-12
- Russell CJ. 1999. Toward a model of life experience learning. See Stennett et al. 1999. In press
- Russell CJ, Dean MA, Broach D. 1998. *Guidelines for Bootstrapping Validity Coefficients in ATCS Selection Research*. Norman, OK: Univ. Okla. 58 pp.
- Russell JA, Carroll JM. 1999. On the bipolarity of positive and negative affect. *Psychol. Bull.* 125:3-30
- Ryan AM, Greguras GJ. 1998. Life is not multiple choice: reactions to the alternatives. See Hakel 1998, pp. 183-202
- Ryan AM, Ployhart RE, Friedel LA. 1998a. Using personality testing to reduce adverse impact: a cautionary note. *J. Appl. Psychol.* 83:298-307
- Ryan AM, Ployhart RE, Greguras GJ, Schmit MJ. 1998b. Test preparation programs in selection contexts: self-selection and program effectiveness. *Pers. Psychol.* 51:599-621
- Sackett PR, Ellingson JE. 1997. The effects of forming multi-predictor composites on group differences and adverse impact. *Pers. Psychol.* 50:707-21
- Sackett PR, Gruys ML, Ellingson JE. 1998. Ability-personality interactions when predicting job-performance. *J. Appl. Psychol.* 83:545-56

- Sackett PR, Roth L. 1996. Multi-stage selection strategies: a Monte Carlo investigation of effects on performance and minority hiring. *Pers. Psychol.* 49:549–72
- Sackett PR, Wanek JE. 1996. New developments in the use of measures of honesty, integrity, conscientiousness, dependability, trustworthiness, and reliability for personnel selection. *Pers. Psychol.* 49:787–829
- Sager CE, Peterson NG, Oppler SH, Rosse RL, Walker CB. 1997. An examination of five indexes of test battery performance: analysis of the ECAT battery. *Mil. Psychol.* 9:97–120
- Salgado JF. 1997a. The five factor model of personality and job performance in the European community. *J. Appl. Psychol.* 82:30–43
- Salgado JF. 1997b. VALCOR: a program for estimating standard error, confidence intervals, and probability of corrected validity. *Behav. Res. Methods Instrum. Comput.* 29:464–67
- Salgado JF. 1998. Big Five personality dimensions and job performance in army and civil occupations: a European perspective. *Hum. Perform.* 11:271–88
- Sanchez J, Levine E. 1999. Is job analysis dead, misunderstood, or both? New forms of work analysis and design. See Kraut & Korman 1999b, pp. 43–68
- Sanchez J, Prager I, Wilson A, Viswesvaran C. 1998. Understanding within-job title variance in job-analytic ratings. *J. Bus. Psychol.* 12:407–19
- Sánchez-Meca J, Marín-Martínez F. 1997. Homogeneity tests in meta-analysis: a Monte Carlo comparison of statistical power and Type I error. *Qual. Quant.* 31:385–99
- Sands WA, Waters BK, McBride JR. 1997. *Computerized Adaptive Testing*. Washington, DC: Am. Psychol. Assoc. 292 pp.
- Saucier G, Goldberg LR. 1998. What is beyond the Big Five? *J. Pers.* 66:495–524
- Schmidt FL, Hunter JE. 1996. Measurement error in psychological research: lessons from 26 research scenarios. *Psychol. Methods* 1:199–223
- Schmidt FL, Hunter JE. 1998. The validity and utility of selection methods in personnel psychology: practical and theoretical implications of 85 years of research findings. *Psychol. Bull.* 124:262–74
- Schmidt FL, Rader M. 1999. Exploring the boundary conditions for interview validity: meta-analytic validity findings for a new interview type. *Pers. Psychol.* 52:445–64
- Schmidt FL, Viswesvaran V, Ones DS. 1997. Validity of integrity tests for predicting drug and alcohol abuse: a meta-analysis. In *Meta-Analysis of Drug Abuse Prevention Programs*, ed. WJ Bukoski, pp. 69–95. Rockville, MD: NIDA National Institute on Drug Abuse Press. 263 pp.
