AL/HR-TP-1994-0009

AD-A280 155

2

BUILDING A JOINT-SERVICE CLASSIFICATION RESEARCH ROADMAP: INDIVIDUAL DIFFERENCES MEASUREMENT

ARMSTRONG LABORATORY

DTIC ELECTE JUN 01 1994 S B D

Teresa L. Russell
Douglas H. Reynolds
John P. Campbell

Human Resources Research Organization (HumRRO)
66 Canal Center Plaza, Suite 400
Alexandria, VA 22314

HUMAN RESOURCES DIRECTORATE
MANPOWER AND PERSONNEL RESEARCH DIVISION
7909 Lindbergh Drive
Brooks Air Force Base, TX 78235-5352

April 1994

Final Technical Paper for Period January 1993 - August 1993

Approved for public release; distribution is unlimited.

94-16167

405 260

94 5 31 016

AIR FORCE MATERIEL COMMAND
BROOKS AIR FORCE BASE, TEXAS

NOTICE

This technical paper is published as received and has not been edited by the technical editing staff of the Armstrong Laboratory.

Publication of this paper does not constitute approval or disapproval of the ideas or findings. It is published in the interest of scientific and technical information (STINFO) exchange.

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Office of Public Affairs has reviewed this paper, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This paper has been reviewed and is approved for publication.

MALCOLM JAMES REE
Scientific Advisor

PATRICK C. KYLLONEN
Technical Director
Manpower & Personnel Research Div

WILLARD BEAVERS, Lt Colonel, USAF
Chief, Manpower & Personnel Research Div

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	April 1994	Final – January - August 1993

4. TITLE AND SUBTITLE
Building a Joint-Service Classification Research Roadmap: Individual Differences Measurement

5. FUNDING NUMBERS
C – F33615-91-C-0015
PE – 62205F
PR – 7719
TA – 25
WU – 01

6. AUTHOR(S)
Teresa L. Russell
Douglas H. Reynolds
John P. Campbell

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Human Resources Research Organization (HumRRO)
66 Canal Center Plaza, Suite 400
Alexandria VA 22314

8. PERFORMING ORGANIZATION REPORT NUMBER
FR-PRD-93-24

9. SPONSORING/MONITORING AGENCY NAMES(S) AND ADDRESS(ES)
Armstrong Laboratory (AFMC)
Human Resources Directorate
Manpower and Personnel Research Division
7909 Lindbergh Drive
Brooks Air Force Base, TX 78235-5352

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AL/HR-TP-1994-0009

11. SUPPLEMENTARY NOTES

Armstrong Laboratory Technical Monitor: Dr. Malcolm J. Ree, (210) 536-3942.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The Armstrong Laboratory, the Army Research Institute for the Behavioral and Social Sciences, the Navy Personnel Research and Development Center, and the Center for Naval Analyses are committed to enhancing the overall efficiency of the Services' selection and classification research. This means reducing redundancy of research across Services and improving inter-service research planning, while ensuring that each Service's priority needs are served. The Roadmap project products describe across-service military classification research issues. This report serves two purposes. First, it is a reference document that selection and classification experts in the Services can use in making decisions about predictor measures. It provides information about operational and experimental predictors. Second, it refines and supplements predictor-related research objectives that emerged from the Services' current selection and classification practices and interviewing military selection and classification experts to identify selection and classification research objectives.

14. SUBJECT TERMS
Classification
Individual differences
Measurement
Roadmap

15. NUMBER OF PAGES 164

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std Z39-18
298-102

BUILDING A JOINT-SERVICE CLASSIFICATION RESEARCH ROADMAP: INDIVIDUAL DIFFERENCES MEASUREMENT

Table of Contents

Page

List of Tables

List of Figures

PREFACE

This technical paper documents research and development performed by the Human Resources Research Organization (HumRRO) for the Armstrong Laboratory, Human Resources Directorate, under Contract No. F33615-91-C-0015, JON 7719 2403. It is the third in a series of six reports to be delivered under this contract.

The Roadmap project products describe across-Service military classification research issues. The key to the success of this effort has been the participation of experts from the Services. We thank the representatives of the Armstrong Laboratory, the Army Research Institute, the Navy Personnel Research and Development Center, the Center for Naval Analyses, the Defense Manpower Data Center, and the Office of the Assistant Secretary of Defense, who answered our questions. They were helpful and supportive. We are especially grateful to Dr. Malcolm Ree, who, as technical monitor of this contract, provided advice and technical support.

