DOCUMENT RESUME

ED 291 793 TM 011 086

AUTHOR Hough, Leaetta M.
TITLE Overcoming Objections to the Use of Temperament Variables in Selection.
INSTITUTION Army Research Inst. for the Behavioral and Social Sciences, Alexandria, Va.
SPONS AGENCY Department of the Army, Washington, D.C.
PUB DATE 87
CONTRACT MDA-903-82-C-0531
NOTE 32p.; Paper presented at the Annual Meeting of the American Psychological Association (95th, New York, NY, August 28-September 1, 1987).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Job Performance; Military Personnel; *Personality Measures; *Personnel Selection; *Predictive Measurement; Predictive Validity; Psychometrics; *Test Construction; Test Validity

ABSTRACT

Much of the scientific community has believed that temperament variables could not be included in batteries of tests to predict job performance because no generalized principles could be discerned from the results. For Project A, a major Army project on the prediction of job performance, a temperament inventory was developed and implemented. This inventory overcomes objections through a carefully chosen research strategy involving: (1) a literature review to study predictor constructs; (2) an inventory of non-sensitive items and scales designed to detect intentional distortion of self-description; (3) a criterion-related validity study of the job-related nature of the temperament scales; and (4) examination of the effects of motivational sets on scale score and criterion-related validities. Through these stringent requirements, objections to the use of temperament variables were overcome for Project A. Tables of constructed scales are included. (SLD)

Overcoming Objections to the Use of Temperament Variables in Selection

Presented by Leaetta M. Hough
Personnel Decisions Research Institute

Symposium: New Perspectives on Personality and Job Performance (Chair, R. C. Page)

95th Annual American Psychological Association Convention
August 1987, New York City

A major part of this research was funded by the Army Research Institute, Contract No. MDA-903-82-C-0531. All statements expressed in this paper are those of the author and do not necessarily express the official opinions of the U.S. Army Research Institute or the Department of the Army.

Overcoming Objections to the Use of Temperament Variables in Selection

In 1982, I was assigned responsibility for developing temperament, biodata, and interest measures for Project A, a major research project funded by the Army Research Institute to improve prediction of job performance of Army enlisted personnel.

When we started, much of the scientific community believed it would be a waste of time to include temperament variables in a selection battery. There were at least five sources of negative opinion. First, in 1966 Guion and Gottier published an article in Personnel Psychology that affected the scientific community's attitude and knowledge about the usefulness of temperament variables for predicting job performance criteria. They reviewed the criterion-related validities of temperament variables and concluded that, though temperament variables have criterion-related validity more often than can be expected by chance, no generalized principles could be discerned from the results.

A second source of negative opinion about temperament variables came in the form of a theoretical challenge. In 1968, Walter Mischel published his highly influential book that caused an intense examination of and debate over trait conceptions. Mischel asserted that the apparent evidence of cross-situational consistency of behavior was a function of the use of self report as the measurement approach, that traits were an illusion. He proposed "situationism," stating that behavior is explained more by differences in situations than differences in people.

Thus, in 1982 much of the scientific community was persuaded by the published literature and believed that temperament measures had little theoretical merit and were of little practical use. Even those who

thought temperament measures might have some merit were concerned that temperament scales might be inappropriate and unfair to people who were protected under the 1964 Civil Rights Act. In addition, many people worried about intentional distortion of self descriptions in an applicant setting.

Equally important and negative was the lay community's perception of temperament inventories. People objected to offensive items and resented being asked to respond to such items. Researchers had been sensitized by the lay community's negative reaction to temperament inventories and were legitimately leery of antagonizing the public.

This was the environment in 1982.

Now, in 1987, Army generals are asking us to implement the temperament inventory we developed. What did we do to bring this about?

RESEARCH STRATEGY

A lot of time and effort was required. We also had a research strategy. That strategy is outlined on page two of your handout. I'd like to describe that approach and some of our findings. The research strategy was construct oriented and included four basic steps: (1) a literature review to identify predictor constructs that were likely to predict job performance criteria important to the Army, (2) the development of a temperament inventory that consisted of nonsensitive items and scales designed to detect intentional distortion of self descriptions, (3) a criterion-related validity study to identify temperament scales that were job-related, and (4) an examination of the effects of motivational sets on scale scores and criterion-related validities.

