

# NEW FRONTIERS IN MANAGEMENT SELECTION SYSTEMS: WHERE MEASUREMENT TECHNOLOGIES AND THEORY COLLIDE

Craig J. Russell\*  
*Purdue University*

Karl W. Kuhnert  
*University of Georgia*

The argument is made that meta-analytic results have liberated the field from an indentured relationship with criterion-related validity. Three major approaches to managerial selection are reviewed with special attention paid to construct validity and theory development efforts. Emphasis is given to meaningful historical trends in the evolution of each approach. Where available, meta-analytic results are reported. Specific directions for future research aimed at increasing our understanding of why current selection systems predict managerial job performance are discussed.

Historically, one of the major demands placed on industrial psychologists has been to provide assistance in the selection of employees. During the long growth period of the U.S. economy between the end of World War II and the 1973 Arab oil embargo, identification and selection of personnel for managerial positions was considered one of the major concerns facing industry (Dunnette, 1971). This economic growth period influenced the field by "pulling" it to develop new and different techniques. At the same time, "push" factors in the form of Title VII of the 1964 Civil Rights Act and the 1968 Age Discrimination Act held existing selection technologies to an entirely new standard of social impact.

Not surprisingly, this was also a period when creative efforts resulted in the development and/or widespread implementation of many new measurement

---

\*Direct all correspondence to: Craig J. Russell, Department of Management, College of Business, Louisiana State University, Baton Rouge, LA 70803-6312.

---

Leadership Quarterly, 3(2), 109-135.

Copyright © 1992 by JAI Press Inc.

All rights of reproduction in any form reserved.

ISSN: 1048-9843

technologies for purposes of managerial selection. Douglas Bray developed and implemented the first private sector assessment center for managerial selection at Michigan Bell (Bray, 1964). Standard Oil of New Jersey validated and used biographical information questionnaires for managerial selection (Standard Oil Company of New Jersey, 1962). Norman Abrahams of the Navy Personnel Research and Development Center applied the Strong Vocational Interest Inventory to predict performance and attrition of future naval leaders at the U.S. Naval Academy (Abrahams & Neumann, 1973).

Simultaneously investigators were concerned with circumstances that might impact the validity of drawing criterion-related inferences from a test. Ghiselli (1956) expressed concern that we needed to pay close attention to the impact of "predictors of predictability" or what Saunders (1956) came to call moderator effects. Dunnette (1963, 1966) formally developed a selection research model that systematically explored alternate moderator effects. The moderator variable with perhaps the greatest impact on selection research efforts during this period was job content. Ghiselli's (1966) exhaustive review of aptitude test criterion-related validities is perhaps the most frequently cited study on the moderating effect of job content. His results suggest that the same test will demonstrate very different criterion-related validities in jobs that otherwise appear to be identical. Further, tests that by all construct validity evidence appear to be measuring the same thing yield different criterion-related validities when applied in the same job. During this period, if senior management wanted to know whether a selection system was discrimination between future high and low performers in an applicant pool, a criterion-related validity effort had to be conducted for each and every job-test combination.

Interestingly, Victor Vroom's preface to Dunnette's (1966) text *Personnel Selection and Placement* noted a change in focus of selection research, specifically that it was "less concerned with techniques for solving particular problems and more concerned with shedding light on the processes that underlie various behavioral phenomena, on the assumption that improvements in technology will be facilitated by a better understanding of these processes" (p. v-vi). Given the "pull" factors personified by business need, the "push" factors captured by emerging EEO doctrine, and the canon of situation specificity, Vroom's statement appears to have been about 20 years premature. Industrial psychologists of this era were too busy developing selection "products," validating them in every employment situation that might come along, and defending them from and/or refabricating them in the face of evolving EEO doctrine (e.g., the 4/5's rule; the Cleary, 1968, model of test bias, etc.).

By the 1980s, at least two of these influences had dwindled. With *Griggs v. Duke Power Company* (1971) and the *EEOC Uniform Guidelines on Employment Selection Procedures* (1978) the impact of EEO tenets on selection procedures were fairly well understood, becoming a somewhat routine component of selection system design and evaluation. Additionally, Schmidt and Hunter (1977) introduced applications of Baye's theorem to permit corrections for differences in sample size, range restriction, and measurement error across criterion-related validity studies. These meta-analytic procedures were quickly applied to the pool of existing criterion-related validity studies (e.g., Hunter & Hunter, 1984; Reilly & Chao, 1982; Schmitt, Gooding, Noe, & Kirsch, 1984). Results suggest that assessment centers biographical information, and cognitive

skill tests (among others) demonstrate substantial criterion-related validity that does not vary across managerial positions.

Thus, it would seem that the time is ripe to return to Vroom's observation of over 20 years ago. Freed from the bonds of situation specificity, the field needs to refocus its energies on basic construct validity and theory building efforts (Burke & Pearlman, 1988; Fleishman, 1988). The purpose of this review is to evaluate what is known about why three common managerial selection applications work. We start with a brief review of validity generalization evidence, followed by an exploration of recent construct validity results and theory building efforts reported for systems based on assessment centers, biographical information, and cognitive skills tests. Specific directions for future construct validity/theory building efforts are explored.

### **VALIDITY GENERALIZATION: A GROUNDED LAUNCH TOWARD THEORY CONSTRUCTION**

In the best of all possible worlds, we could organize the rest of this review around theories of managerial skills and abilities whose operationalizations yield meaningful criterion-related validities with measures of managerial performance. Campbell (1990) noted that taxonomies of individual difference variables in the cognitive, perceptual, spatial, psychomotor, physical, and sensory domains yield approximately fifty distinct constructs. Unfortunately, Campbell notes that none of these investigators had the "applied prediction problem in mind" (Campbell, 1990, p. 692) when constructing these taxonomies. Another way of saying this is that a good taxonomy is not the same as a theory of prediction or job performance. Such theories would make explicit predictions about how earlier constructs (predictors) relate to later constructs (job performance). Taxonomies of individual differences, especially those developed without the "applied prediction problem in mind," simply do not do this (Bobko & Russell, 1991). Unfortunately, no theories exist to guide our review (Burke & Pearlman, 1988; Campbell, 1990; Peterson & Bownas, 1982).

However, a number of important research efforts have been targeted at understanding why various measurement technologies yield criterion-related validity. Hence, we feel the state of theory development (or lack thereof) dictates a focus on the measurement technologies (see Campbell, 1990, for an approach starting with the criterion construct domain). We review efforts at establishing construct validity for each predictor measure and how investigators have attempted to embed the measure in theories or models of the individual and/or job (see Vance, MacCallum, Coovert, & Hedge, 1988, for an example using criterion performance measures). Glaser and Strauss (1967) labeled this "grounded theory building." In the absence of any theory with demonstrated criterion-related validities in applied settings, examination of established measurement technologies' scattered construct validity efforts becomes a promising point of migration toward theory development.

Results from four meta-analyses and a large sample (consortium) validity study were used to select measurement technologies with the highest criterion-related validities in management selection applications (Gaugler, Rosenthal, Thornton, & Bentson, 1987; Hunter & Hunter, 1982; Reilly & Chao, 1982; Richardson, Bellows, Henry, & Co., 1984; and Schmitt, Gooding, Noe, & Kirsch, 1984). All results reported in these studies

**Table 1**  
**Meta-Analytic Results**

	<i>rho</i>	<i>SD</i>	<i>k</i>	<i>N</i>
<i>Assessment Centers</i> <sup>1</sup>				
Gaugler et al. (1987)				
Performance Ratings	.36		44	4,180
Potential Ratings	.53		13	1,338
Dimension Performance	.33		9	748
Training Performance	.35		8	1,062
Career Progress	.36		33	4,907
Schmitt et al. (1984)				
Performance Ratings	.43	.05	6	394
Status Change	.41	.04	8	14,361
Wages	.24	.23	4	301
Achievement/Grades	.31	.26	3	289
<i>Biodata Questionnaires</i> <sup>2</sup>				
Brown (1981)	.26	.04	12	12,453
Hunter and Hunter (1984)	.37	.10	12	4,429
Schmitt et al. (1984)	.18	.01	3	136
Reilly and Chao (1982)	.40	nr	4	2,504
Richardson, Bellows, & Henry (1981, 1984, 1988)	.34	.01	5	11,738
<i>Cognitive Skills Tests</i> <sup>3</sup>				
Hunter and Hunter (1984)				
General Cognitive Ability	.53			
General Perceptual Ability	.43			
General Psychomotor Ability	.26			
Schmitt et al. (1984)	.17	.00	7	14,215

*Notes:* <sup>1</sup> Note that assessment center results from Hunter and Hunter (1984) were not included as these were simply a second reporting of Cohen, Moses, and Byham's (1974) results of averaging validities across studies with no correction for range restriction, criterion unreliability, or sample size.

<sup>2</sup> Only analyses reported by Richardson, Bellows, and Henry (1981, 1984, 1988) and Schmitt et al. (1984) are limited to studies conducted on managers. Neal Schmitt was kind enough to make available their original data, which we reanalyzed to select only those studies focusing on managerial positions using biodata instruments ( $N = 2$  studies containing 3 criterion validities). One of the two studies did not use biodata instruments in the traditional sense (military rank and pay grade as a "biodata" predictor).

<sup>3</sup> Results reported for Hunter and Hunter (1984) were based on a reanalysis of Ghiselli's (1973) original work on mean criterion-related validities with supervisory ratings of proficiency for managers (though one must read Hunter, 1986, to know this is the criterion employed). Results reported for Schmitt et al. (1984) were based on reanalysis of data made available by Neal Schmitt, selecting only those studies focusing on managerial positions and using cognitive skills tests.

were reviewed to determine which measurement technologies had the highest average validity with managerial performance criteria and, in our judgment, were widely used in organizations. Alphabetically, these selection procedures included assessment centers, biodata questionnaires, and cognitive skill tests. Results of these meta-analyses are reported in Table 1.

