PERSONNEL PSYCHOLOGY
1990, 43

PROJECT A VALIDITY RESULTS: THE RELATIONSHIP BETWEEN PREDICTOR AND CRITERION DOMAINS

JEFFREY J. MCHENRY
American Institutes for Research

LEAETTA M. HOUGH, JODY L. TOQUAM,
MARY ANN HANSON, STEVEN ASHWORTH
Personnel Decisions Research Institute

A predictor battery of cognitive ability, perceptual-psychomotor ability, temperament/personality, interest, and job outcome preference measures was administered to enlisted soldiers in nine Army jobs. These measures were summarized in terms of 24 composite scores. The relationships between the predictor composite scores and five components of job performance were analyzed. Scores from the cognitive and perceptual-psychomotor ability tests provided the best prediction of job-specific and general task proficiency, while the temperament/personality composites were the best predictors of giving extra effort, supporting peers, and exhibiting personal discipline. Composite scores derived from the interest inventory were correlated more highly with task proficiency than with demonstrating effort and peer support. In particular, vocational interests were among the best predictors of task proficiency in combat jobs. The results suggest that the Army can improve the prediction of job performance by adding non-cognitive predictors to its present battery of predictor tests.

The purpose of this paper is to report the covariation between the Project A predictor scores and the five criterion scores identified in the previous paper. This paper has five parts. The first part describes the formation of predictor composite scores from the individual test and scale scores. In the second part, the relationship between the predictor composite scores within each predictor domain and the five job performance factors described in Campbell, McHenry, and Wise (1990) are reported. Part three demonstrates how the new predictor tests increment the validity of the current Armed Services Vocational Aptitude Battery

This research was funded by the U. S. Army Research Institute for the Behavioral and Social Sciences, Contract No. MDA903-82-C-0531. All statements expressed in this paper are those of the authors and do not necessarily express the official opinions of the U.S. Army Research Institute or the Department of the Army.

Jeffrey McHenry is now with Allstate Research Center, Menlo Park, CA. Jody Toquam is now with Battelle Research Laboratories, Seattle. Steven Ashworth is now with AllState Research Center, Menlo Park, CA.

TABLE 1

The Number of Incumbents in the Nine Army Enlisted Jobs Studied

Enlisted Job (MOS)	Number of Incumbents
Infantryman (11B)	491
Cannon Crewmember (13B)	464
Armor Crewman (19A)	394
Single Channel Radio Operator (31C)	289
Light Wheel Vehicle Mechanic (63B)	478
Motor Transport Operator (64C)	507
Administrative Specialist (71L)	427
Medical Specialist (91A)	392
Military Police (95B)	597
Total	4,039

(ASVAB). Part four describes the relationship between the new predictor tests and the two "method factors" that were identified in the analysis of the job performance measures. Finally, part five discusses how the predictor-criterion relationships identified in the validity analyses contribute to the understanding of job performance in the population of entry-level skilled jobs.

Formation of Predictor Composites

The preliminary analyses of the new Project A predictor tests indicated that 65 reliable predictor scores and four response validity scores, could be computed from the new predictor battery (Peterson et al., 1990). The six spatial tests provided six scales; the 10 computer tests yielded 20 measures of perceptual-psychomotor abilities; the Assessment of Background and Life Experiences (ABLE) provided measures of 11 temperament/personality traits; the ABLE also included the four response validity scales; the Army Vocational Interest Career Examination (AVOICE) assessed 22 vocational interests; and the Job Orientation Blank (JOB) measured six types of job outcome preferences. In addition, scores from the nine ASVAB subtests were available from Army records.

Several problems precluded using all 74 substantive scores directly. As Table 1 shows, the number of subjects with complete predictor and criterion data within the nine target Project A jobs ranged from 289 for Single Channel Radio Operator to 597 for Military Police (Young, Houston, Harris, Hoffman, & Wise, 1990). Even for Military Police, the ratio of subjects to predictor variables was only 8:1, and the stability of multiple regression estimates could be questioned. Also, scores from many of

the predictor tests were substantially intercorrelated. Composite scores would both be more reliable and exhibit less multicollinearity than any of the individual scores.

Given these considerations, the 74 predictor tests and scale scores were combined into 24 predictor composites before predictor-criterion relationships were calculated. With one exception (noted below), the composites were formed by summing unit weighted standardized scores.

Two sets of data were used to help group the predictor tests and scales into a reduced number of predictor composites. First, a principal components analysis was conducted, and tests and scales with similar patterns of factor loadings were clustered together. Second, expert judgments of predictor-criterion relationships were collected before the predictor and criterion measures were developed (Wing, Peterson, & Hoffman, 1984). This study used a sample of 35 experienced psychologists to make standardized estimates of the corrected (for range restriction and criterion unreliability) correlations between a large array of potential predictor variables and an extensive list of potential performance components. Cluster analysis was then used to identify predictors with similar profiles of expected correlations with the criterion job performance measures. The two data sets resulted in very similar clusters of predictor tests. The predictor development team reviewed the two clusterings and made the final assignment of predictor tests and scales to composites.