Literature Review

Predictor and criterion taxonomies. Since our approach was construct oriented for both predictors and criteria, we needed a taxonomy for both predictors and criteria. The criterion categories were education, training, job involvement, job proficiency, and adjustment. For the predictors, we started with the structure initially found by Tupes and Christal (1961) in the early 60s. Following Hogan's thinking in the early 80s, we split one of the constructs into two. Thus, our predictor taxonomy consisted of six constructs: Surgency, Affiliation, Adjustment, Agreeableness, Dependability, and Intellectance.

Categorization of temperament scales. Once we had a predictor taxonomy, our next step was to categorize existing temperament scales into the classification scheme. From articles and manuals, we obtained hundreds of correlations between temperament scales. We categorized the temperament scales into the six categories and a miscellaneous category, and then refined the classifications through an iterative process of classifying and reclassifying temperament scales to maximize the mean within-category correlations and minimize the mean between-category correlations. The results of this process are shown in Table 1 of your handout. The circles in the diagonal show the mean within-category correlations which are in the .30s and .40s and are, in all cases, higher than the mean between-category correlations.

Meta analysis of criterion-related validities. Our next step was to summarize the criterion-related validities according to these constructs; Table 2 of your handout shows the results. It is a meta analysis of the criterion-related validities of scales within each predictor construct for each criterion construct. As you can see, several temperament constructs correlate with the criteria. Note that there are three

additional predictor constructs. These three, "Achievement," "Masculinity," and "Locus of Control," were all a part of the miscellaneous category. When we summarized the validities for the miscellaneous category, we found respectable validities there too, so we looked more closely at the scales included in the miscellaneous category and found these additional three constructs.

The results in this table are different from the results that Guion and Gottier obtained. We believe that our strategy of summarizing the validities according to both predictor and criterion constructs accounts for the difference in results. To test this hypothesis, we summarized the validity coefficients in our database without regard to construct and obtained a coefficient of essentially zero, quite different from the coefficients in Table 2. We believe this demonstrates the importance of constructs as organizing principles for examining and understanding the literature on the criterion-related validity of temperament variables. We used the results in this table to guide us in selecting predictor constructs to measure.

Development of Temperament Scales

The next step in our research strategy was to develop measures of the constructs that the literature review indicated were likely to predict criteria important to the Army. List 1 of your handout shows the substantive scales we developed for each construct. We developed measures for six constructs: Surgency, Adjustment, Agreeableness, Dependability, Achievement, and Locus of Control. We also developed a "Physical Condition" scale and four response validity scales: Non-Random Response, Social Desirability, Poor Impression, and Self-Knowledge. We developed the Non-Random Response scale to detect inven-

tories that had been completed carelessly, a "Social Desirability" scale to detect intentional distortion that might occur in an applicant setting or a non-draft setting, and a "Poor Impression" scale to detect intentional distortion that might occur in a draft setting. We called the inventory the ABLE, short for Assessment of Background and Life Experiences.

We revised the items and scales in the ABLE many times. People representing a variety of perspectives reviewed the items for sensitive content. We also pretested the scales three times, each time evaluating and revising the items and scales based on soldiers' verbal feedback, item response distributions, internal consistency estimates, and test-retest reliabilities. The scale statistics for the ABLE scales appear in Table 3 of your handout. The average number of items in a scale is 15. The median alpha of the substantive scales is .81, and the median test-retest reliability of the substantive scales is .78. Table 4 summarizes the ABLE substantive scale statistics as well as correlations of the ABLE substantive scales with each other and with other components of the four-hour predictor battery. The only part of the predictor battery that the ABLE substantive scales correlate with in any sizable way are other ABLE substantive scales. The ABLE substantive scales appear to be tapping a part of the predictor domain not tapped by other measures.

Demonstration of Job-Relatedness

The next step in our research strategy was to demonstrate the job-relatedness of our temperament scales. We conducted a concurrent validity study during the summer and fall of 1985. Over 9000 soldiers completed the 4-hour predictor battery that included measures of cognitive

ability, spatial ability, perceptual psychomotor ability, work environment preferences, interests, and temperament.

Criterion-related validities. The criterion measures, the development of which was a major part of the research project, were developed by a different part of the research team. The criterion composites are very briefly described in List 2 of your handout. There are five composites: Core Technical Proficiency, General Soldiering Proficiency, Effort and Leadership, Personal Discipline, and Physical Fitness and Military Bearing. The first two consist mainly of work samples and knowledge tests. The other three consist of supervisory and peer ratings and information obtained from personnel records.