SUMMARY

The Armstrong Laboratory, the Army Research Institute for the Behavioral and Social Sciences, the Navy Personnel Research and Development Center, and the Center for Naval Analyses are committed to enhancing the overall efficiency of the Services' selection and classification research. This means reducing redundancy of research across Services and improving inter-Service research planning, while ensuring that each Service's priority needs are served. With these goals in mind, the Armstrong Laboratory and the Army Research Institute co-sponsored a project to develop a Joint-Service classification research agenda, or Roadmap.

The Roadmap Project has six tasks. The first three tasks have been completed. Task 1 involved documenting the Services' current selection and classification practices and interviewing military selection and classification (S&C) experts to identify S&C research objectives. (Russell, Knapp, & Campbell, 1992). Task 3 which has also been completed was a review and analysis of job analysis methods and procedures (Knapp, Russell, & Campbell, 1992). The current report documents the results of Task 2, a review and analysis of predictor measures.

This report serves two purposes. First it is a reference document that S&C experts in the Services can use in making decisions about predictor measures. It provides information about operational and experimental predictors. Second, it refines and supplements predictor-related research objectives that emerged from Task 1.

BUILDING A JOINT-SERVICE CLASSIFICATION RESEARCH ROADMAP: INDIVIDUAL DIFFERENCES MEASUREMENT

I. INTRODUCTION

Teresa L. Russell

Overview of the Roadmap Project

The Armstrong Laboratory, the Army Research Institute for the Behavioral and Social Sciences, the Navy Personnel Research and Development Center, and the Center for Naval Analyses are committed to enhancing the overall efficiency of the Services' selection and classification research. This means reducing redundancy of research efforts across Services and improving inter-Service research planning, while ensuring that each Service's priority needs are served. With these goals in mind, the Armstrong Laboratory and the Army Research Institute co-sponsored a project to develop a Joint-Service classification research agenda, or Roadmap.

The roadmap project plan has six tasks:

Task 1.	Identify Classification Research Objectives,
Task 2.	Review Classification Tests and Make Recommendations,
Task 3.	Review Job Requirements and Make Recommendations,
Task 4.	Review Criteria and Make Criterion Development Recommendations,
Task 5.	Review and Recommend Statistical and Validation Methodologies, and
Task 6.	Prepare Roadmap.

The first task, Identify Classification Research Objectives is reported in Russell, Knapp, and Campbell (1992). It involved interviewing selection and classification experts and decision-makers from each Service to determine research objectives. Tasks 2 through 5 are systematic reviews of specific predictor, job analytic, criterion, and methodological needs of each of the Services. The final task, Prepare Roadmap, will integrate the findings of Tasks 1 through 5 into a master research plan. The Roadmap is planned to be completed early in 1993.

The third task, Review Job Requirements and Make Recommendations, was reported in Knapp, Russell, and Campbell (1992). The fourth task was reported in Knapp & Campbell (1992), and the fifth task was reported in Campbell (1992). This report fulfills the requirements of Task 2, Review Classification Tests and Make Recommendations.

Predictor-Related Research Objectives

Task 1 yielded a set of research objectives and information about military selection and classification experts' perceptions of the importance and urgency of those objectives. The objectives related to the predictors are described below.

Determine which existing (but not implemented) predictors are most useful for classification purposes (Objective #7).

In the late 1980s, the Services identified nine tests that might be good candidates for inclusion in the Armed Services Vocational Aptitude Battery (ASVAB) or for supplementing the ASVAB. The nine tests measure psychomotor ability, spatial ability, and working memory and are computerized; the battery is called ECAT for Enhanced Computer Administered Tests. NPRDC is the lead organization for ECAT development as well as for the development of the computer adaptive form of the ASVAB (CAT-ASVAB).

The Defense Manpower Data Center (DMDC) is planning to make changes to the ASVAB by the late 1990s. Before then, the Services will determine which ECAT tests (if any) and other experimental tests should indeed be incorporated into the ASVAB. The ASVAB Review Technical (ART) committee has developed a list of criteria against which the tests should be considered (e.g., subgroup differences, differential validity). There are still several missing pieces of information, and ECAT data are currently being collected and analyzed.

Develop and evaluate measures of new predictors likely to be useful for classification purposes (Objective #8).

Basic individual differences research is necessary because changes in military jobs may suggest measurement of new or different attributes. The Air Force's Learning Abilities Measurement Program (LAMP) is probably one of the best known basic abilities research projects. Its goals are to denote the basic parameters of learning ability, develop techniques to assess cognitive ability, and investigate the feasibility of applying a cognitive model-based system to psychological assessment (Kyllonen, 1985a). Selected experimental LAMP measures are currently being converted into experimental tests in a separate Air Force Automated Personnel Testing (APT) project.