In addition to being distinguished by large bodies of criterion-related validity studies, assessment center, biodata, and cognitive skills selection literatures have experienced

increased construct validity efforts in the last decade. Assessment center research has focused on the failure of multi-trait multi-method correlation matrices to support a prior skill and ability dimensions. The biodata literature has been characterized by Owens' (1968, 1971) developmental/integrative model of performance and Mumford and Stokes (in press) ecology model, however, recent work has focused on biodata item content and item generation procedures to shed light on underlying constructs (e.g., Barge, 1987, 1988). Finally, Hulin, R. Henry, and Noon (1990) questioned the stability of relationships between cognitive skill tests and job performance over time, while Deadrick and Madigan (1990) empirically explored alternative explanations for changes in criterion-related validity over time (though in nonmanagerial settings).

Clearly, some very interesting, basic research started to appear in the decade since the first meta-analyses of criterion-related validities. We review these efforts and describe directions for future research below.

### ASSESSMENT CENTER CONSTRUCT VALIDITY

The history of assessment center development has been extensively documented elsewhere (MacKinnon, 1975; OSS, 1948; Thornton & Byham, 1982). We will not review that history here other than to place it in the context of a product life cycle. This selection "product," like so many others, originated in response to immediate war-time selection needs (OSS, 1948). It evolved into its earliest private sector form at AT&T in the late 1950s (Howard & Bray, 1988). Assessment centers in the 1960s through 1970s were characterized by market penetration and market expansion—criterion-related validity studies were conducted resulting in the evidence summarized in the Gaugler et al. (1987) meta-analyses.

Knowledge of why assessment centers demonstrate criterion-related validity has not proceeded at the same pace (Klimoski & Brickner, 1987; Klimoski & Strickland, 1977; Reilly & Warech, 1990). Assessment center architects originally constructed and marketed a product designed to evaluate "assessment dimensions" which were assumed to reflect basic managerial skills and abilities (Byham, 1970; Holmes, 1977; Howard & Bray, 1988). High fidelity exercises and simulations were constructed to capture key tasks and circumstances facing line managers (Crooks, 1977). Based on candidate behaviors in these content valid representations of managerial positions, trained observers (assessors) rated each candidate's managerial skills on the assessment dimensions provided. Candidates were usually evaluated in small groups. The assessors' task was to observe candidate behavior, decide which (if any) of the dimensions it represented, arrive at an evaluation of the total quantity and/or quality of each dimension a candidate had exhibited, and arrive at an overall assessment rating of the candidate's ability to performance in some targeted managerial position. In some instances, dimensional ratings were arrived at following each exercise which were later combined into an "overall" dimensional rating. The last step of arriving at an overall assessment rating usually involved some sort of consensus achieving discussion among multiple assessors, though discussion and consensus achieving efforts may have been involved earlier in the rating process.

Interestingly, in the early 1980s, Byham (1980) provided a different interpretation of constructs underlying assessment center dimensions in order to comply with the

*Uniform Guidelines* that many authors viewed as distinctly different from his earlier "skill and ability" interpretation (Byham, 1970). Specifically, Section 14 of the *Guidelines* stated that content validity validation strategies are not appropriate for methods that claim to measure psychological constructs (e.g., leadership and judgement). However, in Questions and Answers published by the EEOC to help interpret the *Guidelines*, Q & A #75 states "some selection procedures, while labeled as construct measures, may actually be samples of observable work behaviors. Whatever the label, if the operational definitions are in fact based upon observable work behaviors, a selection procedure measuring those behaviors may appropriately be supported by a content validity strategy." Byham (1980) argued that dimensions should be viewed as a "description under which behavior can be reliably classified" (p. 29) and that, when based on a thorough job analysis, content validation strategies are appropriate.

Unfortunately, some authors and practitioners have interpreted this argument to suggest that assessment center dimensions are "construct free," i.e., that dimension designations like "leadership" or "judgment" are just convenient labels for groups of content valid job behaviors. This would suggest that any labels would do, i.e., that the labels "category 1-10" could be used to randomly assign behavioral observations and arrive at ratings. Anyone familiar with the assessment process is aware that assessors take great care in sorting their observations into assessment dimensions. It is the "meaning" ascribed to these dimensions by assessors when sorting their observations that delineates one dimensional construct from another. Hence, while Byham's (1980) argument for the appropriateness of assessment center content validity strategies is sound, it is not appropriate to infer that psychological and/or job-related constructs are not used to sort these content valid behavioral observations.

Evidence regarding the construct(s) underlying test scores and/or ratings can take the form of similarity of test content to the construct domain (content validity), meaningful relationships with performance measures (criterion-related validity), and a pattern of relationships with other phenomena that is expected if ratings truly reflect the construct of interest (e.g., convergent and discriminant validity). As noted above, assessment center architects (at least initially) assumed that dimensional ratings reflect candidates' underlying managerial skills and abilities. The extensive evidence of criterion-related validity is noted above. Based on this mounting evidence of criterion-related validity, Norton (1977, 1981) argued that resemblance of assessment center exercises to the target position was adequate evidence of content validity alone and provided a sufficient basis for implementing an assessment center.

However, Sackett and others argued that rating assessment center dimensions is a very complex judgement task and that the complexity of this task directly affects our ability to draw construct *and* criterion-related validity inferences (Dreher & Sackett, 1981; Sackett, 1982, 1987; Sackett & Hakel, 1979; Sackett & Dreher, 1981; Russell, 1985). Investigators have documented the impact of many characteristics and influences on assessor judgments, including the impact of number of dimensional ratings assessors are required to make (Gaugler & Thornton, 1989), the general failure of assessors to follow judgement processes prescribed in assessor training (Russell, 1985; Sackett & Hakel, 1979), the impact of different quantities of assessor training (Dugan, 1988), the impact of sex and race composition of candidate groups (Schmitt & Hill, 1977), the impact of fixed vs. rotating assessor team membership (Russell, 1983), the impact of

behavioral checklists (Reilly, S. Henry, & Smither, 1990), and the impact of exercise form and content (Schneider, J. & Schmitt, in press).

It is clear the assessor judgments are very susceptible to the context in which their judgments are made. Sackett (1982, 1987; Sackett & Hakel, 1979) argued that the complexity of this judgment task precludes inferences of content validity based on correspondence between exercise content and job requirements—asking assessors to draw inferences about underlying managerial skills and abilities is too intricate to permit simple judgments of correspondence between exercise content and job content as the sole basis for use of an assessment center (see Schippman, Hughes, & Prien, 1987, and Schmitt & Noe, 1983, for a description of appropriate job analytic techniques in exercise development). Indeed, Motowidlo, Dunnette, and Carter (1990) presented evidence suggesting that ratings derived from low fidelity (i.e., low job correspondence) simulations yield comparable criterion-related validities.

Consequently, inferences concerning assessment center criterion-related and construct validity cannot be evaluated on the basis of exercise content alone. Fortunately, a large number of efforts have been mounted in the last 15 years to determine whether dimensional center ratings demonstrate convergent and discriminant validity with one another and/or with other measures of interest. For example, a number of assessment centers require assessors to arrive at dimensional ratings after observing candidates' behavior in each exercise. Post-exercise ratings are subsequently used to arrive at final, "overall" dimensional ratings prior to reaching a single overall assessment rating (consensus achieving discussion among assessors may occur at any point). Hence, if observations of candidate behaviors used to arrive at dimensional ratings truly reflect latent underlying managerial skills and abilities, ratings of dimension A on exercise 1 should be highly correlated with ratings of dimension A on exercise 2 and modestly correlated with ratings of dimensions B, C, and D (i.e., convergent and discriminant validity).

Neidig, Martin, and Yates (1979) were the first to examine these relationships, reporting moderate evidence of convergent validity (mono-dimension, heteroexercise correlations) and little evidence of discriminant validity (hetero-dimension, mono-exercise correlations). Though Sackett and Dreher (1982) found moderate evidence of convergent validity, numerous other authors found only limited evidence of convergent and discriminant validity (Archambeau, 1979; Bycio, Alvares, & Hahn, 1987; Russell, 1987; Schneider, J. & Schmitt, in press; Turnage & Muchinsky, 1982).

Further, if final dimensional ratings represent measures of independent managerial skill and ability constructs, factor analysis of K final (i.e., across exercise) dimensional ratings should yield approximately K independent factors. Factor analytic results consistently suggest loadings on 2-4 factors (Archambeau, 1979; Howard & Bray, 1988; Outlat, 1988; Russell, 1985; Sackett & Hakel, 1979). Two factor interpretations that appear consistently across these results are of "problem solving" and "interpersonal skill" factors.

Russell (1987) and Shore, Thornton, and McFarlane-Shore (1990) presented evidence of the convergence and/or divergence of dimensional ratings with independent measures of personality and/or knowledge. Russell's (1987) results indicated that an independent, self-report measure of interpersonal skills was moderately correlated with ratings derived from an exercise emphasizing interpersonal behaviors (an interview) and

uncorrelated with ratings on the same dimensions derived from an exercise with less emphasis on interpersonal behaviors (an in-basket). Shore et al. (1990) distinguished between "performance-style" and "interpersonal-style" assessment rating. They reported that performance-style dimensional ratings were more strongly correlated with independent measures of cognitive ability, while interpersonal-style ratings were more strongly correlated with conceptually similar personality measures derived from the 16PF.