Table 6 of your handout shows the criterion-related validities of the ABLE scales for these five criteria. The results suggest that Achievement scales are the best predictors of the "Effort and Leadership" criterion; Dependability scales are the best predictors of the "Personal Discipline" criterion; and Physical Condition is the best predictor of the "Physical Fitness and Military Bearing" criterion, though the Achievement scales also correlate with this criterion. These three criteria include the supervisory and peer ratings. The other two criteria Core Technical Proficiency and General Soldiering Proficiency, which consist of work sample and knowledge tests, are not predicted with the ABLE substantive scales.

Table 7 in your handout shows the criterion-related validities of the different types of predictors included in the study. It shows the multiple correlations of each type of predictor with each of the five criteria. As you can see, the best predictors of the supervisory and peer rating criteria, that is, Effort and Leadership, Personal Discipline, and Physical Fitness and Military Bearing, are the ABLE substan-

tive scales. The other conclusion from this table is that the ASVAB mental ability test and the ABLE temperament inventory are the two best predictors of the criterion domain.

Fairness

We next turned to the issue of fairness. Are the items and scales fair for groups protected under the 1964 Civil Rights Act? The mean scores for whites, blacks, and Hispanics appear in Table 8 of your handout. As you can see, minorities do not tend to score lower than whites on the ABLE scales. Our efforts to write items that were not biased against minorities appear to have been successful. We're currently conducting differential validity and fairness analyses; those analyses, however are not yet complete.

Examination of Effects of Motivational Set

The fourth component of our research strategy involved investigating several issues related to motivational set. A frequent criticism of self-report inventories is that respondents can intentionally distort their responses. When respondents are applicants, this is an especially important criticism because the criterion-related validities might be negatively affected by distorted responses. We therefore studied the impact of motivational set on criterion-related validities, the extent to which applicants distort their self descriptions, and the usefulness of the four response validity scales to detect and adjust for motivational set.

Faking study. First, we conducted an experiment in which soldiers were instructed to respond honestly or to distort their responses in a specified way. The participants in the experiment were 245 enlisted

soldiers at Ft. Bragg. The design was a repeated measures with faking and honest conditions counter-balanced. We performed a multivariate analysis of variance on the ABLE scales and found that soldiers can distort their responses when instructed to do so.

We then examined the extent to which the response validity scales detected intentional distortion. Table 9 of your handout shows the results. The last two columns show the effect size of the difference between honest and fake good and honest and fake bad. Effect size can be interpreted in standard deviation terms. Thus, the difference in the honest and fake good condition for Social Desirability is essentially one standard deviation; the Social Desirability scale detects distortion in the fake good condition. As you can see, the Non-Random Response, Poor Impression, and Self-Knowledge scales detect distortion in the fake bad condition.

We next examined the extent to which we could use the response validity scales Social Desirability and Poor Impression to adjust ABLE substantive scales for faking. Table 10 shows the effect of regressing out Social Desirability in the fake good condition and the effect of regressing out Poor Impression in the fake bad condition. Median values are reported in this table. The .49 in the upper left-hand cell indicates that the median difference in ABLE scores between the honest and fake good condition before regressing out Social Desirability is .49 or half a standard deviation. That is, ABLE scale scores differ by about half a standard deviation in the fake good condition as compared to the honest condition. The next number to the right shows that after regressing out Social Desirability from the fake good condition, the ABLE substantive scales differ from the honest condition by only .14 or just over one-tenth of a standard deviation.

The next two values to the right show the results for the honest and fake bad conditions. Clearly, the Social Desirability and Poor Impression scales can be used to adjust substantive scale scores for intentional distortion.

These data demonstrate that: (1) people can distort their responses to temperament scales, (2) response validity scales can detect such distortion, and (3) the response validity scales can be used to adjust temperament scale scores for distortion.

We then asked, to what extent do applicants distort their responses? To answer this question, we compared scale scores of 121 Army applicants with scale scores of two groups of soldiers who had no motive for distorting their responses. Table 11 shows the results. On the substantive scales, applicants actually scored lower than one or both groups of soldiers 9 out of 11 times. These data suggest that applicants do not appear to distort their responses.

Nevertheless, we examined the effects of inaccurate self descriptions, as detected by the response validity scales, on criterion-related validities obtained in the concurrent validity study. Table 12 shows that validities for the group detected as responding in a random way are significantly lower than validities for the group responding conscientiously. Table 13 shows the increment in validity when Social Desirability is used as a moderator variable. Table 14 shows the increment in validity when Poor Impression is used with each substantive scale in a multiple correlation. The data in these three tables indicate that the response validity scales do improve, modestly, the validities of the substantive scales even in a concurrent validity study where there is little motive to distort one's self description.