Both the Navy and the Army conduct basic abilities research as well. Most of the Navy's work has focused on spatial, perceptual, and reaction time measures. The Army is preparing to undertake a project called Expanding the Concept of Quality in Personnel (ECQUIP) that will involve investigation of practical intelligence and social intelligence, among others.

Evaluate alternative fairness models in terms of their effects on selection/classification outcomes across subgroups (Objective #15).
Identify and/or develop classification measures that minimize adverse impact and/or predictive bias (Objective #17).

Adverse impact is "defined as a substantially different rate of selection in hiring, promotion, or other employment decision that works to the disadvantage of members of a race, sex, or ethnic group" (American Institutes for Research, 1992). Adverse impact is not, however, proof of unfairness. Cleary's (1968) model of fairness is currently accepted by both the Uniform Guidelines (1978) and the Society for Industrial and Organizational Psychology (SIOP, 1987). This definition distinguishes between test bias and test fairness: "A test is biased for members of a subgroup of the population if, in the prediction of a criterion for which the test was designed, consistent nonzero errors of prediction are made for members of the subgroup" (Cleary, 1968, p. 115). In other words, a test is biased when prediction from a common regression equation results in either over- or under-prediction of subgroup performance. SIOP (1987) defines fairness as a social rather than a psychometric concept. Fairness is a function of how test scores are used for the job and the population at hand. For example, over-prediction of the performance of a protected group, when a common regression line is used, indicates bias but is generally not considered a fairness problem. With regard to selection and classification measures used by the military, it will be important in future research to consider levels of adverse impact in evaluating tests and to develop new measures that minimize adverse impact and predictive bias.

Goals and Scope of the Current Report

This report has two primary goals: (1) to provide/summarize predictor-related information that selection and classification experts in the Services can use to make decisions about predictor measures and (2) to refine and supplement the predictor-related objectives from Task 1.

Chapter II, Operational Predictors, summarizes information about the ASVAB and several other psychometric measures used to select officers or for other special purposes. Chapter III describes the ECAT battery and summarizes key information about cognitive, psychomotor and physical abilities predictors. The fourth chapter concentrates on personality, interest, and biodata predictors. The fifth chapter outlines basic research that appears promising for the development of novel predictor measures and provides our recommendations for predictor research directions.[1]

An Individual Differences Framework

Kyllonen (1991) pointed out the advantages of studying individual differences measures within a taxonomy. Comparing test batteries against a taxonomy facilitates identification of redundancy in measures, illuminates insufficient or incomplete coverage of a domain, and facilitates advancement of theory. With this in mind, we summarized taxonomical information from each of several individual differences domains; the result appears in Table 1. This framework will serve as a reference point for the discussion of predictor measures.

Cognitive Attributes

Several major models of the intellect have been proposed over the last century. Recent research continues to support Vernon's (1950) hierarchical structure of cognitive abilities (Ackerman, 1987). Vernon proposed that two major group factors emerge in factor analyses, after the extraction of *g*: (1) verbal-numerical *(v:ed)* and (2) practical-mechanical-spatial *(k:m)*. Minor group factors, analogous to Thurstone's (1938) primary mental abilities, are subsumed by the two major group factors. With due respect to Vernon and all of the other major contributors to cognitive abilities research (e.g., Guilford & Lacey, 1947; Spearman, 1927; Thurstone, 1938; Vernon, 1950; and Ekstrom, French, & Harman [1979] who summarized factor-analytic abilities research through the mid-1970s), we chose to summarize cognitive attributes according to Horn's (1989) framework, which is compatible with that of others (e.g., Cattell, 1971; Ekstrom et al., 1979; Toquam, Corpe, & Dunnette, 1989).

[1]We describe tests throughout this report. For each measure, we attempt to report internal consistency reliability, test-retest reliability, and information about sex and race differences in mean scores, validity, and predictive bias. In several cases, however we have been unable to obtain complete information for a test.