Sackett and Dreher (1984) suggested that, in light of weak evidence of convergent and discriminant validity, dimensional ratings may simply represent how well candidates are behaving in ways congruent with managerial role requirements. Neidig and Neidig (1984) suggested that molar exercise-specific skill and ability constructs provide a more appropriate conceptualization. Finally, the results of Russell (1987) and Shore et al. (1990) imply that groups of dimensions (e.g., performance- and interpersonal-style) may be the most appropriate framework for viewing underlying constructs. In order to test these competing explanations, criterion and construct validity evidence should be gathered on assessment centers that have been explicitly designed to capture ratings effective of role requirements and molar vs. micro skills and abilities.

A recent series of studies have explicitly examined two of these alternative explanations. Russell and Domm (1990a) constructed an assessment center in which assessors were explicitly trained to view dimensions as role requirements of the target position (retail store manager). This training focused on substantially reducing the cognitive demands placed on assessors, reducing the rating process to a simple sorting and matching process (i.e., sorting behaviors into groups that "match" role expectations for the target position). Concurrent validity evidence suggested moderate correlations between overall assessment rating and superiors' performance appraisal ratings ( $r = .28$ ). However, in the first reported use of this criterion in the assessment center literature, overall assessment ratings were correlated at  $r = .32$  with net store profit (candidates with a rating of 3 or greater on their overall assessment rating earned profits that average \$1200 higher in the first quarter of 1989 relative to candidates with ratings of 2 or less).

In a second study, Russell and Domm (1991) evaluated an assessment center in which assessors rated 27 traditional assessment center dimensions (e.g., initiative, judgment, listening skill, etc.) and seven role requirements taken directly from a job analysis of the target position (e.g., working with subordinates, maintaining efficient quality production, maintaining equipment and machinery, etc.). Assessors arrived at two separate overall assessment ratings, one consisting of a simple sum of the 27 traditional assessment center dimensions and one based on assessors' clinical judgment based on the seven role requirement dimensions. Concurrent validity evidence was obtained from 170 incumbents in the target foreman position. Results indicated that the seven dimensional ratings based on role requirements had an average correlation of .11 ( $SD = .04$ ) with performance measures. The 27 traditional assessment center ratings had an average correlation of .18 ( $SD = .09$ ) with performance measures. The overall rating derived from the seven role requirement dimensions was correlated .28 ( $p < .01$ ) with supervisors' performance ratings, while the sum of the 27 traditional dimensional ratings did not yield a significant correlation ( $r = .14, p > .05$ ).

Finally, Huck and Russell (1989) reported preliminary evidence of criterion-related validity for an assessment center designed explicitly to evaluate managerial skills and

abilities. The center was designed as a community service provided through an extension management education center in a large southwestern state university. Managers from the surrounding metropolitan area attended a one-day series of exercises and simulations. The purpose of the center was to provide developmental feedback to participants—no one from the participants' parent firms ever received information obtained from the assessment center. Professional assessors from around the country with advanced training in clinical psychology were flown in to observe participants' behaviors and arrive at ratings on traditional assessment center skill and ability-oriented dimensions (e.g., social sensitivity, forcefulness, etc.). Due to the design of this center, none of the assessors were familiar with either the participants' firms *or* jobs. However, it is possible that assessors harbored a common, generalized occupational stereotype of managerial roles. A "management practices" survey was subsequently distributed to participants' subordinates, peers, and superior(s) asking respondents to evaluate the participants' performance on various generic management functions. Predictive validity results suggest meaningful correlations between dimensional ratings and performance ratings (.22-.34). The overall assessment rating was correlated  $r = .26$  ( $p < .05$ ) with supervisors' overall performance ratings.

Hence, it would appear that skill/ability explanations of assessment center construct validity remain viable, though a role congruency explanation receive stronger empirical support. Russell and Domm's (1990a, 1991) results suggest that assessors in an assessment center used to select managerial personnel can sort and match candidate behaviors according to role requirements of the target position in a way that yields criterion-related validity. Conversely, Huck and Russell's (1989) results suggest that in a situation where assessors cannot possibly know the role requirements of the target position (but may hold some general stereotype of what constitutes a managerial role), criterion-related validity is still forthcoming. Almost all the evidence failing to support a "managerial skill and ability" explanation of dimensional ratings has examined assessment centers designed for selection purposes using in-house assessors. It is possible that assessors find assessing abstract "managerial skill and ability" constructs too complex a task, reverting to a simple matching of observations with role requirements to arrive at dimensional ratings. Dugan (1988) reported finding that two weeks of traditional assessor training in rating dimensions as "skills and abilities" resulted in lower criterion-related validity relative to one week of such training, suggesting that training oriented toward concepts of managerial skills and abilities adds unnecessary complexity, and hence error, to the assessors' rating task. A reasonable assessor response to this complexity would be to simply sort and match behavioral observations to role requirements.

Regardless, meta-analyses indicate that ratings derived from behaviors observed in assessment center environments demonstrate criterion-related validity. The field has attempted in the last ten years to evaluate how various antecedent events impact indications of construct validity. The meaning assessors ascribe to their observations, and hence the evidence of construct validity that might be forthcoming, is highly dependent on the context within which their judgments are made.

### Future Research Needs

Future research needs to focus on how different approaches to defining assessment dimensions, assessor training, and exercise design impact evidence of construct and criterion-related validity. Specifically, if assessors can draw inferences about underlying managerial skills and abilities from behaviors observed in assessment exercises, investigators need to explore how evidence of construct validity is related to differences in (1) dimensional definitions; (2) how assessors are trained to infer a dimensional skill or ability from behavioral observations; (3) how exercise construction impacts candidates' opportunity to manifest skill-specific behaviors; and (4) cognitive demands placed on assessors. Results from such efforts will directly contribute to our knowledge of how individuals differ in underlying managerial skills and abilities.

In contrast, evidence supporting a role-congruency explanation of dimensional rating does not hold great promise for theory development—knowing that samples of behavioral observations obtained in assessment exercises “match” role requirements and predict performance criteria does not provide great insight into which individual difference characteristics contribute to managerial performance. However, Russell and Domm's (1990a) results suggest that assessment centers designed to reflect role requirements will continue to have profitable applications.

### BIODATA CONSTRUCT VALIDITY

As with the assessment center literature, the history of biographical information and its use in personnel selection has been extensively documented elsewhere (see the forthcoming *Handbook of Biographical Information* edited by Mumford, Stokes, and Owens) and will not be reviewed here. We focus on the numerous efforts at developing theories or models of constructs underlying biodata items that have been mounted over the last 25 years, though as recently noted by Campbell (1990), no “theory of experience” (p. 692) has been forthcoming. Owens (1968, 1971) was first to present a developmental/integrative (D/I) model, suggesting that prior life events are sources of individual development impacting future knowledge, skills, abilities, and motivation. This approach has been extended by Mumford and Stokes (in press) into an ecology model. This conceptualization describes a longitudinal sequence of interactions between the environment, a person's “resources” (e.g., human capital), and a person's “affordances” (needs, desires, choices).

Schoenfeldt (1974) integrated Owens' D/I model into an Assessment-Classification model, suggesting that models of work performance (i.e., Campbell, Dunnette, Lawler, and Weick's, 1970, person-process-product model) should be thought of as the result of the intersection of constructs taken from job taxonomies and prior life experiences. Finally, Kuhnert and his colleagues (Kuhnert & Lewis, 1987; Kuhnert & Russell, 1990) merged Bass' (1985) notions of transactional and transformational leadership with Kegan's (1982) Constructive/Developmental model of adult development to suggest that life experiences differentially impact the development of leadership skills at unique stages of value formation.

These models are valuable alternate conceptualizations of the construct domain underlying life history items, though most research efforts have not targeted life history

constructs that precede performance in managerial and leadership positions. Given that specific life history constructs are lacking, it is not surprising that these models generally do not provide strong guidance for operationalization. No explicit direction has been forthcoming in how to design paper and pencil life history inventories to test hypotheses about how aspects of the environment, “resources,” and “affordance” evolve and interact to impact criteria of interest (though Mumford, Uhlman, & Kilcullen, in press, provide an interesting description of how this was done with a student sample). Guilford (1959), Dunnette (1963), and E. Henry’s (1966) early criticisms of biodata research as exercises in atheoretical empiricism will continue until specific relationships between theory, item content, and criterion performance measures are identified.

Efforts at exploring these linkages fall predominately into two categories: (1) a number of post hoc efforts involving biodata factor interpretation (including both item loadings and subsequent convergent and discriminant validities with other measures of interest); or (2) some sequence of a prior theory-based effort at item development followed by confirmatory factor analysis.<sup>1</sup> Most of the post hoc factor interpretation efforts have used the same instrument with samples of college students. We briefly review selected results reported from biodata factor interpretations before examining theory-based efforts as item generation.

### **Post Hoc Factor Interpretation**

Owens and Schoenfeldt (1979) conducted perhaps the single most comprehensive empirical investigation of factors underlying responses to biodata items. They administered a biodata instrument to 9764 freshman over a five-year period, identified an underlying factor structure, then reported substantial evidence of construct validity in the relationships of factor scores to over 84 independent measures (including the Strong Vocational Interest Blank, high school grade-point average, SAT scores, Rotter’s I-E scale, the California F scales, etc.). The factor labels for males and females are reported in Table 2.

These factors are very interesting in that evidence suggests they are construct valid dimensions of life experience that are differentially related to a large number of other measures. The measures of most organizational relevance employed by Owens and Schoenfeldt were the Strong Vocational Interest Blank and voluntary turnover (unfortunately, no measure of job performance was available).

Shaffer, Saunder, and Owens (1986) found additional support for these factors in a longitudinal design using independent observer ratings. In an independent sample, Eberhardt and Muchinsky (1982a) replicated the factors for males, while achieving only partial replication for females. They speculated that changes in sex role between the late 1960’s and early 1980’s may have contributed to changes in what constituted meaningful life history constructs for females.