Project A researchers are currently conducting a predictive validity study which will provide an opportunity to evaluate the validities of the ABLE substantive scales and the usefulness of the response validity scales in a selection situation.

Summary

We overcame objections to the use of temperament variables in selection by:

1. reviewing the literature using a construct-based approach to identify useful temperament constructs in previous criterion-related validity studies;

2. focusing scale development on constructs that are likely to predict criteria important to the client;

3. developing scales that consist of items acceptable to the public;

4. developing scales that are not biased against minorities;

5. developing scales that are psychometrically good;

6. developing response validity scales to detect inaccurate self descriptions;

7. evaluating job-relatedness of scales by demonstrating criterion-related validity;

8. developing and evaluating "adjustments" to substantive scale scores based on response validity scale scores, and;

9. evaluating the effect of motivational set on scale scores and criterion-related validities.

REFERENCES

Gui . M., & Gottier, R. F. (1966). Validity of personality measures in personnel selection. Personnel Psychology, 18, 135-164.

Mischel, W. (1968). Personality and assessment. New York: Wiley.

HAND-OUT

Overcoming Objections to Use of Temperament Variables in Selection: Demonstrating Their Usefulness

by Leaetta M. Hough

Personnel Decisions Research Institute

American Psychological Association Convention

New York, August 1987

A major part of this research was funded by the Army Research Institute, Contract No. MDA-903-82-C-0531. All statements expressed in this paper are those of the author and do not necessarily express the official opinions of the U.S. Army Research Institute or the Department of the Army.

RESEARCH STRATEGY: CONSTRUCT ORIENTATION

1. Review Literature
   - o Develop predictor taxonomy
   - o Classify temperament scales
   - o Develop criterion taxonomy
   - o Summarize criterion-related validities according to predictor and criterion constructs
   - o Identify useful predictor constructs
2. Develop Temperament Scales
   - o Examine items for sensitive content
   - o Develop response validity scales to detect intentional distortion
   - o Pretest
   - o Examine psychometric characteristics
   - o Revise
3. Demonstrate Job-Relatedness
   - o Conduct concurrent validity study
   - o Compute criterion-related validities
   - o Conduct differential validity analyses
   - o Conduct fairness analyses
   - o Conduct predictive validity study
4. Examine Effects of Motivational Set
   - o Evaluate fakability of scales
   - o Evaluate response validity scales
   - o Evaluate moderator effects of response validity scales
   - o Develop "adjustment" formula
   - o Assess effects on criterion-related validities

Table 1 Mean Within-Category and Between Category Correlations of Temperament Scales

	Surgency	Adjustment	Agreeableness	Dependability	Intellectance	Affiliation	Miscellaneous
Surgency	Mean$_r$=(.46) SD$_r$=.16 N$_r$=146						
Adjustment	Mean$_r$=.20 SD$_r$=.18 N$_r$=321	Mean$_r$=(.43) SD$_r$=.19 N$_r$=165					
Agreeableness	Mean$_r$=.04 SD$_r$=.17 N$_r$=173	Mean$_r$=.24 SD$_r$=.16 N$_r$=162	Mean$_r$=(.37) SD$_r$=.14 N$_r$= 44				
Dependability	Mean$_r$=-.08 SD$_r$= .16 N$_r$= 286	Mean$_r$=.13 SD$_r$=.20 N$_r$=276	Mean$_r$=.06 SD$_r$=.17 N$_r$=166	Mean$_r$=(.34) SD$_r$=.18 N$_r$=121			
Intellectance	Mean$_r$=.12 SD$_r$=.15 N$_r$=175	Mean$_r$=.02 SD$_r$=.14 N$_r$=193	Mean$_r$=.04 SD$_r$=.16 N$_r$= 94	Mean$_r$=-.12 SD$_r$= .18 N$_r$= 162	Mean$_r$=(.40) SD$_r$=.19 N$_r$= 52		
Affiliation	Mean$_r$=.09 SD$_r$=.21 N$_r$=157	Mean$_r$=.00 SD$_r$=.16 N$_r$=150	Mean$_r$=.10 SD$_r$=.17 N$_r$= 98	Mean$_r$=.08 SD$_r$=.14 N$_r$=160	Mean$_r$=-.14 SD$_r$= .15 N$_r$= 84	Mean$_r$=(.33) SD$_r$=.16 N$_r$= 45	
Miscellaneous	Mean$_r$=.09 SD$_r$=.17 N$_r$=392	Mean$_r$=.12 SD$_r$=.18 N$_r$=419	Mean$_r$=.02 SD$_r$=.18 N$_r$=215	Mean$_r$=.02 SD$_r$=.18 N$_r$=361	Mean$_r$=.04 SD$_r$=.17 N$_r$=242	Mean$_r$=-.04 SD$_r$= .15 N$_r$= 208	Mean$_r$=(.05) SD$_r$=.20 N$_r$=246