Table 2 shows how the individual scale and test scores were combined into the 24 predictor composite scores. The nine ASVAB subtests combined into four composite scores, Technical, Quantitative, Verbal, and Speed. Although ASVAB scores are obtained at the time of application, the scale analyses were performed on the individuals in the concurrent validation sample. In computing the Technical composite score, the Electronics Information subtest received a weight of one-half, whereas the Mechanical Comprehension and Auto Shop subtests received unit weights because a factor analysis indicated that the loading of the Electronics Information on the Technical factor of the ASVAB was only about one-half as large as the loading of the Mechanical Comprehension and Auto Shop subtests. These results are quite consistent with previous factor analyses of the ASVAB (Kass, Mitchell, Grafton, & Wing, 1983).

The six spatial tests were all highly intercorrelated (mean inter r = .46) and were combined into a single unit weighted composite score.

Six composite scales were formed from the 20 perceptual-psychomotor test scores from the computer battery. These six composites were Psychomotor, Complex Perceptual Speed, Complex Perceptual Accuracy, Number Speed and Accuracy, Simple Reaction Speed, and Simple Reaction Accuracy.

TABLE 2

Content of 24 Predictor Composite Scores in Terms of the Individual Test or Scale Scores in the Project A Predictor Battery

Test or Scale Score	Predictor Composite
From the ASVAB	
Mechanical Comprehension	Technical (TCH)
Auto Shop	
Electronics Information	
Math Knowledge	Quantitative (QUN)
Quantitative	
Arithmetic Reasoning	
Verbal	Verbal (VRB)
General Science	
Coding Speed	Speed (SPD)
Number Operations	
From the Paper-and-Pencil Spatial Tests	
Assembling Objects	Spatial (SPT)
Map	
Mazes	
Object Rotation	
Orientation	
Figural Reasoning	
From the Computerized Perceptual/Psychomotor Tests	
Cannon Shoot Test (Time Score)	Psychomotor (PSM)
Target Shoot Test (Time to Fire)	
Target Shoot Test (Log Distance)	
Target Tracking 1 (Log Distance)	
Target Tracking 2 (Log Distance)	
Pooled Mean Movement Time	
Short Term Memory Test (Decision Time)	Complex Perceptual
Perceptual Speed & Accuracy Test (Decision Time)	Speed (CPS)
Target Identification Test (Decision Time)	
Short Term Memory Test (Percent Correct)	Complex Perceptual
Perceptual Speed & Accuracy Test (Percent Correct)	Accuracy (CPA)
Target Identification Test (Percent Correct)	
Number Memory Test (Percent Correct)	Number Speed
Number Memory Test (Initial Decision Time)	and Accuracy (NSA)
Number Memory Test (Mean Operations Decision Time)	
Number Memory Test (Final Decision Time)	
Choice Reaction Time (Decision Time)	Reaction Speed (SRS)
Simple Reaction Time (Decision Time)	
Choice Reaction Time (Percent Correct)	Reaction
Simple Reaction Time (Percent Correct)	Accuracy (SRA)
From the ABLE	
Self-Esteem	Achievement
Work Orientation	Orientation (ACH)

Table 2 (continued)

Test or Scale Score	Predictor Composite
Energy Level	
Conscientiousness Non-Delinquency	Dependability (DEP)
Emotional Stability	Adjustment (ADJ)
Physical Condition	Physical Condition (CND)
From the AVOICE	
Clerical/Administrative Medical Services Leadership/Guidance Science/Chemical Data Processing Mathematics Electronic Communications	Skilled Technical (IST)
Mechanics Heavy Construction Electronics Vehicle/Equipment Operator	Structural/Machines (ISM)
Combat Rugged Individualism Firearms Enthusiast	Combat-Related (ICM)
Drafting Arts Audiographics Aesthetics	Audiovisual Arts (IAV)
Food Service Professional Food Service Employee	Food Service (IFS)
Law Enforcement Protection	Protective Services (IPS)
From the JOB Questionnaire	
Job Status Job Security Serving Others Ambition	Organizational and Co-Worker Support (JSP)
Routine	Routine Work (JRT)
Autonomy	Job Autonomy (JAT)

Four temperament/personality composites were computed from the ABLE scales. The composites were named Achievement Orientation, Dependability, Adjustment, and Physical Condition. Four of the 11 ABLE temperament/personality scales were not included in any composite. These four scales had been developed to predict performance as a manager or supervisor, which was not part of the job responsibilities of

Table 2A

Mean Intercorrelations[a] Among the 24 Predictor Composites and the Five Performance Criterion Scores (29 × 29) Computed on the Full Samples from the Nine Batch A MOS

