Introduction

Skepticism regarding the usefulness of personality measurement for personnel selection reached a peak during the late 1960's. Two critiques were particularly influential. The first was Mischel's (1968) Book which claimed that (1) there is no evidence that personality is consistent over time and that (2) measures of personality explain only a trivial amount of the variance in social conduct. The second was a review by Guion and Gottier (1965) that concluded there was no evidence for the validity of personality tests. On closer analysis and with additional research, these conclusions have been largely reversed.

The renewed interest in personality assessment is based on both qualitative (Goldberg, 1992; R. Hogan, 1991; Schmidt & Ones, 1992) and quantitative (Barrick & Mount, 1991; Tett, Jackson, & Rothstein, 1991) reviews which conclude that personality measures, when organized in terms of the FFM, are consistently related to relevant job performance criteria. Several developments have helped reverse professional opinion on this subject. First, because measures of normal personality, rather than measures of psychopathology, were used in the validity analyses, validities were enhanced. Second, the FFM provides an taxonomic structure for organizing personality scales and when this structure is used to select measures in validation research, correlations with job-related criteria improve. Third, when latent performance constructs are used to align predictors and criteria, the validities associated with personality measures approximate the validities associated with cognitive tests. Finally, unlike cognitive tests, personality measures have no adverse or differential impact on protected groups.

Although other personality inventories based on the FFM have been recently developed (cf. Costa & McCrae, 1985, 1989; Goldberg, 1992; Wiggins, 1991), the HPI is uniquely designed to forecast performance in real world settings. It has been used in over one hundred research studies to predict job performance. This research shows that several scales are consistently related to the performance requirements that are common to many jobs. Three of the six occupational scales assess broad dimensions of organizational effectiveness – Service Orientation, Reliability, and Stress Tolerance. The remaining three scales predict potential success in clerical, sales, and management jobs. These six scales were developed and validated for use in personnel selection.

The occupational scales are composed of HICs from the personality scales. HICs were combined empirically to predict non-test criteria in organizations. Because the occupational scales use the same HICs as the personality scales, they share the same psychometric properties. Therefore, the HIC descriptions presented in Table 2.2 and the reliabilities presented in Table 2.5 apply to the occupational scales. Note, however, the personality scales were constructed through factor analysis of the 41 HICs and that procedure optimizes the internal consistency reliability of the scales. When we recombine HICs to form new occupational scales, their statistical independence reduces the internal consistency of the new scale. The internal consistency of the six occupational scales ranges from .69 to .86 (N=7,583). In our judgment, however, test-retest reliabilities are the estimates that matter in practical application and these range from .77 to .87 (N=150). The following section summarizes the development of each occupational scale.

Description of the Occupational Scales

Service Orientation Scale

Purpose: To identify persons who are pleasant, courteous, cooperative, and helpful in dealing with customers, clients, and co-workers.

Technical Description: The scale consists of 14 items from three HICs which appear on the Adjustment, Likeability, and Prudence scales. Sample items include:

I do my job as well as I possibly can.

I am a relaxed easy going person.

I always try to see the other person's point of view.

Total group scale means and standard deviations appear in Table 6.1. The internal consistency reliability (Cronbach, 1951) of this scale is .69. This moderate internal consistency reflects the fact that Service Orientation is composed of HICs taken from three independent personality scales (Table 2.1). The reliability of the Service Orientation scale, estimated by test-retest reliability over a four-week period is .78.

Normative data are available by total group in Appendix A and by gender, race, and age in Appendix B. To comply with provisions of the Civil Rights Act of 1990, we recommend that total group norms presented in Appendix A be used to convert raw scale scores into percentiles when making personnel decisions.

Developmental Validation Research. The initial concurrent validation research on the Service Orientation scale appears in R. Hogan, J. Hogan, and Busch (1984). A sample of 101 nursing aides at a large Baltimore hospital completed the HPI and were rated by their supervisors' for service orientation on the job. Reliability of the supervisors' ratings was .51. HPI HICs were correlated with supervisors' ratings and HICs from the Adjustment, Sociability, and Likeability scales were significantly related to the criterion measure. These HICs were combined into a single scale, which correlated .61 ($p<.001$) with supervisors' ratings. The current Service Orientation scale is a shortened version of the original 87-item scale. The correlation between the original scale and the revised 14-item scale is .71.

Table 6.1 ***Descriptive Statistics and Reliabilities for HPI Occupational Scales***

Scale	Number of Items	Mean	SD	Alpha	Inter-Item r	r^{tt}	SE
Service Orientation	14	10.36	2.52	.69	.13	.78	1.18
Stress Tolerance	25	19.48	4.53	.86	.20	.87	1.65
Reliability	18	12.35	3.42	.75	.14	.83	1.43
Clerical Potential	24	18.91	3.41	.72	.09	.77	1.64
Sales Potential	67	40.77	8.88	.82	.07	.85	3.43
Managerial Potential	37	30.68	4.63	.79	.10	.83	1.93

Table 6.2 lists additional validation studies using Service Orientation in the predictor battery. This table includes the reference sources, samples studied, criteria used, and validity coefficients for Service Orientation.

Table 6.3 contains adjectival descriptors that correlate with scores on the Service Orientation scale. These data were generated from the peer description validation research presented in Chapter Three (page 42). In this study, subjects (N=128) completed the HPI and two peers who knew each subject well rated the subject using 112 adjectives from Gough and Heilbrun's (1983) Adjective Check List (ACL). As seen in Table 6.3, service oriented persons are described as calm, praising, and soft-hearted and they are not seen as worrying, fault-finding, or bossy.

Stress Tolerance Scale

Purpose: To identify persons who handle pressure well and are not tense or anxious.

Technical Description: The scale consists of 25 items from five HICs which appear on the Adjustment and Ambition scales. Sample items include:

I keep calm in a crisis.

I am seldom tense or anxious.

I am a happy person.

Total group scale means and standard deviations appear in Table 6.1. The internal consistency reliability (Cronbach, 1951) of this scale is .86. This is the highest coefficient alpha among the occupational scales and it reflects the large contribution of Adjustment scale HICs. The reliability of the Stress Tolerance scale, estimated by test-retest reliability over a four-week period, is .87.

Normative data are available by total group in Appendix A and by gender, race, and age in Appendix B. To comply with provisions of the Civil Rights Act of 1990, we recommend that total group norms presented in Appendix A be used to convert raw scale scores into percentiles when making personnel decisions.