Davis (1984) and Stokes, Mumford, and Owens (1989) have demonstrated that factors derived from college student responses to the Owens and Schoenfeldt (1979) instrument were predictive of life history factors derived from a different instrument later in life (i.e., a post-college experience inventory). Neiner and Owens (1982) found that undergraduate responses to the Owens and Schoenfeldt instrument were predictive of job choice six years later (three to four years after college graduation). Finally,

**Table 2**  
**Biodata Factors Derived by Owens and Schoenfeldt (1979)**

<i>Male</i>	<i>Female</i>
1. Warmth of parental relationship	1. Warmth of maternal relationship
2. Intellectualism	2. Social leadership
3. Academic achievement	3. Academic achievement
4. Social introversion	4. Parental control vs. freedom
5. Scientific interest	5. Cultural-literary interests
6. Socioeconomic status	6. Scientific interest
7. Aggressiveness/independence	7. Socioeconomic status
8. Parental control vs. freedom	8. Expression of negative emotions
9. Positive academic attitude	9. Athletic participation
10. Sibling friction	10. Feelings of social inadequacy
11. Religious activity	11. Adjustment
12. Athletic interest	12. Popularity with opposite sex
13. Social desirability	13. Positive academic attitude
	14. Warmth of parental relationship
	15. Social maturity

Eberhardt and Muchinsky (1982b, 1984) found that factors derived from the Owens and Schoenfeldt instrument were predictive of responses and "types" derived from Holland's Vocational Interest scale.

While the greatest number of construct validity efforts have arisen from Owens' pioneering research, a few investigators have developed and examined independent biodata questionnaires and sources of life history information. Russell, Mattson, Devlin, and Atwater (1990) developed a biodata questionnaire to predict officer performance of applicants considered for admission to the U.S. Naval Academy. Incidents were extracted from life history essays written by freshmen at the Academy. Aspects of the criterion construct domain were used to help extract items (no a prior life history constructs were used). Factor analysis resulted in one dominant factor reflecting life problems and difficulties, followed by four smaller factors reflecting samples or signs of the criterion construct domain (aspects of task performance, work ethic/self discipline, assistance from others, and extraordinary goals or effort).

Two recent studies gathered life history information from top level managers through structured interviews (Russell, 1990) and middle level military officers through written life histories (Bettin & Kennedy, 1990). Both studies used experts to (1) sort life history experiences according to the relevance to aspects of the criterion domain and (2) make forecasts of future job performance. Bettin and Kennedy reported a concurrent validity of .43 with superiors' performance ratings, while Russell reported a concurrent validity of .40 with superiors' performance ratings and .49 with their most recent merit pay increase. While no attempt was made to identify constructs from the *life history domain*, evidence from the two studies strongly suggests that raw life history data can be filtered according to constructs from the *criterion domain* in a meaningful way.

In perhaps the only attempts to explore the life history construct domain of top level managers, Lindsey, Homes, and McCall (1987) and Bobko, McCoy, and McCauley (1988) identified a number of "lessons" derived from experience that "successful" top

level executives felt contributed to their development. The 16 “key events” that were derived from a subjective “factoring” of these lessons fell into (1) five different types of developmental job assignments, (2) five different types of hardships or life problems, (3) two types of interaction with other people, and (4) four “other” significant events (i.e., course work, early work, first supervisor, and purely personal experiences). While very suggestive, strong conclusions are precluded by the absence of data to evaluate any relationship between these lessons and organizationally relevant criteria.

### Item Development Efforts

Of course, any post hoc interpretation of item loadings is dependent on the nature of the items analyzed. Investigators who have described item generation procedures tend to have used a team of doctoral students or other investigators to generate a large pool of items based on some broad specifications of either the criterion and/or life history construct domains. Unfortunately, the vast majority of investigators do not describe how their biodata items were developed (Russell, in press).

Owens and Schoenfeldt (1979) initially generated their items from two hypothetical categories of prior life experiences: inputs to the organism and behaviors (signs and samples) of the organism. Inputs were conceived of as treatment variables and can be thought of as all environmental circumstances including “parents, teachers, siblings, close friends, school experience, and free time activities” (p. 574).

Unfortunately, notions of prior “inputs” and “behaviors” are too broad to provide much guidance for management selection. The factors derived from Owens and Schoenfeldt’s input and behavior items *may* provide such guidance in the future. For example, one might suspect a priori that performance of managers in scientific research and development settings might best be predicted by items loading on factors dealing with both science and interpersonal relations (e.g., warmth of parental relationship, intellectualism, academic achievement, social introversion, scientific interest, and sibling friction for males). Mumford, Uhlman, and Kilcullen (in press) described how different predictions drawn from the ecology model could be used to develop scales from the Owens and Schoenfeldt instrument that demonstrate criterion-related validity in student samples. Of course, a complete test of such a hypothesis would require generation of a broader pool of items reflecting these dimensions in later life experiences that are not currently captured by Owens and Schoenfeldt’s instrument.

Russell and Domm (1990b) developed and validated a biodata instrument for the selection of retail store managers. Using Schoenfeldt’s (1974) Assessment/Classification model, they hypothesized that whatever the relevant constructs were underlying prior life experiences, they should (1) be related to current job performance and (2) embody person characteristics, behaviors, task outcomes, and environmental characteristics (Campbell et al.’s person-process-product model). Russell and Domm (1990b) asked current store managers to write life history essays about prior life experiences they felt were related to specific dimensions of their current job. Incidents were coded from these essays under the headings of person characteristics, behaviors, task outcomes, and environmental characteristics. Cross validation indicated an average correlation of .37

with ratings of job performance. Factor analysis resulted in one dominant factor reflecting life problems and difficulties, followed by three smaller factors labeled dealing with groups of people/information, focused work efforts, and dealing with rules and regulations.

Interestingly, even though the samples and biodata items were completely different, Russell and Domm (1990b) and Russell et al. (1990) found items reflecting life problems and difficulties loading dominantly on the first factor extracted. Russell and Domm (1990b) did not expect factor analysis to yield loadings in direct correspondence with Campbell et al.'s (1970) person-process-product constructs *or* with the dimensions of job performance used as stimulus material for the life history essays. Instead, they expected some "hybrid" factors to emerge which contained person characteristics, behavioral processes, task outcomes, and environmental characteristics that capture key aspects of some developmental episode. Hence, they interpreted the dominant life problems factor as a possible example of a life history construct that contributes to managerial development. Russell and Domm (1990b) interpreted the remaining factors as samples of prior experiences that have a close correspondence with managerial job requirements that may or may not constitute key developmental episodes.

Finally, Delery and Schoenfeldt (1990) developed items reflective of four dimensions of managerial role requirements (managerial functions, relational factors, targets or goals, and style elements). Factor analysis performed on responses from 830 entering college students resulted in 11 interpretable factors (technical expertise, management functions, influence/control, evaluation/decision making {intrinsic}, leadership, career forecasting/planning, energy level, evaluation/decision making {advancement}, persuasive communication, evaluation/decision making {college choice}). In a separate sample of 310 students they found statistically significant and practically meaningful correlations between biodata scale scores and questionnaire measures of organizational involvement, athletic involvement, volunteer activity, employment activity, and honors.

### **Future Research Needs**

It has been well established that biographical information items demonstrate criterion-related validity in managerial and nonmanagerial settings. For biodata to achieve its full potential in managerial development and selection, future research must develop items that representatively sample aspects of key developmental processes. Meaningful life history constructs have been strongly suggested in the pioneering work of Owens and Schoenfeldt (1979). Efforts by Delery and Schoenfeldt (1990) and Russell and his colleagues (Russell et al., 1990; Russell & Domm, 1990b) suggest alternate constructs that initially appear to be more relevant to predicting managerial job performance. Regardless, the relevance of these constructs for managerial selection will remain in doubt until new pools of post-college life history items are developed (Mael, 1991; Russell, in press). Identification of relevant life history constructs followed by theory-driven item development strategies should yield major advances in our understanding and prediction of managerial performance.

## EVOLVING CONSTRUCT AND CRITERION-RELATED VALIDITY ISSUES IN TESTS OF COGNITIVE SKILL AND ABILITY

Over the past century there has been great debate on the nature, meaning, measurement and role of intelligence in predicting job performance. Unlike the assessment center and biodata domains, much of this debate was theory driven (e.g., Binet & Henri, 1895) and considerable attention was placed on the various conceptualizations of intelligence *and* how well intelligence tests predicted job performance (Thorndike, Lay, & Dean, 1909). Historically, a key distinction made by researchers is whether intelligence is a general intellectual ability or whether it is composed of many specific aptitudes or skills, such as verbal, quantitative, and technical skills. Over this time, many researchers from Spearman (correlation), to Thurstone (factor analysis), to Hunter (validity generalization) have used new advances in statistics and methodology to support their positions. Hulin has examined the impact of time lag between cognitive predictors and criterion (Henry, R. & Hulin, 1987, 1989; Hulin, Henry, R., & Noon, 1990), adding a new consideration to the historic concerns of construct and criterion-related validity. This debate is fundamental to understanding practical validation attempts, as well as understanding the nature of intellectual ability itself. Because there is no clear answer to whether intelligence is more than a simple sum of skills and aptitudes, much of the research evidence we review continues to follow, for better and worse, in the steps of our predecessors. Brief descriptions of the conceptual issues and evidence of criterion-related validity are presented, followed by a discussion of how time lag between predictor and criterion impacts our interpretation of prior findings.