Table 1	
Individual Differences Attributes and Constructs	
Broad Attributes	Related Constructs
Cognitive	
G_c - Knowledge or Crystallized Intelligence	Knowledge of general information Word knowledge
G_f - Broad Reasoning or Fluid Intelligence	Inductive reasoning Deductive reasoning
G_v - Broad Visual Intelligence	Spatial visualization Spatial orientation
SAR - Short Term Acquisition and Retrieval	Recency memory Word span
TSR - Long Term Storage and Retrieval	Associational fluency Expressional fluency Ideational fluency
G_s - Broad Speediness	Visual scanning Visual matching
G_a - Auditory Intelligence	Discrimination among sound patterns Auditory cognition of relations
G_q - Quantitative Thinking	Computational fluency Numerical computation
Eng - English Adeptness	Word parsing Phonetic decoding
Psychomotor	
Dexterity	Finger dexterity Manual dexterity
Basic Movement Speed and Accuracy	Reaction time Control precision Speed of arm movement
Perceptual-Motor Movement Control	Multi-limb coordination Rate control

(Continued)

Table 1 Individual Differences Attributes and Constructs (Continued)	
Broad Attributes	Related Constructs
Physical	
Muscular Strength	Muscular tension Muscular power Muscular endurance
Cardiovascular Endurance	Cardiovascular endurance
Movement Quality	Flexibility Balance Coordination
Personality	
Extraversion	Sociable, Gregarious, Ambitious, Achievement-oriented
Emotional Stability	Emotional, Anxious, Depressed
Agreeableness	Good-natured, Cooperative
Conscientiousness	Dependable, Responsible
Intellectance	Curious, Broad-minded
Interest	
Realistic	Practical, likes hand-on work
Investigative	Curious, likes academic endeavors
Artistic	Creative, likes self-expression
Social	Friendly, likes people
Enterprising	Ambitious, likes managing & directing
Conventional	Concrete, likes exactness in work

Source: Cognitive (Horn, 1989), Psychomotor (Fleishman, 1967; Imhoff & Levine, 1981; McHenry, 1987); Physical (Hogan, 1991a); Personality (Barrick & Mount, 1991; Digman, 1990; Tett, Jackson, & Rothstein, 1991); Interests (Holland, 1983).

Horn integrated information processing research with traditional factor-analytic results and evidence from physiological studies of brain injury and other impairments to identify broad and narrow cognitive factors. Narrow (or primary) factors are ones for which the intercorrelations among sub-factors are large; broad factors (second-order) are defined by tests that are not as highly intercorrelated. He defines six broad cognitive attributes--G_c, G_f, G_v, SAR, TSR and G_s--and two other factors that are important in specific settings, G_q and Eng.

Knowledge or Crystallized Intelligence, G_c, underlies performance on knowledge or information tests. *Broad Reasoning or Fluid Intelligence*, G_f, subsumes virtually all forms of reasoning--inductive, conjunctive, deductive, and so forth. G_f abilities decline with age while G_c abilities can improve and are less likely to decline. According to Horn (1989), tests (e.g., verbal analogies) are good reasoning measures to the extent that they contain words or materials that are equally familiar, or unfamiliar, for all examinees; otherwise variance due to knowledge (e.g., word knowledge) makes such tests measures of G_c. Individual differences in performance on novel tasks, or in the early stages of learning on even simple tasks, are a function of G_c (Ackerman, 1987; Horn, 1989). When examinees become familiar with the task, other abilities become determinants of performance.

G_f is at the heart of what is typically called intelligence, and it is instrumental to accumulation of crystallized knowledge (Carroll, 1989; Horn, 1989). It is also closely related to Working Memory Capacity (WMC), the central construct involved in information-processing (Kyllonen & Christal, 1990). WMC is the ability to process and store information simultaneously on complex tasks, regardless of content (e.g., math, verbal). Kyllonen and Christal found that General Reasoning correlated .80 to .88 with WMC in four large studies that used a variety of reasoning and WMC measures. They suggest that WMC "is responsible for individual differences in reasoning ability" (p. 427). This hypothesis is compatible with Ackerman's (1988) idea that general ability reflects the availability of attentional resources. In the remainder of this review, we treat WMC as a G_f construct and discuss WMC measures alongside more traditional measures of G_f.

Broad Visual Intelligence, G_v, is Horn's (1989) broad spatial attribute, including all spatial constructs where speed is not important. Complex spatial visualization, mental rotation, and paper form board tests are measures of G_v.

Short-Term Acquisition and Retrieval, SAR, is derived from information processing research. It encompasses tasks that involve sequential processing of information in short term memory. Recency memory, for example, requires recalling the most recently presented stimuli out of a string of stimuli presented in temporal order. SAR and WMC are related but not unitary constructs (Cantor, Engle, & Hamilton, 1991). Tasks that measure SAR appear to focus on recall of information; whereas WMC tasks are more complex and may require transformation or reorganization of information in short term memory. *Long-Term Storage and Retrieval*, TSR, constructs refer to the organization of information or concepts in long-term memory and the fluency of retrieval. TSR is measured by unspeeded fluency tasks that require the individual to produce (retrieve) ideas, expressions, or words given a stimulus or given tasks that require recitation of previously learned material.