Table 6.2 ***Additional Validation Studies for Service Orientation, Stress Tolerance, and Reliability***

Source	Sample	Criteria	*r*
Service Orientation			
Hogan, Hogan, & Busch (1984)	37 nursing students	Faculty ratings	.31*
	30 nursing home aides	Supervisor ratings	.42**
Raza, Metz, Dyer, Coan, & Hogan (1986)	201 hospital service workers	Times counseled for aberrant behavior	-.15*
Montgomery, Butler, & McPhail (1987)	122 nuclear power plant workers	Supervisor ratings of punctuality	.23**
Hogan, Jacobson, Hogan, & Thompson (1987)	76 service operations dispatchers	Number of absences above allowable	-.40**
Muchinsky (1987)	102 customer service representatives	Supervisor ratings of quantity	.21*
		Supervisor ratings of quality	.28*
Muchinsky (1987)	145 telemarketers	Sales performance	.31*
		Lead generation	.30*
		Fund raising	.24*
Muchinsky (1987)	44 field sales representatives	Supervisor ratings of overall performance	.27*
Muchinsky (1987)	50 office managers	Supervisor ratings of follow-through	.33*
		Supervisor ratings of adaptability	.26*
		Supervisor ratings of overall performance	.29*
		Supervisor ratings of planning	.29*
Cage (1989)	20 nannies	Supervisor ratings of appearance	.43*
		Supervisor ratings of common sense	.37*
		Supervisor ratings of overall effectiveness	.38*
Hogan & Arneson (1991)	178 habilitation therapists	Disciplinary notations	-.30**
Curphy, Gibson, Asiu, Horn, & Macomber	312 military attaches	Supervisor ratings of performance	.32*
Muchinsky (1993)	109 insurance employees	Supervisor ratings of interpersonal skills	.25*
		Supervisor ratings of organization	.26*
Hogan & Hogan (1993)	163 truck drivers	Accident rates	-.48**
Hayes, Roehm, & Castellano (1994)	130 machine operators	Supervisor ratings of effectiveness	.61*
Hogan, Hogan, & Brinkmeyer (1994)	255 truck drivers	Supervisor ratings of overall performance	.40**
Landy, Jacobs, & Associates (1994)	588 bus operators	Supervisor ratings of courtesy	.14*
Hogan & Gerhold (1995a)	27 financial consultants	Number of new accounts (annually)	.43*
Klippel (1995)	29 service operations coordinators	Supervisor ratings of social skill	.37*
Gerhold (1995)	23 financial consultants	Annual commission	.51**
Stress Tolerance			
Guier (1984)	65 psychiatric counselors	Supervisor ratings: overall job performance	.25*
Hogan, Peterson, Hogan, & Jones	110 line-haul truck drivers	Number of commendations received	.17*
		Number of claims filed for equipment failure	-.25**
Montgomery, Butler, & McPhail (1987)	122 nuclear power plant workers	Supervisor ratings of dependability	.22**
		Supervisor attitude ratings	.18*
		Supervisor ratings of punctuality	.30**

Source	Sample	Criteria	*r*
Hogan, Jacobson, Hogan, & Thompson (1987)	76 service operations dispatchers	Number of absences above allowable	-.27**
Muchinsky (1987)	102 customer service representatives	Supervisor ratings of quantity	.24*
		Supervisor ratings of quality	.30*
		Supervisor ratings of teamwork	.35*
		Supervisor ratings of overall performance	.31*
Muchinsky (1987)	145 telemarketers	Sales performance	.29*
		Fund raising	.33*
Muchinsky (1987)	44 field sales representatives	Supervisor ratings of overall performance	.25*
Muchinsky (1987)	50 office managers	Supervisor ratings of follow-through	.37*
		Supervisor ratings of adaptability	.39*
		Supervisor ratings of overall performance	.34*
		Supervisor ratings of planning	.34*
		Supervisor ratings of judgment	.31*
Salas, Hogan, Driskell, & Hoskins (1988)	135 Navy electronic students	Academic performance	.19*
		Training completion	.19*
Cage (1989)	20 nannies	Supervisor ratings of appearance	.43*
		Supervisor ratings of common sense	.39*
		Supervisor ratings of overall effectiveness	.42*
Hogan & Hogan (1993)	163 truck drivers	Low accident rates	.41**
Brinkmeyer (1994)	49 job applicants	Recruiter ratings of interpersonal performance	.29*
Hogan & Gerhold (1994)	24 certified nurses aides	Supervisor ratings of overall performance	.50*
Hayes, Roehm, & Castellano (1994)	130 machine operators	Supervisor ratings of effectiveness	.71*
Hogan, Brinkmeyer, & Kidwell (1994)	30 tank truck drivers	Supervisor ratings of overall performance	.30*
Gerhold (1995)	23 financial consultants	Annual commission	.52**
Reliability			
Hogan, Hogan, & Briggs	56 truck drivers	Number of commendations received	0.51**
Hogan, Peterson, Hogan, & Jones (1982)	110 line-haul truck drivers	Number of claims filed for equipment failure	-.25**
		Number of commendations received	.25**
Raza, Metz, Dyer, Coan, & Hogan (1986)	201 hospital service workers	Times counseled for aberrant behavior	-.18*
Hogan, Arneson, Hogan, & Jones (1986)	178 habilitation therapists	Number of injuries sustained	-.17*
		Hours lost to job injury	-.17*
Montgomery, Butler, & McPhail (1987)	122 nuclear power plant workers	Supervisor ratings of dependability	.19*
		Supervisor ratings of punctuality	.23**
		Supervisor ratings of accuracy	.16*
Hogan, Jacobson, Hogan, & Thompson (1987)	76 service operations dispatchers	Number of absences above allowable	-.49**
Salas, Hogan, Driskell, & Hoskins (1988)	135 Navy electronic students	Course completion time	.16*

Source	Sample	Criteria	*r*
Muchinsky (1987)	102 customer service representatives	Supervisor ratings of quantity	.18*
		Supervisor ratings of teamwork	.24*
		Supervisor ratings of overall performance	.19*
Muchinsky (1987)	145 telemarketers	Sales performance	.26*
		Fund raising	.30*
		Lead generation	.34*
Muchinsky (1987)	44 field sales representatives	Supervisor ratings of overall performance	.29*
Muchinsky (1987)	50 office managers	Supervisor ratings of follow-through	.38*
		Supervisor ratings of adaptability	.35*
		Supervisor ratings of overall performance	.25*
Woolley (1990)	131 male students	Counterproductivity at work	-.36*
		Counterproductivity at school	-.22*
	158 female students	Counterproductivity at work	-.33*
		Counterproductivity at school	-.38*
Hogan & Hogan (1993)	163 truck drivers	Low accident rates	.48**
Brinkmeyer (1994)	49 job applicants	Recruiter ratings of social skills	.24*
Hogan & Gerhold (1994)	24 certified nurses aides	Supervisor ratings of overall performance	.52*
Hayes, Roehm, & Castellano (1994)	130 machine operators	Supervisor ratings of effectiveness	.65**
Hogan, Brinkmeyer, & Kidwell (1994)	30 tank truck drivers	Supervisor ratings of overall performance	.61*
Nolan, Johnson, & Pincus (1994)	320 adults	Rated severity of alcoholism	-.62**
Hogan, Hogan, & Brinkmeyer (1994)	255 truck drivers	Supervisor ratings of overall performance	.43**
Landy, Jacobs, & Associates (1994)	588 bus operators	Supervisor ratings of safety	.15*
Hogan & Gerhold (1995b)	90 convenience store managers	Supervisor ratings of overall performance	.27**
Klippel (1995)	20 service operations coordinators	Work attendance Supervisor ratings of professional courtesy	.51* .47*