Scale	VRB	QUN	TCH	SPD	SPT	PSM	CPA	CPS	NSA	SRS	SRA	ACH	DEP	ADJ
VRB														
QNT	52													
TCH	56	45												
SPD	1	26	−7											
SPT	41	56	50	17										
PSM	21	23	33	11	45									
CPA	16	19	13	9	26	15								
CPS	11	11	11	15	27	27	−35							
NSA	28	50	20	35	35	28	10	29						
SRS	8	9	4	13	13	21	−1	34	19					
SRA	13	11	14	3	14	6	16	0	8	5				
ACH	5	11	8	8	7	6	5	3	10	3	0			
DEP	−2	5	−3	8	1	−3	7	−3	2	0	0	58		
ADJ	10	11	14	1	12	12	6	5	9	6	3	58	32	
CND	−6	−2	−3	1	−4	11	−4	7	4	6	−3	37	14	24
IST	−3	9	−6	9	−1	0	2	0	5	2	−5	30	30	15
ISM	−6	−3	22	−10	7	10	−1	1	−2	−2	0	13	0	8
ICM	11	3	26	−6	13	18	1	5	4	0	1	21	1	17
IAV	4	3	−1	0	4	1	1	1	−5	−1	−2	17	18	5
IFS	−10	−3	−14	2	−9	−14	−3	−6	−5	−4	−3	0	7	−4
IPS	−9	−9	−1	−2	−5	3	−1	2	−3	0	−1	15	9	7
JSP	5	0	0	7	−1	0	0	0	4	2	0	30	24	8
JRT	−15	−10	−12	−4	−12	−7	−1	−4	−7	−3	−1	−16	0	−12
JAT	14	9	13	2	7	2	−2	3	5	1	0	15	−2	4
CTP	32	33	34	9	38	20	17	9	22	8	10	11	11	10
GSP	36	38	40	7	47	28	21	12	24	9	10	11	9	12
ELS	7	12	18	8	14	13	7	7	10	2	1	30	22	19
MPD	5	10	8	5	7	0	6	0	3	0	3	18	30	11
PFB	−11	−1	−9	10	−4	1	−1	4	1	6	−5	28	22	17

Note: Abbreviations for the predictor composites (the first 24 scales in the matrix) are given in Table 2. Abbreviations for the job performance factors (the last five scales) are as follows: CTP—Core Technical Proficiency; GSP—General Soldiering Proficiency; ELS

first-tour soldiers included in this study. Nor were any of the four ABLE response validity scales used in computing composites.

Six vocational interest composites were computed from the AVOICE scales: Skilled Technical, Structural/Machines, Combat-Related, Audiovisual Arts, Food Service, and Protective Services. One of the 22 AVOICE scales, Shipping/Warehousing, was not included in any composite.

Finally, the six scales of the JOB were combined into three composites: Organizational and Co-Worker Support, Routine Work, and Job Autonomy.

Table 2A (continued)

CND	IST	SIM	ICM	IAV	IFS	IPS	JSP	JRT	JAT	CTP	GSP	ELS	MPD	PFB
10														
13	33													
15	25	57												
6	65	27	22											
−1	37	25	13	32										
12	29	31	38	16	18									
9	24	4	6	16	0	12								
−5	2	6	−4	0	17	2	−18							
5	6	9	14	11	−2	3	25	−7						
−3	2	9	16	0	−8	0	6	−9	7					
0	2	9	21	3	−8	1	4	−11	6	58				
11	5	8	16	1	−7	6	8	−8	5	28	28			
0	4	−4	0	0	−4	−2	5	−2	−3	19	17	59		
30	11	0	3	7	0	2	9	−3	−1	3	5	46	33	–

—Effort and Leadership; MPD—Personal Discipline; PFB—Physical Fitness and Military Bearing.

[a] Decimals omitted.

Relationships Between Predictor Domains and Job Performance Constructs

Table 2A shows a basic zero-order matrix of the mean intercorrelations, averaged over the nine Batch A MOS, among the 24 predictor scores (as defined above) and the five job performance criterion factors described in the C. H. Campbell et al. (1990) paper. Table 3 again portrays the definition of these five factors.

It was hypothesized that the composite scores measuring cognitive abilities would be useful for predicting scores on the two task performance factors, Core Technical Proficiency and General Soldiering Proficiency, as well as the Effort and Leadership component, since that factor also contained rating scales intended to assess task proficiency performance components. It was further hypothesized that the temperament/interest composites would add significantly to the prediction of the

TABLE 3

A Five-Factor Model of Job Performance for Entry-Level Jobs Based on Analyses of Project A Performance Data

1. *Core Technical Task Proficiency*: The proficiency with which the individual performs the tasks that are specific and "central" to his or her job (MOS). The tasks represent the core content of the job that distinguishes it from other jobs.
2. *General Task Proficiency*: In addition to the core technical content specific to an MOS, individuals in every job are responsible for being able to perform a variety of general or common tasks—e.g., use of basic weapons, first aid. This factor represents proficiency on these general tasks.
3. *Peer Support and Leadership, Effort, and Self Development*: Reflects the degree to which the individual exerts effort over the full range of job tasks, perseveres under adverse or dangerous conditions, and demonstrates leadership and support toward peers.
4. *Maintaining Personal Discipline*: Reflects the degree to which the individual adheres to Army regulations and traditions, exercises personal self-control, demonstrates responsibility in day-to-day behavior, and does not create disciplinary problems.
5. *Physical Fitness and Military Bearing*: Represents the degree to which the individual maintains an appropriate military appearance and bearing and stays in good physical condition.