- Schmit MJ, Kilm JA, Robie C. 1999. Refining a personality test to be used in selection across several cultures. In *Personality and Performance: Boundary Conditions for Measurement and Structural Models*, chair WC Borman, Symp. Annu. Meet. Soc. Ind. Organ. Psychol., 14th Atlanta
- Schmit MJ, Ryan AM. 1997. Applicant withdrawal: the role of test-taking attitudes and racial differences. *Pers. Psychol.* 50:855–76
- Schmitt N, Chan D. 1998. *Personnel Selection*. Thousand Oaks, CA: Sage. 378 pp.
- Schmitt N, Chan D. 1999. The status of research on applicant reactions to selection tests and its implications for managers. *Int. J. Manage. Rev.* In press
- Schmitt N, Jennings D, Toney R. 1999. Can we develop measures of hypothetical constructs? *Hum. Resour. Manage. Rev.* In press
- Schmitt N, Nason E, Whitney DJ, Pulakos ED. 1995. The impact of method effects on structural parameters in validation research. *J. Manage.* 21:159–74
- Schmitt N, Ployhart RE. 1999. Estimates of cross-validity for stepwise regression and with predictor selection. *J. Appl. Psychol.* 84:50–57

- Schmitt N, Pulakos ED. 1998. Biodata and differential prediction: some reservations. See Hakel 1998, pp. 167–82
- Schmitt N, Rogers W, Chan D, Sheppard L, Jennings D. 1997. Adverse impact and predictive efficiency of various predictor combinations. *J. Appl. Psychol.* 82:719–30
- Schneider RJ, Ackerman PL, Kanfer R. 1996. To “act wisely in human relations”: exploring the dimensions of social competence. *Pers. Individ. Differ.* 21:469–81
- Schoenfeldt LF. 1999. From dust bowl empiricism to rational constructs in biographical data. *Hum. Resour. Manage. Rev.* In press
- Segall DO. 1999. General ability measurement: an application of multidimensional adaptive testing. Presented at Meet. Natl. Council Meas. Educ., Montreal, Canada
- Shackleton V, Newell S. 1997. International assessment and selection. See Anderson & Herriot 1997, pp. 81–95
- Sharf JC, Jones DP. 1999. Employment risk management. See Kehoe 1999. In press
- Smith M, DeMatteo JS, Green P, James LR. 1995. *A comparison of new and traditional measures of achievement motivation.* Presented at Annu. Meet. Am. Psychol. Assoc., 103rd, New Orleans
- Snell AF, McDaniel MA. 1998. *Faking: getting data to answer the right questions.* Poster presented at Annu. Meet. Soc. Ind. Organ. Psychol., 13th, Dallas
- Snell AF, Sydell EJ, Lueke SB. 1999. Towards a theory of applicant faking: integrating studies of deception. *Hum. Resour. Manage. Rev.* In press
- Spychalski A, Quiñones M, Gaugler BB, Pohley K. 1997. A survey of assessment center practices in organizations in the United States. *Pers. Psychol.* 50:71–90
- Stajkovic AD, Luthans F. 1998. Self-efficacy and work-related performance: a meta-analysis. *Psychol. Bull.* 124:240–61
- Stanush P, Arthur W, Doverspike D. 1998. Hispanic and African American reactions to a simulated race-based affirmative action scenario. *Hispanic J. Behav. Sci.* 20(1):3–16
- Stennett RB, Parisi AG, Stokes GS, eds. 1999. *A Compendium: Papers Presented at the First Biennial Biodata Conference.* Athens, GA: Univ. Georgia. In press
- Stokes GS, Mumford MD, Owens WA, eds. 1994. *Biodata Handbook.* Palo Alto, CA: Consult. Psychol. Press. 650 pp.
- Stokes GS, Searcy CA. 1999. Specification of scales in biodata form development: rational vs. empirical and global vs. specific. *Int. J. Select. Assess.* 7:72–85
- Stricker LJ, Rock DA. 1998. Assessing leadership potential with a biographical measure of personality traits. *Int. J. Select. Assess.* 6:164–84
- Sundstrom E, ed. 1999. *Supporting Work Team Effectiveness.* San Francisco: Jossey-Bass. 388 pp.