Broad Speediness, G_s, underlies performance on all types of speeded measures, including clerical or perceptual speed and visual matching tasks. According to Horn (1989), almost any task can be made into a measure of G_s by increasing speededness and decreasing knowledge and reasoning requirements. Horn also issues a caveat about the interpretation of G_s measures. Physiological, emotional, or even strategical (i.e., carefulness) differences may influence performance on speeded tests more than on other cognitive measures.

Auditory Intelligence, G_a, "represents a facility for chunking streams of sounds, keeping these chunks in awareness, and anticipating an auditory form that can develop out of such streams" (Horn, 1989, p. 84). Horn believes that G_a is an important form of intelligent behavior that has been largely overlooked in the past. Primary factors that indicate G_a involve detection, transformation, and retention of tonal patterns.

Horn (1989) separates *Quantitative Thinking*, G_q, and *English Adeptness*, ENG, from the other factors because of their importance in educational settings. However, developing measures of G_q independent of G_c and G_f may not be simple. Tests that involve reasoning that can be acquired outside courses in mathematics are likely to measure G_f; G_q tests require subjects to make advanced calculations. As mentioned, Horn distinguishes G_q from G_c and G_f because mathematical aptitude plays a significant role in educational guidance and placement decisions. Likewise, he proposes an *English Adeptness*, ENG, general factor because tests that measure it are useful for diagnosing language difficulties.

It would be unwise to leave the reader with the impression that cognitive abilities are orthogonal to each other. A wealth of evidence supports the existence of one broad general factor (*g*) underlying cognitive test scores (e.g., Jensen, 1986). *g* is "mental energy" (Spearman, 1927), the ability to learn or adapt (Hunter, 1986), not exactly the ability to learn (Tyler, 1986), the availability of attentional resources (Ackerman, 1988), or perhaps working memory capacity (Kyllonen & Christal, 1990). In any case, it is a mathematical factor that can be computed several different ways with essentially the same result (Ree & Earles, 1991a); however, *g* computed with one battery of tests is not identical to *g* from another battery of tests (Linn, 1986). *g* may be defined by elementary mental processes such as decision time on a letter recognition task (Kranzler & Jensen, 1991a, 1991b) or perhaps *g* can be predicted by scores on information processing tasks (Carroll, 1991a, 1991b). Both interpretations have been offered for the same set of data. *g* has a high degree of heritability under the control of many genes (Humphreys, 1979), but is also influenced by the environment (Jensen, 1977, 1992). It is related to educational achievement and socio-economic status in complex ways (Humphreys, 1986, 1992). *g* predicts job performance (Hunter, 1986; Ree, Earles, & Teachout, 1992; Thorndike, 1986) and training success (Ree & Earles, 1991b) and yields small positive correlations with a host of other variables (e.g., Vernon, 1990).

Horn (1989) offers a different interpretation of the observed shared variance across cognitive ability tests. Considering the complexity of inter-relationships among abilities, he suggests that tests are rarely pure measures of an ability. It may be difficult, for example, to construct tests that do not require both knowledge and reasoning or both knowledge and short term memory. Horn's view also takes the population of subjects into account. Whether a test will measure G_c or G_f depends upon the subjects' knowledge in relation to the content of the test. From this perspective, *g* is a result of the application of multiple abilities to perform test items.

Moreover, existence of *g* does not preclude the existence of specific abilities and vice versa. Almost everyone acknowledges that specific abilities have been identified and replicated. The debate surrounds the magnitude and significance of the contribution of specific abilities in predictive validity settings over that afforded by *g*. Experts disagree over the amount of increment that is worthwhile (Humphreys, 1986; Linn, 1986). Although we will revisit this question, we will not answer it in this report.

Psychomotor, Physical, Personality, and Interest Domains

Psychomotor and physical constructs are reviewed in detail in Chapter III, and personality and interest domains are discussed in Chapter IV. We provide a brief overview here.