Table 6.3 ***Adjectival Correlates of the HPI Occupational Scales***[a]

General Occupational Scales					
Service Orientation		**Stress Tolerance**		**Reliability**	
Moody	-.38	Tense	-.47	Self-pitying	-.38
Tense	-.37	Moody	-.42	Stable	.32
Worrying	-.37	Unstable	-.42	Noisy	-.32
Anxious	-.34	Worrying	-.41	Fault-finding	-.31
Fault-finding	-.33	Self-Pitying	-.38	Affectionate	.30
Calm	.32	Nervous	-.35	Curious	-.30
Praising	.31	High-strung	-.35	Irresponsible	-.29
Bossy	-.31	Fearful	-.34	Soft-hearted	.28
Soft-hearted	.30	Temperamental	-.33	Praising	.28
Self-pitying	-.29	Self-punishing	-.32	Unstable	-.27
Specific Occupational Scales					
Clerical		**Sales**		**Manager**	
Active	.39	Talkative	.45	Active	.32
Tense	-.31	Shy	-.40	Polished	.31
Nervous	-.30	Outgoing	.39	Precise	.30
Energetic	.30	Quiet	-.35	Ingenious	.30
Outgoing	.29	Assertive	.33	Wise	.29
Moody	-.27	Reserved	-.33	Planful	.27
Cold	-.27	Spunky	.33	Outgoing	.26
Fearful	-.27	Retiring	-.32	Efficient	.25
Unstable	-.26	Outspoken	.32	Irresponsible	-.25
Polished	.25	Noisy	.31	Unstable	-.25

[a]These correlations are based on data from the peer description validation research presented in tables 3.10, 3.11, and 3.12. All correlations are $p < .01$.

N = 128

Developmental Validation Research. The initial validation research on the Stress Tolerance scale appears in J. Hogan, R. Hogan, and Briggs (1984). A sample of 56 truck drivers employed by a large east coast motor freight company completed the HPI. The employee service file of each driver was reviewed and job performance data from the previous eighteen months were coded and used as criterion measures. The Stress Tolerance scale, composed of HICs from the Adjustment, Prudence, and Ambition scales, correlated with number of days a driver missed work for medical reasons ($r=-.51$, $p<.001$) and number of commendation letters received ($r=.42$, $p=.002$).

The current Stress Tolerance scale is a shortened version of the original 55-item scale. The correlation between the original scale and the revised 25-item scale is .85. McDonald, Beckett, and Hogdon (1988) extended the construct validity of this scale in a study of fitness and performance in male and female personnel on active duty in the U.S. Navy (N=102). Subjects completed the HPI and the Profile of Mood States (POMS; McNair, Lorr, & Droppleman, 1971) and the Tennessee Self Concept Scale (TSCS; Fitts,

1965). The Stress Tolerance scale correlated -.24, -.28, .25, and -.30 respectively with the POMS scales of Tension, Depression, Vigor, and Confusion ($p<.05$). In addition, the Stress Tolerance scale correlated .33 and .63 with the TSCS subscales of physical self-concept and total self-concept ($p<.01$).

Table 6.2 lists additional validation studies using Stress Tolerance. This table includes the reference sources, samples studied, criteria used, and validity coefficients for Stress Tolerance.

Table 6.3 contains adjectival correlates of scores on the Stress Tolerance scale. As seen, persons with low scores on Stress Tolerance are described as tense, moody, and unstable. Persons who score high on the Stress Tolerance scale are seen as stable and even-tempered.

Reliability Scale

Purpose: To identify persons who are honest, dependable, and responsive to supervision.

Technical Description: The scale consists of 18 items from four HICs which appear on the Adjustment, Likeability, and Prudence scales. Sample items include:

> I rarely do things on impulse.
>
> When I was in school, I rarely gave the teachers any trouble.
>
> I am rarely irritated by faults in others.

Total group scale means and standard deviations appear in Table 6.1. The internal consistency reliability (Cronbach, 1951) of this scale is .75. The reliability of the Reliability scale, estimated by test-retest reliability over a four-week period, is .83.

Normative data are available by total group in Appendix A and by gender, race, and age in Appendix B. To comply with provisions of the Civil Rights Act of 1990, we recommend that total group norms presented in Appendix A be used to convert raw scale scores into percentiles when making personnel decisions.

Developmental Validation Research. The initial validation research on the Reliability scale appears in J. Hogan and R. Hogan (1989). Four groups of subjects completed the HPI: (a) 40 incarcerated delinquent persons from Arizona who had been arrested for drug-related offenses; (b) 103 working-class nurse aides from an inner-city hospital; (c) 38 police cadets from a suburban department; and (d) 145 college students. The delinquents were assigned a score of 0, the non-delinquents were given a score of 1. In addition, the college students completed a questionnaire asking about their past delinquent behavior (e.g., drug use, promiscuity, arrests); the questionnaire was scored by summing the number of nondelinquent items endorsed. Nineteen HICs were correlated with the delinquency criteria, 9 of these were chosen on the basis of their correlations with the delinquency criteria and their conceptual correspondence to the structure of deviancy. Although each component HIC correlated with the delinquency criteria, the 'Avoids Trouble' HIC had the highest correlation, with r's ranging from .68 to .73 (N=143). The current Reliability scale is a shortened version of the original 69-item scale. The correlation between the original scale and the revised 18-item scale is .72.

Interest in integrity testing has increased recently due to the expense of counterproductive behaviors in organizations and the national trend toward cost-cutting. Sackett, Burris, and Callahan (1989) distinguish overt integrity tests from personality-based tests, with the former containing transparent items (e.g., "I stole more than $5,000 from my last employer.") and the latter containing genotypic responsibility and conscientiousness items (e.g., "People respect my integrity."). The HPI Reliability scale is a personality-based measure of integrity.