Effort and Leadership factor and would be the best predictors of Personal Discipline and of Physical Fitness and Military Bearing.

Analyses of Predictor Validities

To assess the relationships between predictor domains and job performance factors, the multiple correlation of the predictor composites within each domain with each of the five job performance factors was computed. This was done separately for each of the nine jobs, and each R was corrected for range restriction and adjusted for shrinkage.

The procedure used to correct R for range restriction is described in Lord and Novick (1968). The method adjusted the covariances among the ASVAB subtests so that they matched the covariances obtained in the 1980 youth population (Mitchell & Hanser, 1984). For each MOS, the covariances among the complete set of predictor composite scores and between the predictor composite scores and the performance factor scores were then adjusted according to their covariances with the ASVAB subtests. This procedure takes into account any range restriction related to the abilities measured in the ASVAB, but it fails to consider factors that are unrelated to the abilities tapped by the ASVAB. For example, a number of individuals who enlisted in the Army at the

TABLE 4

Mean Within-Job Corrected and Uncorrected Validities[a] for the Composite Scores Within Each Predictor Domain

Job Performance Factor	Predictor Domain: General Cognitive Ability (K=4)[b]	Spatial Ability (K=1)	Perceptual-Psychomotor Ability (K=6)	Temperament/ Personality (K=4)	Vocational Interest (K=6)	Job Reward Preference (K=3)
Core Technical Proficiency	.63 (.43)	.56 (.38)	.53 (.32)	.26 (.15)	.35 (.24)	.29 (.13)
General Soldiering Proficiency	.65 (.47)	.63 (.47)	.57 (.37)	.25 (.15)	.34 (.25)	.30 (.14)
Effort and Leadership	.31 (.22)	.25 (.14)	.26 (.15)	.33 (.30)	.24 (.20)	.19 (.12)
Personal Discipline	.16 (.11)	.12 (.08)	.12 (.07)	.32 (.31)	.13 (.11)	.11 (.09)
Physical Fitness and Military Bearing	.20 (.16)	.10 (.08)	.11 (.08)	.37 (.36)	.12 (.13)	.11 (.10)

[a]Validity coefficients were corrected for range restriction and adjusted for shrinkage. Uncorrected *R*s are in parentheses.

[b]K is the number of predictor scores.

same time as the concurrent sample left the Army as a result of disciplinary problems. While such problems may be unrelated to abilities tapped by the ASVAB, they might be related to the trait measures in the ABLE. Such attrition means that the variance of the temperament/personality scores in the validation sample is probably less than the variance to be expected in an unselected sample of 18 to 20-year-olds.

The procedure used to adjust R for shrinkage was developed by Claudy (1978). The adjustment is intended to yield an estimate of R that is equal to the expected value of the multiple correlation between the predictor scores and the criterion in the population from which the sample was drawn.

Given six predictor domains and five job performance factors, there were 30 corrected and uncorrected multiple correlations generated for each of the nine jobs. These *R*s were averaged across the nine jobs to obtain the mean validity for each predictor domain by performance factor combination, and the 30 mean corrected and uncorrected *R*s are reported in Table 4. The table shows that averaged across jobs, the hypothesized predictor-criterion relationships were, by and large, confirmed.

The general cognitive ability composites, computed from the ASVAB, were the best predictors of Core Technical Proficiency (mean R = .63) and General Soldiering Proficiency (mean R = .65). Recall that the ASVAB was administered an average of two years prior to the collection

of the criterion data. The spatial ability composite and the perceptual-psychomotor ability composites also provided substantial prediction of Core Technical Proficiency and General Soldiering Proficiency.

As hypothesized, the general cognitive ability composites based on ASVAB also predicted Effort and Leadership (corrected mean $R = .31$). The mean R with Effort and Leadership was only slightly lower for the composite scores from the other two cognitive domains, spatial ability (mean $R = .25$) and perceptual-psychomotor ability (mean $R = .26$). However, the composites from the three cognitive domains did not predict performance on Personal Discipline or Physical Fitness and Military Bearing very well. None of the six average multiple correlations exceeded .20.

The best prediction of Factors 3, 4, and 5 was provided by the temperament/personality composites from the ABLE. The mean R for predicting the Effort and Leadership factor was .33. The ABLE composite that contributed most to this correlation was Achievement Orientation. For predicting Personal Discipline, the mean R was .32, with the ABLE Dependability composite making the largest contribution. Finally, the ABLE composite correlated .37 on average with Physical Fitness and Military Bearing. The key predictor of this performance factor was the ABLE Physical Condition composite. The temperament/personality domain provided poorer prediction of the two task performance criteria than any of the other five predictor domains. The mean R for Core Technical Proficiency was only .26, while the mean R for General Soldiering Proficiency was .25.

The relationships between the vocational interest composites and the job performance factors were somewhat different than expected. The performance factors predicted best from the interest composites were Core Technical Proficiency (mean $R = .35$) and General Soldiering Proficiency (mean $R = .34$). The performance factors predicted least well from the interest composites were Personal Discipline (mean $R = .13$) and Physical Fitness and Military Bearing (mean $R = .12$). The mean R between the interest composites and Effort and Leadership was .24. The pattern of correlations across the five performance factors was more like the pattern for the cognitive predictor domains than the pattern for the temperament/personality domain.