Table 2 Meta Analysis of Criterion-Related Validity Studies [1] That Used Temperament Predictors

	Criterion											
	Educational		Training		Job Involvement		Job Proficiency		Negative Adjustment: Delinquency		Negative Adjustment: Substance Abuse	
Predictor Construct[2]	Number Predictors	mean r	Number Predictors	mean r	Number Predictors	mean r	Number Predictors	mean r	Number Predictors	mean r	Number Predictors	mean r
*Surgency	42	.15	47	.08	21	.04	175	.04	8	[-.29]	30	.06
Affiliation	5	-.04	0	---	4	.06	16	-.01	0	---	4	-.03
*Adjustment	44	[.26]	44	[.16]	21	[.13]	146	[.13]	10	[-.43]	31	-.07
*Agreeableness	9	.01	5	.10	4	.02	48	-.01	1	[-.31]	8	-.04
*Dependability	24	[.15]	26	.11	18	[.17]	102	[.13]	10	[-.27]	25	[-.28]
*Intellectance	6	[.18]	7	.14	8	-.10	32	.01	1	[-.24]	2	[.18]
Achievement	8	[.30]	4	[.33]	4	[.24]	0	---	4	[-.35]	0	---
Masculinity	8	[-.16]	3	.09	10	.10	0	---	3	.02	8	[-.18]
Locus of Contro'	1	.32	2	[.29]	7	[.25]	0	---	0	---	0	---

[1] Time Period 1960-1984.

[2] A star denotes the construct is one of the "Big Five" constructs.

Note: Correlations are not corrected for unreliability or range restrictions.

List 1

ABLE[1] Scales Organized According to Construct Intended to Measure

SUBSTANTIVE SCALES:

Surgency
- Dominance
- Energy Level

Adjustment
- Emotional Stability

Agreeableness (Likeability)
- Cooperativeness

Dependability
- Nondelinquency
- Traditional Values
- Conscientiousness

Achievement
- Work Orientation
- Self Esteem

Locus of Control
- Internal Control

Physical Condition
- Physical Condition

RESPONSE VALIDITY SCALES:

- Non-Random Response
- Social Desirability
- Poor Impression
- Self-Knowledge

[1] Inventory developed by PDRI for the Army Research Institute entitled "Assessment of Background and Life Experience."

Table 3 ABLE Scale Statistics for Total Group[1]
(Concurrent Sample; Revised Trial Battery)

ABLE SUBSTANTIVE SCALES	No. Items	N	Mean	S.D.	Internal Consistency Reliability (Alpha)	Test-Retest Reliability[2]
Emotional Stability	17	8522	39.0	5.45	.81	.74
Self-Esteem	12	8472	28.4	3.70	.74	.78
Cooperativeness	18	8494	41.9	5.28	.81	.76
Conscientiousness	15	8504	35.1	4.31	.72	.74
Nondelinquency	20	8482	44.2	5.91	.81	.80
Traditional Values	11	8461	26.6	3.72	.69	.74
Work Orientation	19	8498	42.9	6.06	.84	.78
Internal Control	16	8485	38.0	5.11	.78	.69
Energy Level	21	8488	48.4	5.97	.82	.78
Dominance	12	8477	27.0	4.28	.80	.79
Physical Condition	6	8500	14.0	3.04	.84	.85
ABLE RESPONSE VALIDITY SCALES						
Social Desirability	11	8511	15.5	3.04	.63	.63
Self-Knowledge	11	8508	25.4	3.33	.65	.64
Non-Random Response[3]	8	9188	7.4	1.19	—	.30
Poor Impression	23	8492	1.5	1.85	.63	.61

[1] Total group after screening for missing data and random responding.
[2] N = 408 - 412 for test-retest correlations (N = 414 for Non-Random Response test-retest correlations).
[3] Screened only for missing data.