Woolley and Hakstian (1992) examined the construct validity of the two types of integrity measures using the HPI Reliability scale, the Personnel Reaction Blank (PRB; Gough, 1972), the Personnel Decisions, Inc., Employment Inventory (PDI-EI; Personnel Decisions Inc., 1985), and the Reid Report (Reid, 1967)—all except the Reid Report are personality-based integrity measures. The correlations among the various scales indicated that the personality-based measures substantially converged. The correlations between the personality-based measures and the Reid Report scales, which are overt integrity measures, were consistently lower than correlations between the personality-based measures. The Reid Report Punitive scale was unrelated to the personality-based measures, suggesting that attitudes toward punishing others are unrelated to responsibility and conscientiousness. Factor analysis of Reliability, PRB, PDI-EI, Reid Report, CPI Class II scales, selected 16 PF scales, and NEO-PI scales yielded four factors. The largest of the four was "Socialized Control," a dimension the authors concluded was the core of what Sackett et al. (1989) referred to as integrity. The Reliability scale loaded .75 on this factor which also included loadings of .76, .48 and .08 for the PRB, PDI-EI, and Reid Report scales, respectively. Woolley and Hakstian conclude that personality-based and overt measures of integrity are factorially and conceptually distinct.

The Reliability scale has been used in a number of meta-analytic reviews – cf. Barrick and Mount (1991), Ones (1992), and Tett et al. (1992). Ones (1992) estimated that the average criterion-related validity of the Reliability scale when predicting counterproductive behaviors is .45, a figure that is somewhat higher than those reported for Conscientiousness measures in general. In their meta-analysis of personality-based integrity measures, Barrick and Mount (1991) conclude, "…the results suggest that if the purpose is to predict job performance based on an individual's personality, then those measures associated with Conscientiousness (Factor III in the five-factor model) are most likely to be valid predictors for all jobs. In fact, it is difficult to conceive of a job in which the traits associated with the Conscientiousness dimension would not contribute to job success" (pp. 21-22).

Table 6.2 lists several studies in which the Reliability scale was used to predict job performance. This table includes the reference sources, samples studied, criteria used, and validity coefficients for Reliability.

Table 6.3 contains adjectival correlates of the Reliability scale. As can be seen, persons with high scores on the Reliability scale are described as stable, affectionate, and praising, while persons with low scores are described as self-pitying, fault-finding, and irresponsible.

Clerical Potential Scale

Purpose: To identify persons who are attentive to detail, congenial, and industrious.

Technical Description: The scale consists of 24 items from five HICs which appear on the Adjustment, Ambition, Likeability, and Prudence scales. Sample items include:

> In a group, I like to take charge of things.
>
> When I was in school, I rarely gave the teachers any trouble.
>
> I am sensitive to other people's moods.

Total group scale means and standard deviations appear in Table 6.1. The internal consistency reliability (Cronbach, 1951) of this scale is .72. The reliability of the Clerical Potential scale, estimated by test-retest reliability over a four-week period is .77.

Normative data are available by total group in Appendix A and by gender, race, and age in Appendix B. To comply with provisions of the Civil Rights Act of 1990, we recommend that total group norms presented in Appendix A be used to convert raw scale scores into percentiles when making personnel decisions.

Developmental Validation Research. The initial concurrent validation research on the Clerical Potential scale appears in J. Hogan and R. Hogan (1986). Women (N=107) in 44 different clerical positions in a large insurance company completed the HPI and were rated for overall job performance by two supervisors. These two sets of ratings correlated .60. Five HICs had the following correlations with the supervisors' ratings: Not Anxious (.26), No Somatic Complaint (.25), Avoids Trouble (.26), Leadership (.12), and Caring (.22). These HICs were combined to form a single scale that correlated .47 with rated performance in one half of the sample, .37 in the second half, and .42 overall (correlations corrected for attenuation).

If an organization uses formal, structured assessments of potential clerical employees, these evaluations will typically be performance tests and paper-and-pencil measures of basic cognitive abilities. To determine whether the Clerical Potential scale adds to the prediction of job performance beyond the variance accounted for by more traditional performance assessments, we correlated the Clerical Potential scale with four subtests of the PSI Basic Skills Tests (shown in Table 3.2) and found no significant relationships. This indicates that the Clerical Potential scale is independent of the cognitive and skill-based assessments typically used to select clerical employees. Rosse, Miller and Barnes (1991) used the Clerical Potential scale to study hospital clerks whose jobs involve patient contact. Clerical Potential correlated .25 and .27 ($p<.05$, corrected for criterion unreliability and restriction in range of predictors) with supervisors' ratings of service and overall job performance.

Table 6.3 contains adjectival correlates of scores on the Clerical Potential scale. As seen, persons with high scores on Clerical Potential are described as active, outgoing, and polished, while persons with low scores are described as nervous, moody, and cold.

Sales Potential Scale

Purpose: To identify persons who are socially skilled, self-assured, assertive, and can create interest in products and services.

Technical Description: The scale consists of 67 items from 14 HICs which appear on the Adjustment, Ambition, Likeability, Sociability, Prudence, and Intellectance scales. Sample items include:

I am a quick-witted person.

I would go to a party every night if I could.

I work well with other people.

Total group scale means and standard deviations appear in Table 6.1. The internal consistency reliability (Cronbach, 1951) of this scale is .82. The reliability of the Sales Potential scale, estimated by test-retest reliability over a four-week period, is .85.

Normative data are available by total group in Appendix A and by gender, race, and age in Appendix B. To comply with the provisions of the Civil Rights Act of 1990, we recommend that total group norms presented in Appendix A be used to convert raw scale scores into percentiles when making personnel decisions.

Developmental Validation Research. The initial validation research on the Sales Potential scale was conducted by Merrill (1992). A sample of executives and top sales representatives (N=67) from a large international trade magazine and publishing company completed the HPI. The sales reprentatives were the top advertising revenue producers in the company, all had at least five years tenure and their annual accounts revenues exceeded one million dollars. Merrill coded sales representatives as “2” and other executives as “1” and correlated the HPI with this dummy coded criterion. Fourteen HICs were correlated with sales criterion and the zero order correlations ranged from .24 to .41. These 14 HICs, mostly from the Ambition, Sociability, and Prudence (scored negatively) scales, were combined into a 67-item scale and the total score correlated .65 ($p<.001$) with the sales criterion.

Table 6.3 contains the adjectival correlates of scores on the Sales Potential scale. As seen, persons with high scores on Sales Potential are described as talkative, outgoing, and assertive, while persons with low scores are described as shy, quiet, and reserved.

Managerial Potential Scale

Purpose: To identify persons who can supervise others in a pleasant and effective fashion.

Technical Description: The scale consists of 37 items from eight HICs which appear on the Adjustment, Ambition, Prudence, and School Success scales. Sample items include:

I strive for perfection in everything I do.

I know what I want to be.

In a group, I like to take charge of things.