The pattern of correlations for the job reward preference composites was similar to that for the vocational interest composites, though the mean Rs were a bit lower for all five performance factors. As a further test of the hypothesized predictor-criterion relationships, the predictor composites were grouped into two sets. The 11 general cognitive ability, spatial ability, and perceptual-psychomotor ability composites were grouped into one set, and the 13 temperament, vocational interest, and

TABLE 5

Mean Validity[a] for the Cognitive Ability, the Temperament-Interest-Reward Preference, and the Combined Predictor Composites

Job Performance Factor	Predictor Composites: Cognitive Ability (K=11)[b]	Temperament-Interest-Reward Preference (K=13)	All (K=24)
Core Technical Proficiency	.65	.44	.67
General Soldiering Proficiency	.69	.44	.70
Effort and Leadership	.32	.38	.44
Personal Discipline	.17	.35	.37
Physical Fitness and Military Bearing	.23	.38	.42

[a] Validity coefficients were corrected for range restriction and adjusted for shrinkage.

[b] K is the number of predictor scores.

job reward preference composites were grouped into a second set. For each set the R was computed with each of the five job performance factors within each of the nine jobs. Mean Rs from these analyses are presented in Table 5.

The obtained pattern was very similar to that predicted. The ability composites provide the best prediction of Core Technical Proficiency (mean R = .65) and General Soldiering Proficiency (mean R = .69). The temperament-interest-reward preference composites provided the best prediction of Personal Discipline (mean R = .35) and Physical Fitness and Military Bearing (mean R = .38). The temperament-interest-reward preference composites also predicted Effort and Leadership better than did the ability composites, though the difference was not very large (mean Rs = .38 and .32, respectively).

Table 5 also shows that, when all 24 predictor composite scores were used to predict each performance factor, the mean Rs were .67 for Core Technical Proficiency, .70 for General Soldiering Proficiency, .44 for Effort and Leadership, .37 for Personal Discipline, and .42 for Physical Fitness and Military Bearing. These results indicated that for at least two of the job performance factors—Effort and Leadership and Physical Fitness and Military Bearing—the best prediction was obtained when all predictors were used.

The one surprising result in Table 5 was the high correlation between the temperament-interest-reward preference predictors and the two task performance factors. For both factors, the mean R was .44. In fact, the temperament-interest-reward preference composites predicted the first two performance factors better than they predicted the last three. While

TABLE 6

Mean Incremental Validity[a,b] for the Composite Scores Within Each Predictor Domain

Job Performance Factor	Predictor Domain: General Cognitive Ability (K=4)[c]	General Cognitive Ability Plus Spatial Ability (K=5)	General Cognitive Ability Plus Perceptual-Psychomotor Ability (K=10)	General Cognitive Ability Plus Temperament/ Personality (K=8)	General Cognitive Ability Plus Vocational Interest (K=10)	General Cognitive Ability Plus Job Reward Preference (K=7)
Core Technical Proficiency	.63	.65	.64	.63	.64	.63
General Soldiering	.65	.68	.67	.66	.66	.66
Effort and Leadership	.31	.32	.32	.42	.35	.33
Personal Discipline	.16	.17	.17	.35	.19	.19
Physical Fitness and Military Bearing	.20	.22	.22	.41	.24	.22

[a] Validity coefficients were corrected for range restriction and adjusted for shrinkage.

[b] Incremental validity refers to the increase in R afforded by the new predictors above and beyond the R for the Army's current predictor battery, the ASVAB.

[c] K is the number of predictor scores.

differential criterion reliabilities might account for some of the difference, it seems unlikely. The five criterion factor scores are themselves composites of scores that have individual reliabilities of .60 or higher, and there should be very little differential attenuation. Also, as will be seen later, there are differential patterns of results across variables that are not compatible with a differential reliability explanation.

The Incremental Validity of the Project A Predictor Tests

To address the question of incremental validity, the validity of the general cognitive ability composite scores (computed from the ASVAB) was compared with the validity obtained when the composite scores from other predictor domains were added to the regression equation. That is, for each estimate OLS weights were computed and the resulting R was adjusted for shrinkage. This was done for each performance factor within each of the nine jobs, and validities were averaged across jobs. The resulting mean validities are reported in Table 6.

They indicate that none of the predictor domains added more than .02 to the general cognitive ability composites' validity for predicting

TABLE 7

Mean Validity and Incremental Validity[a,b] for the Project A Cognitive Ability and the Project A Temperament-Interest-Reward Preference Predictor Composites

	Predictor Composites				
		New Cognitive Ability Predictor Variables		New TIR[d] Variables	
Job Performance Factor	General Cognitive Ability (ASVAB) Composites (K=4)[c]	New Project A Cognitive Composites (K=7)	New Project A Cognitive Composites Plus ASVAB Composites (K=11)	New Project A TIR Composites (K=13)	New Project A TIR Composites Plus ASVAB Composites (K=17)
Core Technical Proficiency	.63	.59	.65	.44	.65
General Soldiering Proficiency	.65	.65	.69	.44	.67
Effort and Leadership	.31	.27	.32	.38	.43
Personal Discipline	.16	.13	.17	.35	.37
Physical Fitness and Military Bearing	.20	.14	.23	.38	.41

[a] Validity coefficients were corrected for range restriction and adjusted for shrinkage.