Table 4 ABLE Substantive Scales: Summary

(Revised Trial Battery)

	Range	Median
Reliability:		
Internal Consistency (Alpha)	.69 - .84	.81
Test-Retest	.69 - .85	.78
Relationship to Predictor Variables:		
Correlation ABLE Substantive Scales	.00 - .73	.30
Correlation Interest Scales	.00 - .43	.09
Correlation Preferred Work Environment Scales	.00 - .35	.13
Correlation Perceptual/Psychomotor Measures	.00 - .13	.03
Correlation Cognitive Measures	.00 - .20	.05
ASVAB[1] Adj. R^2	.01 - .04	.01

[1] Mental ability test currently used by military.

List 2

Criterion Composites[1]

Core Technical Proficiency - a) hands-on tests of MOS-specific technical knowledge and skills; and b) tests of school and job knowledge.

General Soldiering Proficiency - a) hands-on tests of general soldiering skill; and b) general soldiering knowledge and skill test items.

Effort & Leadership - a) supervisory and peer ratings of effort and leadership, overall effectiveness, MOS effectiveness and predicted combat effectiveness; and b) letters and certificates of commendation and other achievements.

Personal Discipline - a) supervisory and peer ratings of personal control and discipline; and b) disciplinary actions and other negative indicators in personnel files.

Physical Fitness & Military Bearing - a) supervisory and peer ratings of physical fitness and military bearing; and b) physical readiness tests.

[1]Data gathered at same time as Trial Battery was administered, i.e., summer and fall of 1985.

Table 6 Validities of ABLE Scales for Job Performance Criteria: Zero-Order Correlations (Revised Trial Battery; Concurrent Validity Study)

Predictor	Criterion: Core Technical Proficiency	General Soldiering Proficiency	Effort & Leadership	Personal Discipline	Physical Fitness & Military Bearing
Surgency:					
• Dominance	.01	.01	.15	.02	.18
Achievement:					
• Self Esteem	.02	.01	[.20]	.13	[.20]
• Work Orientation	.02	.02	[.23]	.18	[.21]
• Energy Level	.02	.02	[.22]	.14	[.25]
Adjustment:					
• Emotional Stability	.02	.02	.17	.12	.16
Agreeableness (Likeability)					
• Cooperativeness	.01	.02	.15	[.21]	.14
Dependability:					
• Traditional Values	.03	.06	.13	[.25]	.16
• Non-delinquency	.05	.07	.12	[.29]	.14
• Conscientiousness	.02	.02	.18	[.23]	.22
Others:					
• Internal Control	.04	.05	.13	.13	.13
• Physical Condition	-.04	-.05	.09	-.03	[.29]
Response Validity Scales:					
• Non-Random Response[1]	.13	.14	.07	.10	.02
• Social Desirability	-.07	-.06	.02	.05	.07
• Poor Impression	-.04	-.05	-.15	-.15	-.16
• Self-Knowledge	-.04	-.03	.07	.05	.13

[1]Correlations are based on unscreened data for this scale. N varies from 8424 to 9322 for this scale.

Note: N varies from 7666 to 8477.

Note: A box indicates notable predictor/criterion construct relationships.

Table 7

Multiple Correlations[1] of Six Independent Predictor Composites with each of Five Job Performance Criteria

(Concurrent Validity Study)

Predictor Composites	Criterion Composites: Core Technical Proficiency	General Soldiering Proficiency	Effort & Leadership	Personal Discipline	Physical Fitness & Military Bearing
ASVAB[2] (mental ability test)	[.62]	[.64]	[.35]	.20	.14
Spatial Abilities	.56	.62	.26	.14	.11
Perceptual/Psychomotor Abilities (computerized)	.54	.58	.30	.12	.10
Work Environment Preferences	.28	.27	.20	.10	.11
Temperament (and physical activities scale)	.26	.24	[.34]	[.33]	[.36]
Interests	.34	.34	.26	.14	.13

[1]Multiple Rs are adjusted for shrinkage and corrected for restriction in range, but not corrected for criterion unreliability.

[2]Mental ability test currently used by military.

Note: Entries in table are averaged across 9 Army military occupational specialties (MOS) with complete criterion data. Total sample is 3902. Sample sizes range from 281 to 570; median = 432.

Note: Boxes denote the two best predictors of the criterion space.