Total group scale means and standard deviations appear in Table 6.1. The internal consistency reliability (Cronbach, 1951) of this scale is .79. The reliability of Managerial Potential, estimated by test-retest reliability over a four-week period is .83.

Normative data are available by total group in Appendix A and by gender, race, and age in Appendix B. To comply with provisions of the Civil Rights Act of 1990, we recommend that total group norms presented in Appendix A be used to convert raw scale scores into percentiles when making personnel decisions.

Developmental Validation Research. The initial validation research for the Managerial Potential scale was conducted at a large east coast motor freight company (J. Hogan, R. Hogan, & Murtha, 1993). This company has a policy of promoting from within so that drivers become coordinators (first-line supervisors who coordinate work on the loading docks), who may become terminal managers, who then may become regional managers. Employees (N=372) completed the HPI, including 56 drivers, 219 coordinators, 83 terminal managers, and 14 regional managers. Drivers were assigned a score of 1; coordinators a score of 2; terminal managers, 3; and regional managers, 4. This variable was called Organizational Status, and it was assumed to reflect overall managerial potential. In addition, managers were rated for performance by their supervisors. Correlations were computed between HPI scores and Organizational Status for all cases (N=372). Eleven HICs, primarily from the Adjustment and Ambition scales, were combined to form a single scale which correlated .66 ($p<.001$) with the Organizational Status criterion and .25 ($p<.01$) with the overall supervisors' ratings.

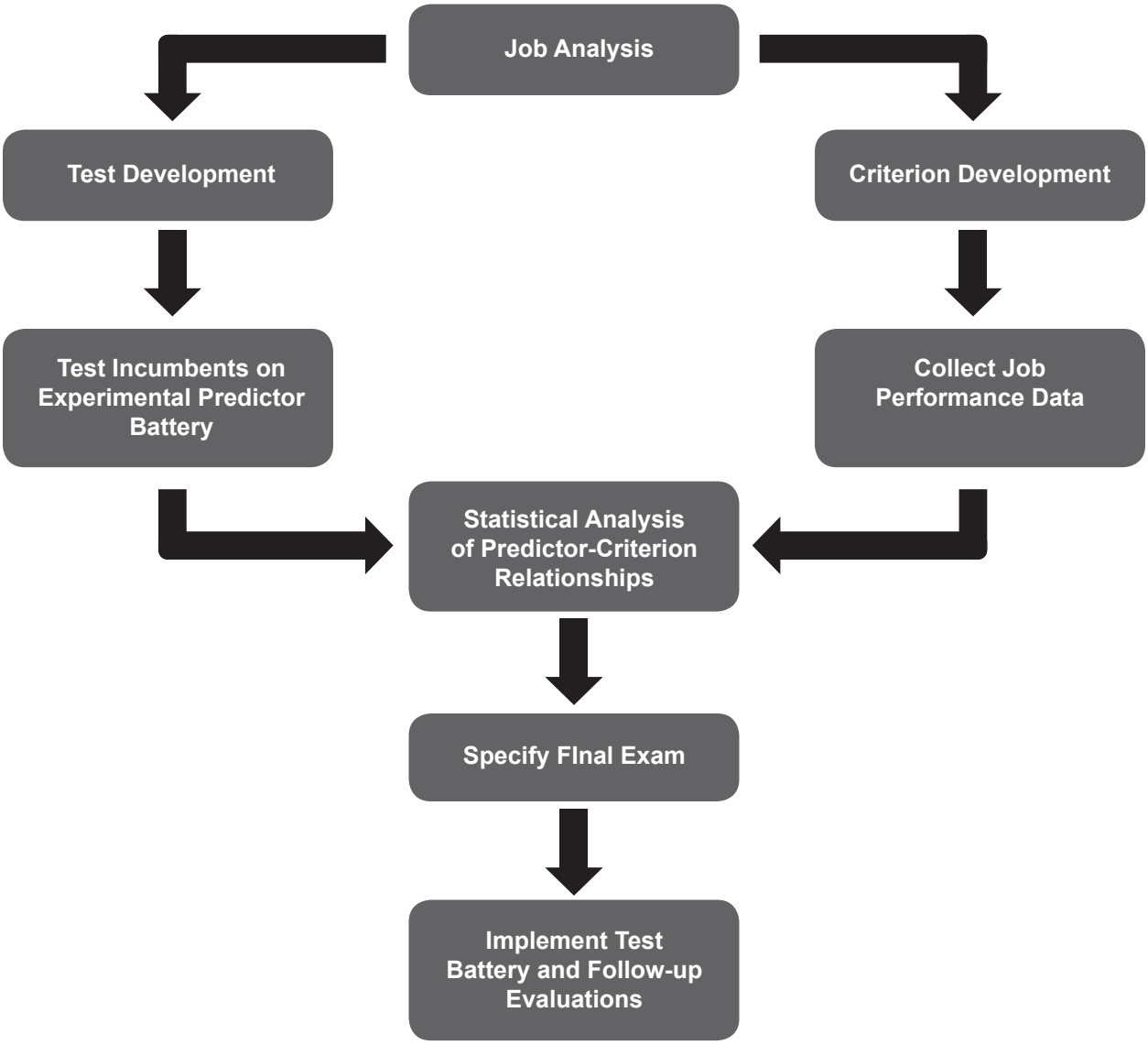
It is often suggested that the characteristics associated with managerial potential are situation or industry specific. Data using the Managerial Potential scale suggest that the characteristics assessed by this measure predict managerial performance across settings. The Managerial Potential scale is associated with: (1) city police officers' performance ($r=.39$, $p<.01$) (Quigley, Millikin-Davies, Francis, & J. Hogan, 1989); (2) superintendents' ratings ($r=.28$, $p<.01$) of school principals (J. Hogan & Zenke, 1986); (3) supervisors' composite evaluations ($r=.39$, $p<.01$) of long-distance telephone company managers (Muchinsky, 1987); (4) supervisors' ratings ($r=.26$, $p<.01$) of insurance underwriters and claims analysts with managerial responsibilities (Muchinsky, 1993); and (5) supervisors' performance ratings ($r=.22$, $p<.01$) of convenience store managers (R. Hogan, J. Hogan, Lock, & Brinkmeyer, 1994).

Table 6.3 contains adjectival correlates of the Managerial Potential scale. As seen, persons with high scores on Managerial Potential are described as polished, ingenious, wise, and planful, while persons with low scores are described as irresponsible and unstable.