[b] Incremental validity refers to the increase in R afforded by the new predictors above and beyond the R for the Army's current predictor battery, the ASVAB.

[c] K is the number of predictor scores.

[d] Temperament-Interest-Reward Preference Variables

Core Technical Proficiency. Similarly, no predictor domain added more than .03 to the general cognitive ability composites' validity for predicting General Soldiering Proficiency. However, in both instances, the predictor composite that did add the greatest increment was spatial ability.

The greatest amount of incremental validity was generated by the ABLE. The four temperament/personality composite scores from the ABLE added .11 to the validity for predicting Effort and Leadership, .19 to the validity for predicting Personal Discipline, and .21 to the validity for predicting Physical Fitness and Military Bearing.

Table 7 provides another means for looking at the incremental validity and shows that the seven new cognitive scores (i.e., spatial ability plus the six perceptual-psychomotor composites) predicted job performance almost as well as the four composite scores from the ASVAB. For Core Technical Proficiency and General Soldiering Proficiency, the validity of the Project A cognitive composites was quite high (mean R = .59 and .65, respectively). However, the increments provided by the new tests over

the ASVAB's prediction of Core Technical and General Soldiering were only .02 and .04. As a practical matter, there perhaps is not much room for improvement when the validity of the predictor has already reached .63 or .65.

Table 7 also shows that the temperament-interest-reward preference composite predicted Effort and Leadership (mean R = .38), Personal Discipline (mean R = .35), and Physical Fitness and Military Bearing (mean R = .38) better than did the ASVAB above. When the ASVAB was added, the mean validity increased .05 for Effort and Leadership, .02 for Personal Discipline, and .03 for Physical Fitness and Military Bearing.

A comparison of Tables 6 and 7 shows that almost all of the incremental validity was provided by the ABLE. When the AVOICE composites and the JOB composites were added to the ABLE and ASVAB composites, the mean validity increased only by .01. Similarly, the AVOICE and JOB composites added only .02 to the prediction of Personal Discipline and added nothing to the prediction of Physical Fitness and Military Bearing.

Relationships Between Predictor Domains and "Method Factors"

J. P. Campbell et al. (1990) described written test and rating method factors that emerged from a structural analysis of the job performance measures. However, the term "method factor" may be a misnomer. The written test factor may reflect comprehension of the manuals, instructions, and other materials that must be read on the job. For several of the jobs that were studied, excerpts from technical manuals and other learning aids were incorporated into the written knowledge tests. The rating factor represents the unique variance that rating scales contributed to the assessment of performance. It is similar to what many researchers might term "halo error" (cf. Saal, Downey & Lahey, 1980). However, it is also possible that the rating factor represents a global assessment of performance that is an important component of effectiveness (Cooper, 1981; Feldman, 1986; Landy & Farr, 1980; Murphy, 1982). The Project A data base does provide an opportunity to study the relationships between this rating factor and individual difference variables from several domains.

Table 8 shows the multiple correlations between the predictors within each domain and the two method factors. The mean Rs for the written test factor were much greater than the mean Rs for the rating factor across all six predictor domains.

TABLE 8

Mean Validity[a] for the Composite Scores Within Each Predictor Domain for the "Method Factor" Scores

Criterion Method Factor	Predictor Domain: General Cognitive Ability (K=4)[b]	Spatial Ability (K=1)	Perceptual-Psychomotor Ability (K=6)	Temperament/ Personality (K=4)	Vocational Interest (K=6)	Job Reward Preference (K=3)
Written Test	.62	.55	.54	.21	.32	.28
Rating	.15	.07	.08	.18	.09	.08

[a] Validity coefficients were corrected for range restriction and adjusted for shrinkage.

[b] K is the number of predictor scores.

The best predictors of the written test factor were the general cognitive ability composites (mean R = .62), and across the nine jobs the ASVAB verbal composite was the most consistent predictor. However, the spatial ability composite and the perceptual-psychomotor ability composites, using very different formats, had mean correlations of .55 and .54, respectively. While correlations were lower for the composites within the three non-cognitive domains, they were not trivial, ranging from .21 for the temperament/personality composites to .32 for the vocational interest composites. This pattern of correlations suggests variance that is much more general in meaning than "test taking" skill.

The best predictors of the rating factor were the temperament/personality composites (mean R = .18). Within the temperament/personality domain, the most consistent predictor of the rating factor was the ABLE dependability composite. After the temperament/personality composites, the second best predictors were the general cognitive ability composites (mean R = .15). Mean correlations for the composites within the remaining four domains all were less than .10. This pattern of correlations suggests that the rating factor taps dependability and competence on the job, but much more evidence would be needed to confirm this interpretation.