Table 8

ABLE Scale Means and Standard Deviations Separately for Race (Trial Battery) (Revised)

	Black (N = 2227 - 2256)		Hispanic (N = 284 - 292)		White (N = 5614 - 5673)		Other (N = 328 - 332)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
ABLE Substantive Scales								
Emotional Stability	39.3	4.97	38.7	5.25	38.9	5.63	38.2	5.47
Self-Esteem	28.7	3.32	28.7	3.49	28.4	3.83	27.8	4.02
Cooperativeness	42.6	5.02	41.9	4.92	41.6	5.38	41.6	5.18
Conscientiousness	35.7	3.68	36.1	4.08	34.7	4.53	35.7	3.80
Nondelinquency	45.4	5.18	45.0	5.96	43.7	6.11	44.8	5.93
Traditional Values	27.2	3.11	27.0	3.16	26.3	3.95	26.7	3.42
Work Orientation	43.1	5.51	43.5	5.44	42.8	6.31	43.2	5.80
Internal Control	37.8	4.55	38.2	4.50	38.1	5.37	38.4	4.54
Energy Level	48.6	5.35	49.6	5.49	48.3	6.21	48.2	5.92
Dominance	27.7	3.86	27.3	4.09	26.8	4.42	26.5	4.15
Physical Condition	14.4	2.84	14.0	3.11	13.8	3.10	13.6	3.09
ABLE Response Validity Scales								
Social Desirability	15.8	3.05	17.4	3.60	15.2	2.91	17.0	3.50
Self-Knowledge	26.2	3.10	25.4	3.12	25.1	3.39	25.5	3.11
Non-Random Response	7.6	0.65	7.6	0.68	7.7	0.54	7.6	0.62
Poor Impression	1.4	1.66	1.4	1.57	1.5	1.94	1.6	1.91

Note: A box indicates a difference from the white mean of approximately one-half standard deviation or more.

Table 9

ABLE Response Validity Scales:
Effects of Honest* and Faking* Conditions
Ft. Bragg

ABLE Response Validity Scale	Honest First*			Fake Good First*			Fake Bad First*			Effect Size Honest vs. Fake Good	Effect Size Honest vs. Fake Bad
	N	M	S.D.	N	M	S.D.	N	M	S.D.		
Social Desirability (Unlikely Virtues)	109	15.8	3.1	57	20.1	5.8	56	17.8	4.8	-1.02	- .53
Self-Knowledge	109	29.6	3.6	57	29.7	4.1	56	21.8	5.2	- .03	1.85
Non-Random Response	109	7.6	1.0	57	7.0	1.8	56	2.8	2.2	.45	3.16
Poor Impression	109	1.5	2.1	57	1.7	2.2	56	14.6	7.9	- .09	-2.67

*Values are based on the sample that completed the questionnaires under the condition of interest first.

TABLE 10

Effects of Regressing Out Response Validity Scales
(Social Desirability and Poor Impression)
in Faking Conditions for ABLE

	Honest vs. Fake Good Effect Size		Honest vs. Fake Bad Effect Size	
	Before Adjustment	After Adjustment	Before Adjustment	After Adjustment
ABLE Substantive Scales	.49	.14	2.10	.45

Table 11

Comparison of Ft. Bragg Honest*, Ft. Knox, and MEPS (Applicants) ABLE Scales

ABLE Scale	Ft. Bragg (Honest)* N	Ft. Bragg (Honest)* Mean	MEPS (Applicants) N	MEPS (Applicants) Mean	Ft. Knox N	Ft. Knox Mean	Total S.D.
Response Validity Scales							
Social Desirability	116	15.91	121	16.63	276	16.60	3.21
Self-Knowledge	116	29.54	121	28.03	276	29.64	3.63
Non-Random Response	116	7.58	121	7.79	276	7.75	.64
Poor Impression	116	1.50	121	1.05	276	1.54	1.84
Substantive Scales							
Emotional Stability	112	66.22	118	66.03	272	65.05	7.86
Self-Esteem	112	34.77	118	34.04	272	35.12	5.00
Cooperativeness	112	53.33	118	54.60	272	54.19	6.05
Conscientiousness	112	46.37	118	46.49	272	48.97	5.86
Non-Delinquency	112	53.24	118	54.36	272	55.49	6.91
Traditional Values	112	36.67	118	36.97	272	37.28	4.50
Work Orientation	112	59.71	118	58.37	272	61.40	7.73
Internal Control	112	49.48	118	51.90	272	50.37	6.13
Energy Level	112	57.56	118	56.67	272	57.19	6.95
Dominance	112	35.54	118	32.84	272	35.41	6.05
Physical Condition	112	32.96	118	28.27	272	31.08	7.49

*Scores are based on persons who responded to the honest condition first.