### Conceptualizations of Cognitive Skills and Ability

Beginning as early as 1904, Charles Spearman proposed a theory of intellectual functioning which was based on the observation that all intelligence tests correlated to a greater or lesser extent with each other. Spearman (1927) hypothesized that the proportion of the variance that the tests had in common accounted for a general (or *g*) factor of intelligence. The remaining portions of the variance were either accounted for by some specific (*s*) component of this general factor or by error (*e*). Tests that correlated with other tests of cognitive ability were thought to be highly saturated with the *g* factor. Those with no *g* factor were not considered to assess intelligence or cognition, but rather were viewed as tests of pure sensory, motor, or personality traits. Thus it was *g* rather than *s* that was assumed to afford the best measure of overall intelligence.

Continued support for Spearman's initial observations comes from the almost universal finding that batteries of tests designed to tap specific cognitive skills tend to be highly positively correlated in large, representative samples of the population. These matrices of positive correlations came to be labeled the "positive manifold," and dictate that hierarchical factor analytic techniques will yield a single dominant factor solution that can be readily interpreted as *g* (Jensen, 1986). As a result, Spearman (1927) and others have been very careful to define *g* in terms of a factor analytic result unique to a given battery of test items.

Today the term cognitive ability or *g* is often used synonymously with intelligence and is defined as a very general capacity for abstract thinking, problem solving, and

learning complex things (Snyderman & Rothman, 1986), while *s* is defined as a specific skill or aptitude that is acquired after training, and consists of competent, expert, rapid, and accurate performance (Welford, 1968). General cognitive ability is thought to change slowly and relate to the ease and speed with which knowledge is gained, while *s* relates to skills that are acquired less slowly and with training (Adams, 1987). Kanfer and Ackerman (1989), building on the work of classical learning theorists, developed and provided initial support for a model of skill acquisition that positions *g* as a central, early determinant in some “near-term” skill acquisition process.

### Criterion-Related Validities

Most of the evidence for *g* as a predictor in job performance derives from research conducted on the Armed Services Vocational Aptitude Battery (ASVAB) and the General Aptitude Test Battery (GATB). Prospective new recruits in all the armed services are administered the ASVAB, a test that consists of 334 multiple-choice items organized into 10 different subtests. In general, the test has been deemed to be quite useful in making selection and placement decisions in the armed forces (Grunzke, Gunn, & Stauffer, 1970). The GATB is a tool used to identify aptitudes for occupations, and is available for use by state employment services and other governmental agencies. In recent years the GATB has evolved from a test with multiple cutoffs to one that employs regression and validity generalization for making recommendations based on test results. The rationale and process by which the GATB has made this evolution has been described by Hunter (1980, 1986) and his associates (Hunter & Schmidt, 1983; Hunter & Hunter, 1984).

In contrast to the use of global measures of *g* in selection, Hull (1928) suggested that prediction could be improved with composite scores derived from optimally weighted combinations of measures of specific aptitudes and abilities (*s*). Multiple regression procedures were used to achieve optimal weights. Many investigators have failed in their attempts to discover such “optimal” weights that differentially predict performance across jobs (see Humphreys, 1986, for his description of unsuccessful efforts at discovering stable differential validity in the U.S. Air Force in the early 1950s). Thorndike (1986) and Hunter (1986) present compelling evidence (with the GATB and ASVAB, respectively) that optimally weighted composites do not explain meaningful variance in performance criteria beyond that explained by *g*.

Cognitive ability tests, however, do not predict performance equally well in all jobs (Schmitt, Gooding, Noe, Kirsch, 1984) or across all criteria (Zeidner, 1987), but the pattern of predictive validities is neither random, nor sensitive to a job’s task content. In fact, validities fall into a very interesting pattern—the higher the job level or complexity (regardless of job content), the better the cognitive tests predict performance (Hunter, 1986).

Consequently, when considering the full range of jobs in industrialized economics, but particularly more complex managerial jobs, cognitive ability tests are the most important among known predictors of job performance (Hunter & Hunter, 1984). Meta-analyses of relationships between cognitive ability measures and job performance have been well documented by Schmidt, Hunter and their colleagues (Pearlman, Schmidt, & Hunter, 1980; Schmidt, Gast-Rosenberg, & Hunter, 1980). After the data for

thousands of validity studies had been assembled and meta-analyses performed, these authors concluded that since most variance across studies can be linked to sampling error, there is no real variation in the validity of tests across settings.

Regardless, at least two concerns remain before the unqualified use of cognitive ability tests can be endorsed for all managerial selection applications.

### **Level of Analysis, Time Lag, and Labor Pool as a Moderator of Validities**

Perhaps the most cited reason why early managerial performance is not correlated with later job performances is that the nature of managerial work varies according to organizational level (Jaques, 1976, 1989; Mintzberg, 1973; Simon, 1977; Torbert, 1987). According to these authors, moving from one level of management to another requires more than a change in one's tasks and responsibilities, but to a fundamental transition entailing a qualitative change in the conceptual demands of the required work. In terms of managerial selection this means predictor (and criterion) constructs are likely to be different as individuals move from one organizational level to another and may explain why validities change over time.

Regardless, it has long been known that early performance is not necessarily correlated with latter performance, in learning or job performance situations (Blankenship & Taylor, 1983; Kornhauser, 1923; Henry, R. & Hulin, 1987, 1989). Of greater concern is Hulin, R. Henry, and Noon's (1990) report of temporal trends in predictive validity across 41 studies and 77 validity sequences. Predictors spanned a wide variety of cognitive and psychomotor tests ranging from Law School Admissions Test and undergraduate GPA to rotary pursuit tasks and two-handed coordination tests. Hulin et al. reported substantial evidence of decreasing trends in validity across all tasks and/or jobs over time. Their results suggest that over time either (1) combinations of abilities required for job performance change and/or (2) incumbents' skills and abilities change. Unfortunately, Hulin et al. (1990) could not address these competing explanations as their pool of studies did not differentiate between (1) measures of *g* vs. *s* or (2) complex vs. less complex jobs.

Howard and Bray (1988) and Bray, Campbell, and Grant (1974) report validities between measures of cognitive ability and level of management obtained at 8 and 20 year time lags in the AT&T management progress study (results not included in the Hulin et al. meta-analysis). It is probably safe to assume that rate of promotion is a good measure of performance in a career track involving complex managerial positions. Bray et al. (1974) report a correlation of .19 ( $p < .05$ ) between the School and College Ability test (SCAT) and level of management obtained at year 8, while Howard and Bray report a correlation of .38 in year 20! The year 8 correlation was probably attenuated due to range restriction in the criterion. Curiously, when the sample at year 20 is broken into college and noncollege graduate subsamples (not done at year 8), Howard and Bray (1988) report that validities fall to .20 and .28, respectively. Regardless, it certainly does not suggest a decreasing trend in validity. It is clear that more research needs to address these findings at managerial levels.

Unfortunately, the issues become even more complex when one relaxes the assumption of a large heterogeneous applicant pool that characterizes most of the military and civil service applications underlying Hunter's efforts (Hunter & Hunter,

1984; Hunter, 1986). Many jobs are characterized by highly pre-screened, homogeneous labor markets in which zero or negative correlations among cognitive skills tests might be evident. For example, many candidates for middle- and upper-level management positions come from different functional or operational areas of the firm. A pool of candidates for upper-level management positions in finance and operations at a large multi-national corporation may consist of individuals from financial or accounting functions (high quantitative, low verbal) and retail marketing functions (low quantitative, high verbal). This could lead to a violation of the positive manifold described by Jensen (1986), resulting in the absence of an empirically identifiable  $g$  factor (even though use of an instrument endowed with a meta-analytic endorsement, like the GATB, would still result in a composite  $g$  score). In addition, substantial differential criterion-related validity may be obtained from composite scores derived to predict performance in upper-level financial management positions vs. upper-level operations positions. Different composites of  $s$  might exhibit differential criterion-related validity with middle- or upper-level management positions due to severe range restriction on  $g$  and different levels of variation on  $s$  factors by subgroups in the labor market. Statistically “correcting” for such range restriction would be inappropriate, since the range restriction is a true characteristic of the labor market and not a sampling artifact. Reanalysis of the Schmitt et al. (1984) data reported in Table 1 resulting in a substantially reduced average  $r$  of .17 lends some support to this speculation as Schmitt et al.’s data did not contain any studies using the GATB or ASVAB. Instead, their studies focused on more homogeneous labor pools faced by private sector firms.<sup>2</sup>

### Future Research Needs

Numerous questions remain concerning the long term relationship of  $g$  and  $s$  to managerial job performance. Initially, we need to determine whether decrements in criterion-related validity over time are a function of job complexity,  $g$ ,  $s$ , and/or alternate composites of  $s$ . Ackerman (1987) and Kanfer and Ackerman (1989) provide substantial evidence that changes in criterion-related validity in skill acquisition environments are related to  $g$ ,  $s$ , and job complexity—it remains to be seen whether their results generalize to job performance in managerial positions.

Second, we need to explore the effects of nonheterogeneous labor markets on variance in  $g$ ,  $s$ , and criterion-related validities over time. Evidence supporting Hull’s (1928) original hypothesis that optimally weighted composites of  $s$  exhibit differential validity *may* be forthcoming in highly pre-screened, nonheterogeneous labor markets characterizing middle- and upper-level management positions.

## SUMMARY

### Predictor Domain

Our purpose has been to examine construct validity evidence pertaining to three dominant managerial selection technologies. Evidence suggests two of these technologies may be directly tapping personal skill and ability constructs (cognitive ability tests and assessment centers) while the third is capturing antecedent life history

constructs that causally impact development of and/or are correlationally related to cognitive ability constructs. It is the simultaneous exploration of each technology's research needs that holds the most promise for developing a theory of selection, and paying dividends by augmenting criterion validities produced by meta-analyses.