How to Conduct Validation Research Using the HPI

This section offers some recommendations that might be useful for persons who are unfamiliar with personality constructs and their measures but who want to use the HPI in test validation research. With personality variables in mind, we will discuss (1) job analysis; (2) test specification; (3) criterion measures; (4) statistical analyses; and (5) implementation. Figure 6.1 illustrates the sequence and relation of these steps. Our discussion assumes the use of an empirical validation strategy because content validity is inappropriate

Figure 6.1 ***Concurrent Validation Strategy***



for demonstrating job-relatedness of a personality measure (Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, and Department of Justice, 1978, p. 38302). In addition, recommendations are consistent with the Uniform Guidelines on Employee Selection Procedures (Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, and Department of Justice, 1978), Principles for the Validation and Use of Employment Testing Procedures (Society for Industrial and Organizational Psychology, Inc., 1987) and Standards for Educational and Psychological Testing, (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1985).

Job Analysis

Themes involving personality constructs inevitably emerge during job analysis, however, most structured job analysis procedures will not capture them (Guion, 1992). Incumbents and supervisors typically describe effective and ineffective job performance in terms of personality characteristics. In the past, job analysts have deliberately ignored such information because the descriptions do not refer directly to observed behavior. In fact, an entire strategy of job analysis focuses on "job-oriented" as opposed to "worker-oriented" job requirements. Nevertheless, when the job information provided by incumbents and supervisors is recorded, it will inevitably contain information about personal characteristics that are associated with varying degrees of effective job performance.

Many job analysis procedures first identify tasks of the job – what the job requires – and then identify the knowledge, skills, and abilities (KSA's) necessary to perform the tasks. The importance of these tasks and KSA's for job performance is evaluated using questionnaire methods. This leads to a final set of task and KSA items that, when analyzed through a linkage process, provide the data important for test and criterion specification.

It is in the process of identifying the abilities required for task performance that personality or personal characteristics emerge. For example, when fire fighters are asked about the abilities necessary to extinguish fires, they say such things as "You have to stay calm under pressure," or "You have to be able to trust your partner" (J. Hogan & Stark, 1992). Similarly when asking sales representatives about the abilities necessary for calling on prospective customers they report "You have to be persuasive about the product," or "You have to be confident about your pitch," or "You can't let those rejections get you down." All of these statements concern personality, and are essential for performance of job tasks.

Figure 6.2 presents some ability statements that reflect the personality dimensions associated with the FFM. We included these statements in numerous job analysis questionnaires for jobs ranging from police officer to secretary and the results are very similar. The noncognitive abilities – personality – are among the most frequently endorsed as critical for job performance.

J. Hogan and Arneson (1987) evaluated the usefulness of personality characteristics in job analysis and concluded (1) personality characteristics could be reliably evaluated by incumbents; (2) the FFM provided an adequate taxonomy of characteristics related to job performance; (3) personality characteristics were differentially associated with different occupations; and (4) if important personality characteristics are

identified in a job analysis, that information leads to tests that are valid predictors of job performance. It is important to note that a job analysis should provide information about valid job requirements. The researcher then finds measures or develops tests that assess individual differences in the performance of these job requirements.

Figure 6.2 ***Personality-related Job Analysis Abilities***

Adjustment	Ability to stay calm under pressure and deal with stress.
Ambition	Ability to assume responsibility and take charge in ambiguous situations.
Sociability	Ability to interact with a variety of people.
Likeability	Ability to tolerate provocative and/or difficult people.
Prudence	Ability to follow and carry out orders promptly.
Intellectance	Ability to generate useful ideas for completing work and problem solving
School Success	Ability to learn new material quickly.

Test Specification

When the results of a job analysis indicate that noncognitive abilities—such as those in Figure 6.2—are important, then the HPI personality scales and their constituent HICs will be appropriate experimental predictors. Using all the scales of the HPI, rather than one or two, will allow one to use the full range of the FFM for prediction; this strategy also will allow one to use combinations of predictor variables—assuming an adequate sample size. The predictor domain of the HPI contains seven personality scale variables and 41 HIC variables. In addition, combinations of HICs from the primary scales predict specific syndromes of occupational performance. These scales are described in the previous section.

The results of a job analysis often suggest other predictors of job performance. Personality, as assessed by the HPI, is relatively unrelated to cognitive abilities and therefore will not be useful for predicting learning, memory, computational, or analytic abilities. Nor will personality be the best predictor of interests, motivation, or specialized occupational skills. Finally, measures of normal personality will not be useful for screening out aberrant job-related personality disorders or psychopathological tendencies. The predictor battery for empirical validation research should be chosen on the basis of a careful linkage of job requirements to the individual difference measures most appropriate for evaluating those requirements.

Criterion Measures

In a discussion of the performance domain, Campbell (1990) emphasizes that the latent structure of constructs should extend across both the predictor and the criterion space. In short, job requirements that drive the choice of predictors should also drive the choice of criteria that best reflect job performance. Although it is logical to expect that the same constructs that underlie predictor measures will underlie job performance measures, this is frequently not the case. Researchers typically review experimental tests rather carefully but, job performance criteria are often chosen on the basis of convenience, without much concern for what is really being assessed. When this happens, the constructs underlying the predictor domain will be unrelated to the constructs underlying the job performance domain and the result is that there will be no relation between test measures and job performance criteria. Then researchers often conclude that the test “doesn’t work.”

In personality research, a major task is to identify and assess, with good psychometric quality, relevant indices of job performance. Figure 3.2 in Chapter 3 provides examples of criteria appropriate for the FFM constructs. Using this table, we can surmise that if one wants to predict managerial potential, one should assess ambition, competitiveness, persuasiveness, and the ability to get along with others. Appropriate criterion measures might include supervisor, peer, and subordinates ratings of leadership, ability to build a team, good judgment, and the desire to succeed. Matching test and criterion measures on the basis of a common construct is the key to maximizing the prediction of job performance.

If a number of relevant and reliable criterion variables could be included in a validity study, then the researcher might consider factor analyzing the variables of interest and developing criterion scores from the factors that emerge. Using the factor scores as dependent variables will result in non-redundant, reliable, and comprehensive coverage of the criterion space.

Statistical Analysis

Pearlman (1985) reviewed average validity coefficients for a number of predictor types and reported that the average validity of personality measures is .10. In this review, he made no distinction among the constructs evaluated by the personality or the job performance measures. Research by Hough, Eaton, Dunnette, Kamp, and McCloy (1990) evaluates the statistical consequences of aligning predictor and criterion measures of the same construct. Hough and her colleagues show that as constructs and their measures become better defined and delineated, validity coefficients increase. They found that, in military performance, aggregated personality scales predicted general criteria on the order of .20. With refinements in the predictor and criterion measures, validity coefficients increased to about .40. In a sense, the correlation coefficient is a conservative estimate of the latent structure similarity between the predictor and criterion construct.