For Table 9, the predictor composites again were grouped into the ability and temperament-interest-reward preference sets. For the written test factor, the mean Rs across the nine jobs were .64 for the 11 scales in the ability set scores, .40 for the 13 scales in the temperament-interest-reward preference, and .65 across all 24 predictor scales. For the rating factor, the mean Rs were .16, .22, and .26, respectively.

The pattern of correlations for the rating factor is similar to the pattern for the Effort and Leadership performance factor. This suggests that the rating factor obtained in this study reflects raters' global assessment of soldiers' overall competency and dependability. That is, when

TABLE 9

Mean Validity[a] for the Cognitive Ability, the Temperament-Interest-Reward Preference, and All Predictor Composites for the "Method Factor" Scores

	Predictor Composites		
Criterion Method Factor	Cognitive Ability (K=11)[b]	Temperament-Interest-Reward Preference (K=13)	All (K=24)
Written Test	.64	.40	.65
Rating	.16	.22	.26

[a] Validity coefficients were corrected for range restriction and adjusted for shrinkage.
[b] K is the number of predictor scores.

raters were asked to evaluate a soldier on a particular rating dimension, they considered the soldier's performance on that dimension and two other factors as well. The first factor was their general assessment of how well the soldier was capable of performing the job. The second was their judgment of the soldier's dependability (cf. Feldman, 1986).

Another method of studying the two method factors is to examine how the pattern of predictor-criterion relationships changes when the variance attributable to the method factors is removed from the performance factor scores. These results are presented in Table 10.

The validity coefficients presented for the "raw" performance factor scores in Table 10 are the same as those presented in Table 4. To compute residual performance factor scores, the variance attributable to the written test factor was partialed from the scores for Core Technical Proficiency and General Soldiering Proficiency, and the variance attributable to the rating factor was partialed from the scores for Effort and Leadership, Personal Discipline, and Physical Fitness and Military Bearing. Written knowledge tests were not used in computing scores for Effort and Leadership, Personal Discipline, or Physical Fitness and Military Bearing. Nor were rating scales used in computing scores for Core Technical Proficiency or General Soldiering Proficiency.

The table shows that the residual scores for Core Technical Proficiency and General Soldiering Proficiency were less predictable than the raw scores, by all six predictor domains. The decrease in the mean R was greater for the ability predictor domains than for the temperament-interest-reward preference domains. It is interesting that the difference between the mean R for raw and residual scores for Core Technical Proficiency and General Soldiering Proficiency was approximately .16 across all three cognitive ability domains. Given a "method variance" interpretation, one might expect that the decrease in the mean R would be somewhat greater for the general cognitive ability composites (which included

TABLE 10

Mean Validity[a] for the Composite Scores within Each Predictor Domain for the Raw and Residual Job Performance Factor Scores

Job Performance Factor	Predictor Domain: General Cognitive Ability (K=4)[b]	Spatial Ability (K=1)	Perceptual-Psychomotor Ability (K=6)	Temperament/ Personality (K=4)	Vocational Interest (K=6)	Job Reward Preference (K=3)
Core Technical Proficiency						
Raw Score	.63	.56	.53	.26	.35	.29
Residual Score	.47	.37	.37	.22	.28	.21
General Soldiering Proficiency						
Raw Score	.65	.63	.57	.25	.34	.30
Residual Score	.49	.48	.41	.21	.26	.22
Effort and Leadership						
Raw Score	.31	.25	.26	.33	.24	.19
Residual Score	.46	.41	.38	.31	.32	.27
Personal Discipline						
Raw Score	.16	.12	.12	.32	.13	.11
Residual Score	.19	.15	.13	.28	.15	.10
Physical Fitness and Military Bearing						
Raw Score	.20	.10	.11	.37	.12	.11
Residual Score	.21	.11	.14	.35	.14	.10

[a] Validity coefficients were corrected for range restriction and adjusted for shrinkage.

[b] K is the number of predictor scores.

verbal ability) than for the perceptual-psychomotor ability composites (which were computed from computer tests that required relatively little reading).

For the Effort and Leadership factor, the cognitive ability composites predicted the residual performance scores better than they predicted the raw scores. For example, the mean R of the general cognitive ability composites with the raw score was .31, while the mean R with the residual score was .46. The increase was .16 for the spatial ability composite and .12 for the perceptual-psychomotor ability composites. For the temperament/personality composites, the results were the opposite. The mean multiple correlation of the temperament/personality composites with the raw Effort and Leadership score was .33, while the mean R with the residual score was .31.

The vocational interest composites and the job reward preference composites actually "behaved" similarly to the cognitive ability composites. For both predictor domains, the mean R was greater for the residual Effort and Leadership score than for the raw Effort and Leadership score.

This pattern of correlations for Effort and Leadership suggests two interesting conclusions. First, the pattern provides additional evidence

that the vocational interest composites are more similar to cognitive predictors than to temperament/personality predictors. Second, the changes in the pattern of predictor correlations for raw versus residual criterion scores suggest that the Effort and Leadership factor becomes more like a task performance factor when the rating method factor is partialed out. Also, the average multiple correlations between the ability predictor composites and the residual Effort and Leadership score are very similar to the multiple R's obtained when the same predictors are used to predict the residual scores for the task performance factors.