Table 12. Moderating[1] Effects of Random Responding on Correlations Between ABLE Scales and Job Performance Criteria

ABLE SCALE	CRITERION: Effort/Leadership Low (Random)	Effort/Leadership High (Non-Random)	Personal Discipline Low (Random)	Personal Discipline High (Non-Random)	Physical Fitness/Bearing Low (Random)	Physical Fitness/Bearing High (Non-Random)
Surgency:						
Dominance	.06	.15	.05	.02	.18	.18
Achievement:						
Self-Esteem	-.00	.15	.03	.09	.08	.18
Work Orientation	.05	.23	.08	.18	.12	.21
Energy Level	.07	.22	.10	.14	.20	.25
Adjustment:						
Emotional Stability	.11	.17	.08	.12	.09	.16
Agreeableness:						
Cooperativeness	.13	.15	.17	.21	.10	.14
Dependability:						
Traditional Values	.07	.13	.19	.25	.18	.16
Nondelinquency	.09	.12	.22	.29	.14	.14
Conscientiousness	.05	.18	.11	.23	.16	.22
Others:						
Internal Control	.00	.13	.03	.13	.05	.13
Physical Condition	-.03	.09	-.00	-.03	.16	.29

N ranges from 659 to 675 for group scoring low on "Non-Random Response" scale
N ranges from 8336 to 8477 for group scoring high on "Non-Random Response" scale
Note: Statistically significant differences at $P \leq .05$ is approximately .04.

[1] We performed a split group analysis rather than a moderated regression because the variable of interest had a highly skewed distribution.

Table 13. Moderating[1] Effects of "Social Desirability" Scale on Correlations Between ABLE Scales and Job Performance Criteria

ABLE SCALE	CRITERION: Effort/Leadership Non-High[2]	Effort/Leadership High[3]	Personal Discipline Non-High[2]	Personal Discipline High[3]	Physical Fitness/Bearing Non-High[2]	Physical Fitness/Bearing High[3]
Surgency:						
Dominance	.15	.14	.00	.06	.18	.17
Achievement:						
Self-Esteem	.21	.18	.12	.12	.21	.17
Work Orientation	.25	.20	.17	.16	.22	.17
Energy Level	.23	.20	.13	.15	.27	.20
Adjustment:						
Emotional Stability	.17	.16	.11	.12	.16	.13
Agreeableness:						
Cooperativeness	.16	.13	.20	.21	.14	.12
Dependability:						
Traditional Values	.14	.11	.26	.22	.18	.11
Nondelinquency	.13	.12	.28	.29	.14	.11
Conscientiousness	.19	.14	.22	.22	.24	.14
Others:						
Internal Control	.13	.12	.12	.15	.15	.08
Physical Condition	.08	.09	-.03	-.02	.28	.29

[1] We performed a split group analysis rather than a moderated regression because the variable of interest had a highly skewed distribution.

[2] N ranges from 5896 to 5997 for group scoring Non-High on "Social Desirability" scale

[3] N ranges from 2428 to 2480 for group scoring high on "Social Desirability" scale

Note: A statistically significant difference at $p \leq .05$ is approximately .03

Table 14. Incremental Validities of ABLE Scales When "Poor Impression" Scale is Included in Predictor Equation (Linear Model)[1]

ABLE SCALE	CRITERION: Effort/Leadership r	Effort/Leadership R	Personal Discipline r	Personal Discipline R	Physical Fitness/Bearing r	Physical Fitness/Bearing R
Surgency:						
Dominance	.15	.19	.02	.15	.18	.22
Achievement:						
Self-Esteem	.20	.22	.12	.17	.20	.22
Work Orientation	.23	.25	.18	.20	.21	.23
Energy Level	.22	.22	.14	.17	.25	.26
Adjustment:						
Emotional Stability	.17	.18	.12	.16	.16	.18
Agreeableness:						
Cooperativeness	.15	.18	.21	.22	.14	.17
Dependability:						
Traditional Values	.14	.17	.25	.26	.17	.20
Nondelinquency	.13	.17	.29	.29	.14	.18
Conscientiousness	.18	.20	.23	.24	.22	.23
Others:						
Internal Control	.13	.17	.13	.17	.13	.17
Physical Condition	.09	.16	.03	.16	.29	.31

N = 8400

Note: A statistically significant difference at $p \leq .05$ is approximately .02

[1] A linear model was used because the zero-order correlations of the "Poor Impression" scale with the criteria are approximately -.15.