Correlational analyses using the HPI should first select valid cases where the VALIDITY scale scores are 10 and above. Next, specify all HIC variables followed by all personality and occupational (if desired) scale variables. The scoring program for the HPI produces a variable list ordered in this sequence. Therefore, correlations analysis using a statistical package such as SPSS need only to specify “EMP to MANAGER” to identify 41 HIC variables, a validity scale, seven primary scales, and six occupational scales. Figure 6.3 provides the variable listing, variable names, variable order, and data definition.

Similarly, it is useful to specify individual criterion measures as well as composites. Composites might be factor scores or ratings of “overall job performance.” Then Pearson product-moment correlations can be computed between all HPI HICs and scales and all individual and composite criterion measures.

This analysis will show which HPI scales best predict the various criteria. Alternatively, a salient and reliable criterion can be specified and the HPI HICs can be regressed against it to define an optimal composite of personality predictors. Using the HICs rather than the scales will yield higher validity coefficients but reduced predictor homogeneity. The increased validity is a function of the heterogeneity of the test components (the range of FFM reflected in the 41 HICs) which allows the prediction of complex, multidimensional criterion syndromes. On the other hand, combining HICs across the HPI personality scales using regression will, by definition, reduce the internal consistency of the predictors entering the equation. Consequently, cross validation is always desirable.

If regression analyses are used, the HICs should be unit weighted unless the sample size is large and cross validation supports the beta weights. Cut-off scores should be set at a ŷ value that reflects acceptable levels of proficiency in the workplace.

Figure 6.3 ***Data File Variable Specifications***

Raw Data File

Record	Variable	Columns	Coding Values
All	Case #	1-3	
All	Record #	5-6	
1	Name	8-37	Last name, First name, no comma, no middle initial, all caps
1	Id Number	38-46	Social security number, 9 digits only
1	Gender	48	1 = Male, 2 = Female
1	Age	50-51	2 digits
1	Race	53	1 = American Indian, 2 = Asian, 3 = Black, 4 = White, 5 = Hispanic, 6 = Other
2	|1 to |52	8-59	0 = False,
3	|53 to |104	8-59	1 = True,
4	|105 to |156	8-59	blank = Missing.
5	|157 to |206	8-57	

Scored Data File

Record	Variable	Columns	Coding Values
1	(same as above)	(same as above)	(same as above)
2	VALID EMP ANX GUILT CALMEVT SOM TRUST		
	ATT COMPETE CONF DEPR LEAD IDEN SAX	8-67	HIC Scores
3	PARTIES CROWDS EXS EXH	8-67	HIC Scores
	ENT EASY SENS CARE		
	PEOPLE HOST MORAL MAST		
	VIRT AUTON SPONT		
4	IMPCON TRO SCI CUR	8-67	HIC Scores
	THS GAMES IDEAS CUL		
	EDUC MATH MEM READ		
	FOCUS IMPMAN APP		
5	VAL ADJ AMB SOC	8-63	Scale Score
	LIK PRU INT SCH		
	SOI STR REL CLERK		
	SALES MANAGER		
6	PVAL PADJ PAMB PSOC	8-63	Scale Score Percentiles
	PLIK PPRU PINT PSCH		
	PSOI PSTR PREL PCLERK		
	PSALES PMANAGER		

HIC and Scale Score Variable Names & Labels

HIC Labels		Scale Labels	
VALID	Validity	VAL	Validity
EMP	Empathy	ADJ	Adjustment
ANX	NotAnxious	AMB	Ambition
GUILT	NoGuilt	SOC	Sociability
CALM	Calmness	LIK	Likeability
EVT	EvenTempered	PRU	Prudence
SOM	NoSomaticComplaints	INT	Intellectance
TRUST	Trusting	SCH	School Success
ATT	GoodAttachment	SOI	Service Orientation
COMPETE	Competitive	STR	Stress Tolerance
CONF	SelfConfidence	REL	Reliability
DEPR	NoDepression	CLERK	Clerical Potential
LEAD	Leadership	SALES	Sales Potential
IDEN	Identity	MANAGER	Managerial Potential
SAX	NoSocialAnxiety	PVAL	Validity
PARTIES	LikesParties	PADJ	Adjustment
CROWDS	LikesCrowds	PAMB	Ambition
EXS	ExperienceSeeking	PSOC	Sociability
EXH	Exhibitionistic	PLIK	Likeability
ENT	Entertaining	PPRU	Prudence
EASY	EasyToLiveWith	PINT	Intellectance
SENS	Sensitive	PSCH	School Success
CARE	Caring	PSOI	Service Orientation
PEOPLE	LikesPeople	PSTR	Stress Tolerance
HOST	NoHostility	PREL	Reliability
MORAL	Moralistic	PCLERK	Clerical Potential
MAST	Mastery	PSALES	Sales Potential
VIRT	Virtuous	PMANAGER	Managerial Potential
AUTON	NotAutonomous		
SPONT	NotSpontaneous		
IMPCON	ImpulseControl		
TRO	AvoidsTrouble		
SCI	ScienceAbility		
CUR	Curiosity		
THS	ThrillSeeking		
GAMES	IntellectualGames		
IDEAS	GeneratesIdeas		
CUL	Culture		
EDUC	Education		
MATH	MathAbility		
MEM	GoodMemory		
READ	Reading		
FOCUS	SelfFocus		
IMPMAN	ImpressionManagement		
APP	Appearance		

Implementation

Cut-off Scores. At the conclusion of the statistical analysis one must decide how to set cut-off scores. Essentially, there are two choices, either of which is acceptable under the Uniform Guidelines provisions for criterion-related validation studies. Because higher test scores are associated with higher levels of job performance (and vice versa), candidates can be rank-ordered by test scores and then chosen in that order. This strategy is useful for public jurisdictions who test large numbers of candidates periodically and then develop an eligibility list for hiring over a one to three year time frame. On the other hand, this is a difficult strategy for organizations that evaluate candidates daily in multiple locations and hire personnel on a need basis.

An alternative to ranking candidates is use of an expectancy analysis to determine appropriate cut-off scores. Consider the following example: Expectancy tables developed by Taylor and Russell (1939) allow us to determine the expected proportion of successes among persons hired on the basis of test scores, given the validity of the selection procedure, the selection ratio, and the base rate of success.

Examination of expectancy tables reveals several trends. First, regardless of the base rate or validity, the use of selection procedures is least helpful when most of the applicants will be hired anyway (i.e., the selection ratio is high) and most helpful when only a few applicants are to be hired (i.e., the selection ratio is low). As validity increases, the expected proportion of successful employees generally increases. The rate of increase is greatest when both the base rate and selection ratio are low; it is least when the base rate is low and the selection ratio is high. When validity is low, the proportion of successful new hires grows as the selection ratio decreases and the base rate increases.