Psychomotor

The labels for psychomotor constructs in Table 1 are derived from the classic work of Fleishman and his colleagues (e.g., Fleishman, 1954; Fleishman & Ellison, 1962; Fleishman & Hempel, 1954a, 1954b, 1955, 1956), Imhoff and Levine (1981), and McHenry (1987). The more recent works have focused on hierarchical models of psychomotor abilities--models that are compatible with Fleishman's taxonomy. In an extensive review of the psychomotor, perceptual, and cognitive ability literature, Imhoff and Levine (1981) proposed two higher-order dimensions of Fleishman's psychomotor ability factors: (1) Basic Movement Speed and Accuracy and (2) Perceptual-Motor Movement Control. Basic Movement Speed and Accuracy includes Fleishman's Control Precision, Speed of Arm Movement, and Reaction Time abilities--abilities that are highly structured and require speed and accuracy with little processing. Fleishman's Multilimb Coordination, Response Orientation, and Rate Control are subsumed by Perceptual-Motor Movement Control. This is a category of abilities that requires continuously or periodically adjusting movements in response to sensory or perceptual feedback. McHenry (1987) extended Imhoff and Levine's (1981) work, adding a third second-order dimension, Dexterity, to include manual and finger dexterity, and he posited a general factor underlying all tests of psychomotor ability.

Physical

Fleishman's (1972) taxonomy had nine physical proficiency constructs: (1) Static Strength, (2) Explosive Strength, (3) Dynamic Strength, (4) Trunk Strength, (5) Extent Flexibility, (6) Dynamic Flexibility, (7) Gross Body Coordination, (8) Gross Body Equilibrium, and (9) Stamina. Hogan (1991a) adapted and revised Fleishman's dimensions to better reflect physiological functioning and work performance. Her categories are seven-fold: (1) Muscular Tension, (2) Muscular Power, (3) Muscular Endurance, (4) Cardiovascular Endurance, (5) Flexibility, (6) Balance, and (7) Coordina-tion. In Hogan's model, Muscular Tension, Muscular Power, and Muscular Endurance are organized into a broader Muscular Strength construct.

Similarly, Flexibility, Balance, and Coordination are included in a broader Movement Quality construct. Cardiovascular Endurance has no higher-order counterpart.

Personality

Personality research has begun to converge on the number and content of replicable factors in personality instruments (Barrick & Mount, 1991; Digman, 1990; Tett, Jackson, & Rothstein, 1991). The "big five" replicable factors are: (1) Extraversion (sociable, ambitious), (2) Agreeableness (amiable, cooperative), (3) Emotional Stability (well-adjusted, calm), (4) Conscientiousness (trustworthy, persistent), and (5) Intellectance (thinking, creative).

Interest

The most widely-used occupational taxonomy, not based on cognitive requisites for jobs, is probably Holland's interest-based scheme (1983). Holland found that four to eight categories of interests subsume most scales in interest inventories and that the different interest constructs have relatively stable relationships with one another. He named the primary six interest themes--Realistic, Investigative, Artistic, Social, Enterprising, and Conventional--or RIASEC. More recent occupational interest measurement research suggests that the Holland factors form the top of an interest hierarchy with the 20 basic interest factors from the Strong-Campbell Interest Inventory constituting the next level (Campbell & Hansen, 1981).

Inter-Relationships Among Constructs

Performance on a cognitive test is largely attributable to general cognitive ability (Ree & Earles, 1991b). Remaining reliable variance may be a result of specific cognitive, psychomotor, physical, personality, and interest variables and perhaps some others (e.g., luck, according to Sternberg, in press). But, different individual differences domains are rarely studied simultaneously. One exception is the work of Snow and his colleagues (Snow, 1989). They have begun mapping relationships between cognitive, personality, and interest constructs. Snow administered the California Personality Inventory (CPI), Strong-Campbell Interest Inventory (SCII), and the Wechsler Adult Intelligence Test to Stanford students. Preliminary results suggested that relationships between personality and cognitive variables are curvilinear and that sex differences in interest and personality add complexity to the model. Cross-domain research will be addressed again in Chapter V.

II. OPERATIONAL PREDICTORS

Teresa L. Russell and Felicity A. Tagliareni

Military testing has a long and distinguished tradition, beginning during World War I. Research since then has culminated in the development of several test batteries that are currently in use or will be operational very soon. This chapter provides an overview of several operational tests: the Armed Services Vocational Aptitude Battery (ASVAB), the Air Force Officer Qualifying Test (AFOQT), the Officer Selection Battery (OSB), the Aviation Selection Test Battery (ASTB), the Basic Attributes Test (BAT) battery, the Multi-Track Test Battery, the Defense Language Aptitude Battery (DLAB), and tests for aptitudes related to intelligence jobs. The ASVAB is the Services' primary enlisted selection and classification tool. The AFOQT, OSB, ASTB, BAT and Multi-Track Test Battery are used for officer selection and classification. The DLAB is a special aptitude test used to identify enlisted personnel and officers for foreign language skills training.