On the other hand, the multiple correlation of the temperament/personality composites with the residual Effort and Leadership score is much higher than for the two residual task proficiency factor scores (mean R = .31 for Effort and Leadership, .22 for Core Technical Proficiency, and .21 for General Soldiering Proficiency). This indicates that, even after the rating factor is partialed from the raw Effort and Leadership score, the residual Effort and Leadership score continues to reflect the more "will do" portion of the job performance space. In contrast, partialing the rating factor from the Personal Discipline and Physical Fitness and Military Bearing scores had little impact on the correlations of these scores with the predictor composites.

Summary and Conclusions

The pattern of predictor-criterion relationships presented in this paper was consistent with the pattern that was expected. Across nine very different jobs, the mean R for the complete set of 11 cognitive ability composite scores was .65 for Core Technical Proficiency and .69 for General Soldiering Proficiency. Clearly, these provide excellent prediction of critical task proficiency for Army enlistees. The temperament/personality variables were the best predictors of Personal Discipline and Physical Fitness and Military Bearing. The best prediction of Effort and Leadership was obtained when both cognitive ability and temperament/personality predictors were used.

The pattern of predictor-criterion relationships enhanced understanding of both the predictor space and the job performance space. On the predictor side, the vocational interest composites provided surprisingly good prediction of Core Technical Proficiency and General Soldiering Proficiency. Because the interest and job performance data were collected concurrently, it is not possible to be certain that prior interest will predict subsequent performance. However, the results do suggest the value of investigating the relationship between vocational interests and various job performance criteria, including job satisfaction and retention, in a longitudinal study.

On the criterion side, the pattern of predictor-criterion correlations added considerable meaning to the interpretation of the five-component job performance model. The pattern of correlations also enhanced understanding of the Effort and Leadership factor, the written test and rating method factors, and the relationship between raw and residual performance factor scores.

In sum, small but significant improvements in the prediction of job-specific and general soldiering task proficiency can be realized through the use of the new spatial and perceptual-psychomotor tests. However, potentially the largest gains in validity can be obtained by using the temperament/personality scales from the ABLE to improve the prediction of Effort and Leadership, Personal Discipline, and Physical Fitness and Military Bearing. These are critical components of overall performance and should not be overlooked by a personnel selection and classification system. Performance is more than being able to perform critical tasks under standardized conditions. It is not one thing.

REFERENCES

Campbell CH, Ford P, Rumsey MG, Pulakos ED, Borman WC, Felker DB, de Vera MV, Riegelhaupt, BJ. (1990). Development of multiple job performance measures in a representative sample of jobs. PERSONNEL PSYCHOLOGY, *43*, 277–300.

Campbell JP, McHenry JJ, Wise LL. (1990). Modeling job performance in a population of jobs. PERSONNEL PSYCHOLOGY, *43*, 313–333.

Claudy JG. (1978). Multiple regression and validity estimation in one sample. *Applied Psychological Measurement*, *2*, 595–601.

Cooper WH. (1981). Ubiquitous halo. *Psychological Bulletin*, *90*, 218–244.

Feldman JM. (1986). A note on the statistical correction of halo error. *Journal of Applied Psychology*, *71*, 173–176.

Kass RA, Mitchell KJ, Grafton FC, Wing H. (1983). Factor structure of the Armed Services Vocational Aptitude Battery (ASVAB) Forms 8, 9, and 10: 1981 Army applicant sample. *Educational and Psychological Measurement*, *43*, 977–1088.

Landy FJ, Farr JL. (1980). Performance rating. *Psychological Bulletin*, *87*, 72–107.

Lord P, Novick M. (1968). *Statistical theory of mental test scores*. Reading, MA: Addison-Wesley Publishing Company.

Mitchell KJ, Hanser LM. (1984). *The 1980 youth population norms: Enlisted and occupational classification standards in the Army* (ARI Technical Report in preparation). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Murphy KR. (1982). Difficulties in the statistical control of halo. *Journal of Applied Psychology*, *67*, 161–164.

Peterson NG, Hough LM, Dunnette, MD, Rosse, RL, Houston JS, Toquam JL, Wing H. (1990). Project A: Specification of the predictor domain and development of new selection/classification tests. PERSONNEL PSYCHOLOGY, *43*, 247–276.

Saal F, Downey R, Lahey M. (1980). Rating the ratings: Assessing the psychometric quality of rating data. *Psychological Bulletin*, *88*, 413–428.

Wing H, Peterson NG, Hoffman RG. (1984, August). *Expert judgments of predictor-criterion validity relationships.* Paper presented at the 92nd Annual Convention of the American Psychological Association, Toronto, Canada.

Young WY, Houston JS, Harris JH, Hoffman RG, Wise LL. (1990). Large-scale predictor validation in Project A: Data collection procedures and data base preparation. PERSONNEL PSYCHOLOGY, *43*, 301–311.