The results reviewed herein suggest a preliminary theory of managerial selection would look very similar to Mumford et al.'s (in press) ecology model, involving an iterative sequence of life events (school, siblings, early job experiences, etc.) impacting the development of *g* and specific profiles of *s*'s. This sequence would, over time, contain paths of reciprocal causality where individuals' levels of *g* and/or profiles of *s*'s subsequently impact the types of life experiences they are exposed to. Ryanen (1988) argued that the "push" factors of civil rights legislation and guidelines had the inadvertent effect of steering personnel selection efforts away from the "research-oriented personnel selection" needed to flesh-out approaches like the ecology model. Instead, emphasis was shifted to "compliance-oriented personnel selection" efforts (p. 380). It is our belief that a collision between (1) the need for a theory of personnel selection and (2) existing measurement technologies geared to meet *EEOC Guidelines* should prove fruitful and is long overdue.

Specifically, Hulin et al.'s (1990) meta-analyses forcefully impose on future investigators the task of examining relationships between alternate measures of *g*, *s*, and job performance longitudinally. The AT&T management progress study is the only longitudinal effort to date containing multiple measures of skills and abilities (using paper and pencil cognitive ability tests and assessment center ratings) that also tracks critical historical events for a cohort of managers. While very suggestive, these results will always be open to question in the absence of any convergent and discriminant validity evidence supporting the interpretation of constructs drawn from assessment center ratings.

Many other studies, taken as a whole, are very suggestive of the types of life experiences that precede the development of managerial capacity (i.e., *g* and *s* profile). It should be clear that future research must focus on how constructs derived from Owens and Schoenfeldt's (1979) factor interpretations or Lindsey et al.'s (1987) key events in managerial lives impact the development of *g* and profiles of *s*'s *as well as* variance in managerial performance over time. Further, it should also be clear that Hull's (1928) hypothesis that *s* composites will amplify the predictive power of *g* needs to be re-evaluated in the prescreened, nonheterogeneous internal labor markets found at middle- and upper-level management ranks. Bobko, Colella, and Russell (1990) and Russell, Colella, and Bobko (1991) recently extended traditionally utility formulations to account for differential predictions of criteria over time, suggesting substantial modifications in how a selection system's value to the organization is determined. They incorporated (1) Hulin et al.'s (1990) results suggesting decreasing criterion-related validities over time with (2) alternate strategic directions adopted by the firm. Bobko et al. (1990) and Russell et al. (1991) demonstrate common circumstances where the predictors described above could yield *negative* economic outcomes for the firm.

### Performance Domain

Noticeable by its absence in the above discussion is a thorough exploration of the criterion performance domain. The casual reader will note this is not due to an absence

of deliberation about managerial performance in the literature. Management "functions" have been the source of ongoing debate for years (see Carrol & Gillen, 1987, for an interesting comparison of Fayol's [1949] taxonomy and Mintzberg's [1973] observations). Zalesnik (1977, 1990) has noted the distinction between "management" and "leadership." Further, an emerging literature on entrepreneurship is currently debating trait vs. behavior oriented construct definitions and predictors (Carland, Hoy, & Carland, 1988).

Campbell (1990) presented perhaps the most thorough discussion of these issues and a preliminary taxonomy of the criterion performance domain. He defines performance as behavior, effectiveness as an evaluation of performance relative to some goal, and productivity as a measure of effectiveness relative to resources used. Descriptions of various approaches to managerial positions are brimming with references to "behaviors," "tasks," "roles," "responsibilities," and "duties" (see Carrol & Gillen's, 1987, discussion of managerial positions). Some authors have offered prescriptions regarding the types of skills found in effective administrators (Katz, 1974). Very few authors have explicitly attempted to link performance requirements (behavioral or nonbehavioral) with alternate measures of specific cognitive skills (see Fleishman & Mumford, 1991, for an exception).

Regardless of the various ways one might partition managerial performance, the relative importance and contribution of any managerial behavior will be dictated by how that managerial position contributes to organizational goals and objectives. We are aware of no studies that look at organizational goals and objectives as a moderator of criterion-related validities obtained. This leads directly to the specter of situation specificity, albeit where the "situation" is the firms' strategic goals and objectives (or whatever are most salient for the position at hand). Smith (1976) reviewed the various calls for weighted criteria throughout this century. However, with the notable exception of Schneider, B. (1987), few investigators even mention the broader organizational context. Again, Bobko et al. (1990) and Russell et al. (1991) demonstrated that firms' strategic goals and objectives can have an extreme impact on the value added of a selection system.

### Conclusion

It is tempting to return to Vroom's preface to Dunnette (1966) and suggest that the recent developments bode well for future theory development in selection. In fact, it looks as if the U.S. Congress, in raising one of the old "push" factors of civil rights legislation, is currently raising issues that are best met by construct validity and theory development efforts. It certainly would be nice to answer questions raised by the judicial and legislative branches with a model explaining *why* individuals with certain attributes perform better at work, rather than to watch laymen's eyes glaze over as we explain technical distinctions between the 4/5's rule, differential prediction, content, and construct validity. Recent attempts by Binning and Barrett (1989) and Landy (1986) to embed traditional trinitarian approaches to test validation (Guion, 1980) in a more global construct validation and theory testing approach provide explicit direction on how to start.

We feel the 25 years of research since Vroom's comments has placed the field in a position where the problems faced by our organizations and society can be best addressed by focusing on "the processes that underlie various behavioral phenomena, on the assumption that improvements in technology will be facilitated by a better understanding of these processes" (Vroom, as cited in Dunnette, 1966, pp. v-vi). It remains to be seen whether progress toward this objective is made in the next 25 years.

## ACKNOWLEDGMENTS

We would like to thank Rebecca Henry and Kevin Mossholder for comments on an earlier version of this manuscript and Stephen K. Markham for his data analysis assistance.

## NOTES

1. Many empirical efforts focus on identifying homogeneous groups of subjects who have similar profiles of prior life experiences, usually using a clustering algorithm to identify groups with similar patterns of factors scores derived from some life history inventory. Unfortunately, almost all of this research as focused on adolescents and/or young adults (Owens & Schoenfeldt, 1979). Further, as noted by Barge (1988), none of these efforts have related subgroup membership to performance criteria (at managerial *or* nonmanagerial levels). Consequently, while we feel that *configurations* of life history constructs will provide meaningful information above and beyond individual life history constructs themselves, research to date has not provided insight into what these configurations might be.

2. The largest data sets containing *N*'s of 8,885 males and 4,846 females found in the Schmitt et al. (1984) data came from a study reported by Moses and Boehm (1975). The 4,936 females consist of a highly prescreened sample being considered for promotion into management positions for purposes of compliance with a U.S. Justice Department consent decree with AT&T. The sample of 8,885 males was originally reported by Moses (1972). It can be considered moderately screened in that it consisted of males attending an operating assessment center at AT&T, the vast majority of whom were recommended (prescreened) for assessment by their immediate superior.

## REFERENCES

- Abrahams, N.M. & I. Neumann. (1973). *The validation of the Strong Vocational Interest Blank for predicting naval academy disenrollment and military aptitude*. (Technical Bulletin STB 73-3). San Diego: Naval Personnel and Training Research Laboratory.
- Ackerman, P.L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin* 102, 3-27.
- Ackerman, P.L. (1989). Within-task intercorrelations of skilled performance: Implications for predicting individual differences? *Journal of Applied Psychology* 74, 360-364.
- Adams, J.A. (1987). Historical review and appraisal of research on the learning, retention and transfer of human motor skills. *Psychological Bulletin* 101, 41-74.
- Archambeau, D.J. (1979). Relationships among skill ratings assigned in an assessment center. *Journal of Assessment Center Technology* 2, 7-20.
- Barge, B.N. (1987, August). Characteristics of biodata items and their relationships to validity. Presented at the 95th annual meeting of the American Psychological Association, New York.