- Sutton vs United Airlines, Inc.* No. 97–1943, 119 S. Ct. 2139, (June 22, 1999)
- Task Force on Assessment Center Guidelines. 1989. Guidelines for ethical considerations. *Public Pers. Manage.* 18:457–70
- Terpstra DE, Mohamed AA, Kethley B. 1999. An analysis of federal court cases involving nine selection devices. *Int. J. Select. Assess.* 7:26–34
- Terris W. 1997. The traditional regression model for measuring test bias is incorrect and biased against minorities. *J. Bus. Psychol.* 12:25–37
- Thorsteinson TJ, Ryan AM. 1997. The effect of selection ratio on perceptions of the fairness of a selection test battery. *Int. J. Select. Assess.* 5:159–68
- Tippins NT. 1999. The Americans with Disabilities Act and employment testing. See Ekstrom & Smith 1999. In press
- Tisak J, Tisak MS. 1996. Longitudinal models of reliability and validity: a latent curve approach. *Appl. Psychol. Meas.* 20(3):275–88
- Tokar DM, Fischer AR, Subich LM. 1998. Personality and vocational behavior: a selective review of the literature, 1993–1997. *J. Voc. Behav.* 53:115–53
- Truxillo DM, Bauer TN. 1999. Applicant reactions to test score banding in entry-level

- and promotional contexts. *J. Appl. Psychol.* 84:322-39
- U.S. Equal Employment Opportunity Commission. 1999. *Enforcement Guidance: Reasonable Accommodation and Undue Hardship under the Americans with Disabilities Act*. Washington, DC: Equal Empl. Opportun. Comm.
- U.S. Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, and Department of Justice. 1978. Uniform guidelines on employee selection procedures. *Fed. Regist.* 43:38290-315
- Van Scotter JR, Motowidlo SJ. 1996. Interpersonal facilitation and job dedication as separate facets of contextual performance. *J. Appl. Psychol.* 81:525-31
- Verive JM, McDaniel MA. 1996. Short-term memory tests in personnel selection: low adverse impact and high validity. *Intelligence* 23:15-32
- Vinchor AJ, Schippmann JS, Switzer FS, Roth PL. 1998. A meta-analytic review of predictors of job performance for salespeople. *J. Appl. Psychol.* 83:586-97
- Viswesvaran C, Ones DS. 1999. Meta-analyses of fakability estimates: implications for personality measurement. *Educ. Psychol. Meas.* 59:197-210
- Viswesvaran C, Ones DS, Schmidt FL. 1996. Comparative analysis of the reliability of job performance ratings. *J. Appl. Psychol.* 81:557-74
- Wang T. 1998. Weights that maximize reliability under a congeneric model. *Appl. Psychol. Meas.* 22:179-87
- West M, Allen N. 1997. Selecting for teamwork. See Anderson & Herriot 1997, pp. 493-506
- Whitney DJ, Diaz J, Mineghino ME, Powers K. 1999. Perceptions of overt and personality-based integrity tests. *Int. J. Select. Assess.* 7:35-45
- Whitney DJ, Schmitt N. 1997. Relationship between culture and responses to biodata employment items. *J. Appl. Psychol.* 82:113-29
- Whyte G, Latham G. 1997. The futility of utility analysis revisited: when even an expert fails. *Pers. Psychol.* 50:601-10
- Wiggins JS, Trapnell PD. 1997. Personality structure: the return of the Big Five. See Hogan et al 1997, pp. 737-65
- Wilk SL, Sackett PR. 1996. Longitudinal analysis of ability-job complexity fit and job change. *Pers. Psychol.* 49:937-67
- Williams LJ, Burroughs SM, chairs. 1999. *New developments using conditional reasoning to measure employee reliability*. Symp. Annu. Meet. Soc. Ind. Organ. Psychol., 14th, Atlanta
- Williamson LG, Campion JE, Malos SB, Roehling MV, Campion MA. 1997. Employment interview on trial: linking interview structure with litigation outcomes. *J. Appl. Psychol.* 82:900-12
- Woehr DJ, Arthur W Jr. 1999. The assessment center validity paradox: a review of the role of methodological factors. See Quiñones 1999
- Wright TA, Cropanzano R. 1998. Emotional exhaustion as a predictor of job performance and voluntary turnover. *J. Appl. Psychol.* 83:486-93
- Wright TA, Staw BM. 1999. Affect and favorable work outcomes: two longitudinal tests of the happy-productive worker thesis. *J. Organ. Behav.* 20:1-23
- Zickar MJ, Drasgow F. 1996. Detecting faking on a personality instrument using appropriateness measurement. *Appl. Psychol. Meas.* 20:71-87

Copyright of Annual Review of Psychology is the property of Annual Reviews Inc. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.