Hypothetical examples, using actual validity data presented earlier in this chapter, illustrate how the HPI occupational scales can be useful in personnel selection. First, consider a trucking company that wants to hire 25 drivers. The 60 applicants for the positions have taken the HPI. Employment records indicate that 60% of the drivers hired in the past five years have been successful employees, with the criterion of success being at least one commendation for good job performance. The company has decided to use the Reliability scale to select new drivers. The Reliability scale scores correlate .51 with commendations in a sample of 56 truck drivers. The Taylor-Russell table for a base rate of .60 is used to estimate how helpful the Reliability scale will be in hiring successful drivers. Using a selection ratio of .40 (25/60=.41) and a validity coefficient of .51, the table indicates that 79% of the 25 drivers selected on the basis of Reliability scores will be successful. This can be compared with the base rate expectancy that 60% of those selected would be successful using the company's previous selection procedures. Use of the Reliability scale leads to a 19% increase in the expected proportion of success and it represents the scale's incremental validity.

Another example using actual validity data shows that even scales with relatively low validity can contribute to the selection process. Assume that a large corporation needs to hire 10 secretaries from a pool of 200 applicants. All applicants have taken the HPI which is scored for the Clerical Potential scale, and all have demonstrated adequate typing and filing skills. In addition to secretarial skills, the job requires the ability to represent the company favorably to the public. A review of past employment evaluations of the secretarial staff indicates that only 40% received a "favorable" or better rating for interacting with the public.

Therefore, the organization decides to incorporate the Service Orientation scale into the selection process. The correlation between this scale and supervisors' job performance ratings of 100 clerical personnel is .25.

For this situation, the Taylor-Russell table for a base rate of .40 should be used to determine the proportion of the 10 hired who could be expected to be successful. Given a selection ratio of .05 (10/200) and a validity of .25, 61% (six of the ten hired) could be expected to represent the company effectively when dealing with the public. This percentage compares favorably with the base rate expectation that only four of the ten would be successful if the selection process leads to a 20% increase in the proportion of success among those selected.

Note that the group used to establish the base rate and the group used to establish the validity of the selection procedure to be used in the selection process must be similar to obtain an accurate evaluation of the selection procedure's utility. In other words, in addition to taking the HPI, applicants should go through the same selection procedures as those in the base rate group (e.g., personal interviews, reference checks).

Multiple Hurdles. A second set of decisions concerns the sequence of the decision making process for evaluating applicants. We recommend a multiple hurdle approach in which the most valid and least expensive selection tools are used near the beginning of the decision making process and the more expensive and time consuming procedures used at the end of the process. The candidates who must be screened using the final procedures will be a more highly qualified subset of the original applicant pool, therefore fewer candidates who are acceptably well-qualified will complete the more expensive procedures of the selection process.

Costing the Selection Process. Methods for estimating the costs and benefits of personnel programs and personnel selection procedures were developed in the late 1940's (Brogden, 1949); however, these methods have been virtually unused until recently. With a greater emphasis on the "bottom line," Brogden's work, which allows us to measure the cost of human resources in financial terms, has been rediscovered. These computations are termed "utility analyses," and Cascio (1982, p. 127) defines the concept as "the determination of institutional gain or loss (outcomes) anticipated from various courses of action."

We conducted four utility analyses using the HPI occupational scales in order to answer this question: What are the financial consequences to an organization of valid employment tests such as the HPI? (J. Hogan & R. Hogan, 1986, p. 20). Because all the analyses use the same formula, only the values for the terms in the equation differ.

We used procedures described by Cascio (1982) to estimate total dollar gains in improved performance over random selection. We adopted the Brogden-Cronbach-Gleser model (Brogden, 1949; Cronbach & Gleser, 1965) because this procedure integrates dollars gained or lost into the evaluation. Data necessary to determine utility include the validity of the selection procedure (concurrent validity), the selection ratio (the ratio of applicants hired to total applicants), and the standard deviation of job performance estimated in dollars. The most difficult term to obtain is the standard deviation of the dollar value of job performance (cf. Roche, 1961; Weekly, Frank, O'Connor, & Peters, 1985). Several investigations consider alternatives to time-consuming cost-accounting procedures (Bobko, Karren, & Parkington, 1983; Cascio & Silbey, 1979;

Schmidt, Hunter, McKenzie, & Muldrow, 1979). Hunter and Schmidt (1983) conclude that the standard deviation of the dollar value of job performance can be estimated conservatively as 40% of the average entry-level annual salary. This finding and the use of a "salary percentage" have promoted the use of utility analysis in personnel selection and human resources decision making.

The utility of the HPI occupational scales over random selection can be calculated using the following equation (Cascio & Ramos, 1986):

$$\Delta\mu = N_s r_{xy} SD_y \gamma/\phi - N_s C_y/\phi$$

Where Δμ = total gain in utility from use of the procedure over random selection; Ns = number of applicants selected; rxy = validity coefficient of the selection procedure; SDy = standard deviation of the job performance criterion in terms of dollars; γ = ordinate of the normal curve at the cutting score of the test procedure; ϕ = selection ratio; and Cy = cost of putting one person through the selection procedure.

Table 6.4 ***Sample Utility Results for HPI Occupational Scales***

Equation Components for Four Selection Methods:[a] $\Delta\mu = N_s r_{xy} SD_y \gamma/\phi - N_s C_y/\phi$

Terms	Service Orientation (Combination Driver)	Reliability (Line-haul Driver)	Sales Potential (Salesperson)	Managerial Potential (School Principal)
Ns	200	3	3	7
rxy	.34	.25	.25	.28
SDy	10,920	15,400	15,400	13,800
γ	.10	.04	.04	.12
ϕ	.05	.01	.01	.06
Cy	325.00	142.06	142.06	5.00
Utility per Selectee[a]	$926.00	$1,194.00	$1,194.00	7,085.00

[a] $\Delta\mu/N_s$

Results using this formula for four jobs and individual HPI scales are presented in Table 6.4. The data for the terms in the equations are based on a single hiring year. We used the 40% statistic to calculate standard deviation of the dollar-value of job performance and the criterion relationships. Validities ranged from .18 for the Reliability scale to .71 for Sales Potential. The dollar-value of job performance also varied with estimated standard deviations ranging from $10,920 for truck drivers to $14,850 for sales representatives. These two terms are directly related to utility. In each example, the gains are impressive, ranging from $926 to $14,470 per new hire. Moreover, even with modest validity coefficients, the economic benefits of valid selection procedures over random methods are considerable. Also worth noting is that, in most instances, the cost of the testing materials represented only a small fraction of the total cost of the selection process.

Finally, no system of personnel decision making is perfect. Decision errors will occur because neither tests nor people are perfect actuarial devices. Therefore, the need to maintain records and follow-up on both hits and misses in the decision making process is crucial to continual improvement of the evaluation process.