- Barge, B.N. (1988). Characteristics of biodata items and their relationship to validity. Unpublished doctoral dissertation, University of Minnesota.
- Bass, B.M. (1985). *Leadership and performance beyond expectations*. New York: Free Press.
- Bettin, P.J. & J.K., Kennedy, Jr. (1990). Leadership experience and leader performance: Some empirical support at last. *Leadership Quarterly* 1(4), 219-228.
- Binet, A. & V. Henri. (1895). La psychologie individuelle. *L'Annee Psychologique* 1, 1-23.
- Binning, J.F. & G.V. Barrett. (1989). Validity of personnel decisions: A conceptual analysis of the inferential and evidential bases. *Journal of Applied Psychology* 74, 478-494.
- Blankenship, A.B. & H.R. Taylor. (1938). Prediction of vocational proficiency in three machine operations. *Journal of Applied Psychology* 22, 518-526.
- Bobko, P. & C.J. Russell. (1991). A review of the role of taxonomies in human resource management. *Human Resources Management Review* 1, 293-316.
- Bobko, P., A. Colella & C.J. Russell. (1990). *Estimation of selection utility with multiple predictors in the presence of multiple performance criteria, differential validity through time, and strategic goals*. Battelle Contract No. DAAL03-86-D-0001.
- Bobko, P., C. McCoy, & C. McCauley. (1988). Towards a taxonomy of executive lessons: A similarity analysis of executives' self-reports. *International Journal of Management* 5, 375-384.
- Bray, D.W. (1964). The management progress study. *American Psychologist* 19, 41-429.
- Bray, D.W., R.J. Campbell & D.L. Grant. (1974). *Formative years in business: A long-term AT&T study of managerial lives*. Malabar, FL: Robert E. Krieger Publishing.
- Burke, M.J. & K. Pearlman. (1988). Recruiting, selecting, and matching people with jobs. In J.P. Campbell & R.J. Campbell (Eds.), *Productivity in organizations* (pp. 97-142). San Francisco, CA: Jossey-Bass.
- Bycio, P., K.M. Alvares, & J. Hahn. (1987). Situation specificity in assessment center ratings: A confirmatory factor analysis. *Journal of Applied Psychology* 72, 463-474.
- Byham, W.C. (1970). Assessment centers for spotting future managers. *Harvard Business Review* 48(4), 150-160.
- Byham, W.C. (1980, February). Starting an assessment center the right way. *Personnel Administrator*, pp. 27-32.
- Campbell, J.P. (1990). Modeling the performance prediction problem in industrial-organizational psychology. In M.D. Dunnette and L.M. Hough (Eds.), *The handbook of industrial and organizational psychology* (vol. 1, pp. 687-732). Palo Alto, CA: Consulting Psychologists Press.
- Campbell, J.P., M.D. Dunnette, E.E. Lawler, III, & K. Weick. (1970). *Managerial behavior and performance effectiveness*. New York: McGraw-Hill.
- Carland, J.W., F. Hoy & J.C. Carland. (1988). "Who is an entrepreneur?" is a question worth asking. *American Journal of Small Business* 12(4), 33-39.
- Carrol, S.J. & D.J. Gillen. (1987). Are the classical management functions useful in describing managerial work? *Academy of Management Review* 12, 38-51.
- Cleary, T.A. (1968). Test bias: Prediction of grades of Negro and white students in integrated colleges. *Journal of Applied Psychology* 5, 115-124.
- Cohen, B.M., J.L. Moses & W.C. Byham. (1974). *The validity of assessment centers: A literature review*. Monograph II. Pittsburgh, PA: Development Dimensions Press.
- Crooks, L.A. (1977). The selection and development of assessment center techniques. In J.L. Moses and W.A. Byham (Eds.), *Applying the assessment center method* (pp. 69-98). New York: Pergamon Press.
- Davis, K.R. (1984). A longitudinal analysis of biographical subgroups using Owens' developmental-integrative model. *Personnel Psychology* 37, 1-14.

- Deadrick, D.L. & R.M. Madigan. (1990). Dynamic criteria revisited: A longitudinal study of performance stability and predictive validity. *Personnel Psychology* 43, 717-744.
- Delery, J.E. & L.F. Schoenfeldt. (1990). The initial validation of a biographical inventory to identify management talent. Presented at the 50th annual meetings of the Academy of Management, San Francisco.
- Dreher, G.F. & P.R. Sackett. (1981). Some problems with applying content validity evidence to assessment center procedures. *Academy of Management Review* 6, 551-560.
- Dugan, B. (1988). Effects of assessor training on information use. *Journal of Applied Psychology* 73, 743-748.
- Dunnette, M.D. (1963). A modified model for test validation and selection research. *Journal of Applied Psychology* 47, 317-323.
- Dunnette, M.D. (1966). *Personnel selection and placement*. Belmont, CA: Wadsworth.
- Dunnette, M.D. (1971). Multiple assessment procedures in identifying an developing managerial talent. In P. McReynolds (Ed.), *Advances in psychological assessment* (vol. 2). Palo Alto, CA: Science and Behavior Books.
- Eberhardt, B.J. & P.M. Muchinsky. (1982a). An empirical investigation of the factor stability of Owens' biographical questionnaire. *Journal of Applied Psychology* 67, 138-145.
- Eberhardt, B.J. & P.M. Muchinsky. (1982b). Biodata determinants of vocational typology: An integration of two paradigm. *Journal of Applied Psychology* 67, 714-727.
- Eberhardt, B.J. & P.M. Muchinsky. (1984). Structural validation of Holland's hexagonal model: Vocational classification through the use of biodata. *Journal of Applied Psychology* 69, 174-181.
- Fayol, H. (1949). *General and industrial management* (C. Storrs, Translator). London, England: Pitman.
- Fleishman, E.A. (1988). Some new frontiers in personnel selection research. *Personnel Psychology* 41, 679-701.
- Fleishman, E.A. & M.D. Mumford. (1991). Evaluating classifications of job behavior: A construct validation of the ability requirements scale. *Personnel Psychology* 44, 523-576.
- Gaugler, B.B. & G.C. Thornton, III. (1989). Number of assessment center dimensions as a determinant of assessor accuracy. *Journal of Applied Psychology* 74, 611-618.
- Gaugler, B.B., D.B. Rosenthal, G.C. Thornton, III, & C. Benston. (1987). Meta analysis of assessment center validity. Monograph. *Journal of Applied Psychology* 72, 493-511.
- Ghiselli, E.E. (1956). Differentiation of individuals in terms of their predictability. *Journal of Applied Psychology* 40, 374-377.
- Ghiselli, E.E. (1966). *The validity of occupational aptitude tests*. New York: Wiley.
- Ghiselli, E.E. (1973). The validity of aptitude tests in personnel selection. *Personnel Psychology* 26, 461-477.
- Glaser, B.G. & A.T. Strauss. (1967). *The discovery of grounded theory: Strategies for qualitative research*. New York: Aldine Publishing.
- Griggs v. Duke Power Company*, 410 U.S. 424 (1971).
- Grunzke, N., N. Gunn, & G. Stauffer. (1970). *Comparative performance of low-ability airmen* (Technical Report 70-4). Lackland AFB, TX: Air Force Human Resources laboratory.
- Guion, R.M. (1980). On trinitarian doctrines of validity. *Professional Psychology* 11, 385-398.
- Henry, E. (1966). Conference on the use of biographical data in psychology. *American Psychologist* 21, 247-249.
- Henry, R.A. & C.L. Hulin. (1987). Stability of skilled performance across time: Some generalizations and limitations on utilities. *Journal of Applied Psychology* 72, 457-462.
- Henry, R.A. & C.L. Hulin. (1989). Changing validities: Ability-performance relations and utilities. *Journal of Applied Psychology* 74, 365-367.

- Holmes, D.S. (1977). How and why assessment works. In J.L. Moses & W.C. Byham (Eds.), *Applying the assessment center method* (pp. 127-143). New York: Pergamon Press.
- Howard, A. & D.W. Bray. (1988). *Managerial lives in transition: Advancing age and changing times*. New York: Guilford Press.
- Huck, J. & C.J. Russell. (1989, October). Assessment center validity in a continuing education setting. Presented at the International Institute of Research, London, England.
- Hulin, C.L., R.A. Henry, & S.L. Noon. (1990). Adding a dimension: Time as a factor in the generalizability of predictive validity. *Psychological Bulletin* 107, 328-340.
- Hunter, J.E. (1980). *Validity generalization for 12,000 jobs: An application of synthetic validity and validity generalization to the General Aptitude Test Battery (GATB)*. Washington, DC: U.S. Employment Service, U.S. Department of Labor.
- Hunter, J.E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior* 29, 340-362.
- Hunter, J.E. & R.F. Hunter. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin* 96, 72-98.
- Jaques, E. (1976). *A general theory of bureaucracy*. Portsmouth, NH: Heinemann Educational Books.
- Jaques, E. (1989). *Requisite organization: The CEO's guide to creative structure and leadership*. Arlington, VA: Cason Hall and Co.
- Jensen, A.R. (1986). g: Artifact or reality? *Journal of Vocational Behavior* 29, 301-331.
- Katz, R.L. (1974). Skills of an effective administrator. *Harvard Business Review* 52(1), 90-102.
- Kegan, R. (1982). *The evolving self: Problem and process in human development*. Cambridge, MA: Harvard University Press.
- Klimoski, R.J. & M. Brickner. (1987). Why do assessment centers work? The puzzle of assessment center validity. *Personnel Psychology* 40, 243-260.
- Klimoski, R.J. & W.J. Strickland. (1977). Assessment center—valid or merely prescient. *Personnel Psychology* 30, 353-361.
- Kornhauser, A.W. (1923). A statistical study of a group of specialized office workers. *Journal of Personnel Research* 2, 103-123.
- Kuhnert, K.E. & P. Lewis. (1987). Transactional and transformational leadership: A constructive/developmental analysis. *Academy of Management Review* 12, 648-657.
- Kuhnert, K.W. & C.J. Russell. (1990). Using constructive developmental theory and biodata to bridge the gap between personnel selection and leadership. *Journal of Management* 16, 1-13.
- Landy, F.J. (1986). Stamp collecting versus science: Validation as hypothesis testing. *American Psychologist* 41, 1183-1192.
- Lindsey, E., Z. Homes, & M. McCall. (1987). *Key events in executive lives*. Greensboro, NC: Center for Creative Leadership.
- MacKinnon, D.W. (1975). *An overview of assessment centers*. (CCL Tech Rep. 1). Berkeley, CA: University of California, Center for Creative Leadership.
- Mael, M.E. (1991). A conceptual rationale for the domain and attributes of biodata items. *Personnel Psychology* 44, 763-792.
- Mintzberg, H. (1973). *The nature of managerial work*. New York: Harper & Row.
- Moses, J.L. (1972). Assessment center performance and managerial progress. *Studies in Personnel Psychology* 4, 7-12.
- Moses, J.L. & V.R. Boehm. (1975). Relationship of assessment-center performance to management progress of women. *Journal of Applied Psychology* 60, 527-529.
- Mumford, M.D. & W.A. Owens. (1984). Individuality in a developmental context: Some empirical and theoretical considerations. *Human Development* 27, 84-108.