The Armed Services Vocational Aptitude Battery (ASVAB)[1]

The ASVAB is a differential aptitude battery, philosophically a descendent of Thurstone's (1938) research to define primary mental abilities. The content of the ASVAB stems from modifications of the Army General Classification Test (AGCT) and the Navy General Classification Test (NGCT) that were employed during World War II (Schratz & Ree, 1989). These tests were measures of general learning ability and were designed to aid in assigning new recruits to military jobs (Eitelberg, Laurence, Waters, & Perelman, 1984). The tests resembled each other in content--covering such cognitive domains as vocabulary, mathematics, and spatial relationships (Eitelberg et al., 1984). Separate batteries were used until the late-1960s when the Services developed a joint testing program. The resulting multiple-aptitude, group-administered ASVAB is now the primary enlisted personnel selection test used by the military.

ASVAB implementation is currently directed by the Manpower Accession Policy Steering Committee. Research and development work associated with the ASVAB is led by the Testing Division of the Department of Defense (DoD) Defense Manpower Data Center and is complemented by work from the Services' research laboratories. Finally, the Military Entrance Processing Command (MEPCOM) executes the Army's responsibility for handling the system assoc-iated with ASVAB testing and score processing (Department of Defense [DoD], 1984a).

When the joint-Service project to develop the ASVAB began, the Services defined four requirements for the joint-Service measure (Eitelberg et al., 1984). First, it was to provide a global assessment of ability, covering ground previously assessed by the AGCT. Second, it was to include items that covered the topics of vocabulary, mathematical reasoning, and spatial relations, covering ground previously assessed by the individual Services' tests. Third, it was

[1]Welsh, Kucinkas, & Curran (1990) prepared an extensive review of ASVAB validity studies and ASVAB information. We highlight some information from the review here. Please refer to Welsh et al. (1990) for greater detail.

required that the test not penalize individuals who are slow to take such examinations, disallowing the use of speeded tests. Finally, the difficulty of the verbal instructions was to be minimized. The battery's content has been modified since these requirements were set forth. Subtests of spatial ability are not now included, there is increased emphasis on verbal and quantitative items, and speeded tests are used. Each of the subtests is built to provide information useful in predicting success in certain specific jobs and not in others, thereby providing both differential measurement and differential validity (Ree & Earles, 1992). Clerical speed subtests are, for example, designed to predict performance in Administrative jobs. However, it is important to note that the *magnitude of differential validity of the* ASVAB *is* modest (Ree & Earles, 1992).

ASVAB Subtests

The ASVAB that has been administered since 1980 includes ten subtests, eight of which are power tests and two of which are speeded (Welsh et al., 1990). Short test descriptions appear in Table 2. The number of items that are included in each subtest, the amount of time it takes to administer each, and the internal consistency and alternate forms reliabilities of each are provided in Table 3. As noted, the average internal consistency reliability for the subtests is .86. The average alternate forms reliability is .79. With instructions, the battery takes anywhere from 3 to 3 1/2 hours to administer.

A listing of the type of content covered in the ASVAB is provided in the *ASVAB Information Pamphlet* (DoD, 1984b). The first subtest administered in the set is General Science (GS). This test is made up of multiple choice questions about general science, including biology, physics, and earth science. As an example, a question may ask the individual to choose the chief nutrient in lean meat or the correct chemical formula for water. The second subtest, Arithmetic Reasoning (AR), assesses the applicant's ability to solve arithmetic problems. The individual is allowed to use scratch paper to solve such problems as "How many 36-passenger busses will it take to carry 144 people?" The third subtest in the battery, Word Knowledge (WK), assesses the individual's knowledge of the meaning of words. Here, the applicant chooses the word that most nearly approaches the same meaning as that of a given word--"impair" for example. The individual then decides whether "direct," "weaken," "improve," or "stimulate" is closest to the meaning of "impair." The fourth subtest, Paragraph Comprehension (PC), measures the applicant's ability to understand what he or she reads. Here, the individual is required to read one or more paragraphs which are then followed by incomplete statements or questions. The applicant chooses from among the options the response that either best completes the last statement in the paragraph or best answers the question posed about the material.