- Mumford, M.D. & W.A. Owens. (1987). Methodology review: Principles, procedures, and findings in the application of background data measures. *Applied Psychological Measurement* 11, 1-31.
- Mumford, M.D. & G.S. Stokes. (in press). Developmental determinants of individual action: Theory and practice in the application of background data. In M.D. Dunnette (Ed.), *The handbook of industrial and organizational psychology* (2nd edition). Orlando, FL: Consulting Psychologists Press.
- Mumford, M.D., G.S. Stokes, & W.A. Owens. (in press). *Handbook of biographical information*. Palo Alto, CA: Consulting Psychologists Press.
- Mumford, M.D., C.E. Uhlman, & R.N. Kilcullen. (in press). The structure of life history: Implications for the construct validity of background data scales. *Human Performance*.
- Neiner, A.G. & W.A. Owens. (1982). Relationships between two sets of biodata with 7 years separation. *Journal of Applied Psychology* 67, 145-150.
- Nickels, B.J. & M.D. Mumford. (1989). Exploring the structure of background data measures: Implications for item development and scale validation. Unpublished manuscript.
- Norton, S.D. (1977). The empirical and content validity of assessment centers vs. traditional methods for predicting managerial success. *Academy of Management Review* 2, 442-453.
- Office of Strategic Services Assessment Staff (1948). *Assessment of men: Selection of personnel for the Office of Strategic Services*. New York: Rinehart.
- Outclat, D. (1988). A research program on General Motors' foreman selection assessment center: Assessor/assesssee characteristics and moderator analysis. Presented at the 16th International Congress on the Assessment Center Method, Tampa, FL.
- Owens, W.A. (1968). Toward one discipline of scientific psychology. *American Psychologist* 23, 782-785.
- Owens, W.A. (1971). A quasi-actuarial prospect for individual assessment. *American Psychologist* 26, 992-999.
- Owens, W.A. & L.F. Schoenfeldt. (1979). Toward a classification of persons. *Journal of Applied Psychology* 64, 569-607.
- Peterson, N.G. & D.A. Bownas. (1982). Skill, task structure, and performance acquisition. In M.D. Dunnette & E.A. Fleishman (Eds.), *Human performance and productivity [vol. 1]: Human capability assessment*. Hillsdale, NJ: Erlbaum.
- Reilly, R.R. & G.T. Chao. (1982). Validity and fairness of some alternative employee selection procedures. *Personnel Psychology* 35, 1-63.
- Reilly, R.R. & M.W. Warech. (1990). *The validity and fairness of alternatives to cognitive tests*. Berkeley, CA: Commission on Testing and Public Policy.
- Reilly, R.R., S. Henry, & J.W. Smither. (1990). An examination of the effects of using behavior checklists on the construct validity of assessment center dimensions. *Personnel Psychology* 43, 71-84.
- Richardson, Bellows, Henry, & Co. (1984). *Technical reports: The managerial profile record*. Washington, DC: Author.
- Russell, C.J. (1983, July). An examination of assessors' clinical judgments. Presented at the 11th International Congress on the Assessment Center Method, Williamsburg, VA.
- Russell, C.J. (1985). Individual decision processes in an assessment center. *Journal of Applied Psychology* 70, 737-746.
- Russell, C.J. (1987). Person characteristic vs. role congruency explanations for assessment center ratings. *Academy of Management Journal* 30, 817-826.
- Russell, C.J. (1990). Selection top corporate leaders: An example of biographical information. *Journal of Management* 16, 71-84.

- Russell, C.J. (in press). Biodata item generation procedures: A point of departure for the new frontier. In G.S. Stokes, M.D. Mumford, & W.A. Owens (Eds.), *Handbook of biographical information*. Orlando, FL: Consulting Psychologists Press.
- Russell, C.J. & D.R. Domm. (1990a). On the construct validity of biographical information: Evaluation of a theory based method of item generation. Presented at the 5th annual meetings of the Society of Industrial and Organizational Psychology, Miami Beach.
- Russell, C.J. & D.R. Domm. (1990b). On the validity of role congruency-based assessment center procedures for predicting performance appraisal ratings, sales revenue, and profit. Presented at the 50th annual meetings of the Academy of Management, San Francisco, CA.
- Russell, C.J. & Domm, D.R. (1991). Simultaneous ratings of role congruency and person characteristics, skill, and abilities in an assessment center: A field test of competing explanations. Presented at the 6th meeting of the Society of Industrial and Organizational Psychology, St. Louis, MO.
- Russell, C.J., A. Colella, & P. Bobko. (1991). Expanding the dimensions of utility: The financial and strategic impact of personnel selection. Unpublished manuscript.
- Russell, C.J., J. Mattson, S.E. Devlin, & D. Atwater. (1990). Predictive validity of biodata items generated from retrospective life experience essays. *Journal of Applied Psychology* 75, 569-580.
- Ryanen, I.A. (1988). Commentary of a minor bureaucrat. *Journal of Vocational Behavior* 33, 379-387.
- Sackett, P.R. (1982). A critical look at some common beliefs about assessment centers. *Public Personnel Management* 11, 140-146.
- Sackett, P.R. (1987). Assessment centers and content validity. *Personnel Psychology* 40, 13-25.
- Sackett, P.R. & G.F. Dreher. (1981). Some misconceptions about content oriented validation: A rejoinder to Norton. *Academy of Management Review* 6, 567-568.
- Sackett, P.R. & G.F. Dreher. (1982). Constructs and assessment center dimensions: Some troubling findings. *Journal of Applied Psychology* 67, 401-410.
- Sackett, P.R. & G.F. Dreher. (1984). Situation specificity of behavior and assessment center validation strategies: A rejoinder to Neidig and Neidig. *Journal of Applied Psychology* 69, 187-190.
- Sackett, P.R. & M.D. Hakel. (1979). Temporal stability and individual differences in using assessment center information to form overall ratings. *Organizational Behavior and Human Performance* 23, 120-137.
- Sackett, P.R. & M.M. Harris. (1988). A further examination of the constructs underlying assessment center ratings. *Journal of Business and Psychology* 3, 214-229.
- Saunders, D.R. (1956). Moderator variables in prediction. *Educational and Psychological Measurement* 16, 209-222.
- Schippman, J.S., G.L. Hughes, & E.P., Prien. (1987). The use of structural multi-domain job analysis for the construction of assessment center methods and procedures. *Journal of Business and Psychology* 1, 353-366.
- Schmidt, F.L. & J.E. Hunter. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology* 62, 529-540.
- Schmitt, N. & T.E. Hill. (1977). Race and sex composition of assessment center groups as determinants of peer and assessor ratings. *Journal of Applied Psychology* 62, 261-264.
- Schmitt, N., R.Z. Gooding, R.A. Noe, & M. Kirsch. (1984). Meta-analysis of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology* 37, 407-422.
- Schneider, B. (1987). The people make the place. *Personnel Psychology* 40, 437-454.
- Schneider, J.R. & N. Schmitt. (in press). An exercise design approach to understanding assessment center dimension and exercise constructs. *Journal of Applied Psychology*.

- Schoenfeldt, L.F. (1974). Utilization of manpower: Development and evaluation of an assessment-classification model for matching individuals with jobs. *Journal of Applied Psychology* 59, 583-595.
- Shaffer, G.S., V. Saunder & W.A. Owens. (1986). Additional evidence for the accuracy of biographical information: Long-term retest and observer ratings. *Personnel Psychology* 39, 791-809.
- Shore, T.H., G.C. Thornton, III, & L.M. Shore. (1990). Construct validity of two categories of assessment center dimension ratings. *Personnel Psychology* 43, 101-116.
- Simon, H.A. (1977). The new science of management decision making: Further thoughts after the Bradford studies. *Journal of Management Science* 26, 29-46.
- Smith, P.C. (1976). The problem of criteria. In M.D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 745-775). Chicago: Rand McNally.
- Snyderman, M. & S. Rothman. (1986, Spring). Science, politics and the IQ controversy. *The Public Interest* 83, 79-98.
- Spearman, C.E. (1927). *The abilities of man*. New York: Macmillan.
- Standard Oil Company of New Jersey (1962). *Social science research reports: Selection and placement* (vol. 11). New York: Author.
- Standards for Ethical Considerations for Assessment Center Operations. (1977). In J.L. Moses & W.C. Byham (Eds.), *Applying the assessment center method* (pp. 303-310). New York: Pergamon Press.
- Stokes, G.S., M.D. Mumford, & W.A. Owens. (1989). Life history prototypes in the study of human individuality. *Journal of Personality* 57, 509-545.
- Thorndike, E.L., W. Lay, & P.R. Dean. (1909). The relation of accuracy in sensory discrimination to general intelligence. *American Journal of Psychology* 20, 364-369.
- Thorndike, R.L. (1986). The role of general ability in prediction. *Journal of Vocational Behavior* 29, 332-339.
- Thornton, G.C., III & W.C. Byham. (1982). *Assessment centers and managerial performance*. New York: Academic Press.
- Torbert, W. (1987). *Managing the corporate dream*. Homewood, IL: Dow Jones-Irwin.
- Turnage, J.J. & P.M. Muchinsky. (1982). Transituational variability in human performance within assessment centers. *Organizational Behavior and Human Performance* 30, 174-200.
- Uniform guidelines on employee selection procedures*. Federal Register, Friday, August 25, 43(166), 38294-38309.
- Vance, R.J., R.C. MacCallum, M.D. Coovert, & J.W. Hedge. (1988). Construct validity of multiple job performance measures using confirmatory factor analysis. *Journal of Applied Psychology* 73, 74-80.
- Welford, A.T. (1968). *Fundamentals of skill*. London: Methuen.
- Zalesnik, A. (1977). Managers and leaders: Are they different? *Harvard Business Review* 55(5), 67-80.
- Zalesnik, A. (1990). The leadership gap. *Academy of Management Executive* 4, 7-22.