Numerical Operations (NO), the fifth subtest, is a speeded test that assesses how rapidly and accurately the individual can complete simple arithmetic problems. As such, the applicant attempts to solve as many questions as possible as the time permits without making any mistakes. Given "2 + 3 =," for example, the individual chooses the response that is correct from four alternatives. The other speeded test, Coding Speed (CS), is designed to evaluate how quickly and accurately the applicant can find a number in a table. Each question in the test is a coded word taken from a key above it. The individual is required to find the correct code number for the given word from among five alternatives.

Table 2 ASVAB Subtest Content (Forms 8-17)	
ASVAB Subtest	Content Description
General Science (GS)	Knowledge of the physical and biological sciences
Arithmetic Reasoning (AR)	Word problems emphasizing mathematical reasoning
Word Knowledge (WK)	Understanding the meaning of words (vocabulary)
Paragraph Comprehension (PC)	Presentation of short paragraphs followed by one or more multiple choice items
Numerical Operations (NO)	A speeded test of four arithmetic operations (addition, subtraction, multiplication, & division)
Coding Speed (CS)	A speeded test of matching words and four-digit numbers
Auto and Shop Information (AS)	Knowledge of auto mechanics, shop practices, and tool functions depicted in verbal and pictorial items
Mathematics Knowledge (MK)	Knowledge of algebra, geometry, and fractions
Mechanical Comprehension (MC)	Understanding mechanical principles, such as gears, levers, pulleys, and hydraulics, depicted in verbal and pictorial items
Electronics Information (EI)	Knowledge of electronics and radio principles, depicted in verbal and pictorial items

Sources: Welsh, Kucinkas, & Curran (1990) and DoD (1984b)

The seventh subtest, Auto and Shop Information (AS), has multiple choice questions that cover information about automobiles, shop practices, and the use of tools. The individual may, for example, be asked to identify the correct use of a chisel or identify the tool pictured. Eighth in the series is a test of the individual's ability to solve problems using high school mathematics--Mathematics Knowledge (MK). As with the other math test in the battery, Arithmetic Reasoning, the use of scratch paper is permitted. Individuals choose from multiple alternatives the correct response to such questions as "If 3x = -5, then x = ..." and "The initial digit of the square root of 59043 is ..." The ninth subtest, Mechanical Comprehension (MC), presents diagrams and pictures that are used to assess the individual's knowledge of general mechanical and physical principles. Given the pictorial choice, for example, of a book, a pair of scissors, a rocking chair, and a suit jacket, the individual chooses which of these objects would feel the coldest if all are the same temperature. The last subtest in the battery, Electronics Information (EI), presents items either verbally or pictorially and evaluates the applicant's knowledge of electrical, radio, and electronics information. Here, the individual responds to questions regarding, for example, the safest way to run an extension cord to a lamp or the correct symbol for a transformer.

Table 3 Content and Reliability of ASVAB Subtests				
			Reliability	
Subtest	Number of Items	Test Time (Minutes)	Internal Consistency	Alternate Forms
General Science (GS)	25	11	.86	.83
Arithmetic Reasoning (AR)	30	36	.91	.87
Word Knowledge (WK)	35	11	.92	.88
Paragraph Comprehension (PC)	15	13	.81	.72
Numerical Operations (NO)	50	03	*	.70
Coding Speed (CS)	84	07	*	.73
Auto & Shop Information (AS)	25	11	.87	.83
Mathematics Knowledge (MK)	25	24	.87	.84
Mechanical Comprehension (MC)	25	19	.85	.78
Electronics Information (EI)	20	09	.81	.72
Total/Average	334	144	.86	.79

*Internal consistency reliability not computed for speeded tests.
Sources: DoD (1984a) and Waters et al. (1988)

ASVAB Composites

The Services combine subtests to form composites for selection and classification purposes. The subtest scores that go into the various composites are transformed first using a common standard score metric (i.e., a T-score with a mean of 50 and standard deviation of 10). The common metric is based on a representative sample of 1980 American Youth, ages 18 to 23. (DoD, 1982). The same subtest often may be found in more than one composite. Schratz & Ree (1989) note that even when subtests are not repeated throughout the composites used, the inter-composite correlations are typically high--about 0.7. Ree and Earles (1991b) suggest this is due to psychometric *g* which is discussed later in this chapter.

Each Service forms it own composites (see Table 4). The Air Force, for example, uses Mechanical (M), Administrative (A), General (G), and Electronics (E) composites for their selection and classification purposes. The composite score used by all of the Services as an assessment of general trainability is named the Armed Forces Qualification Test score or AFQT. In the early 1980s, AFQT was composed of three power tests and one speeded test. The current

AFQT composite was derived from a study by Wegner and Ree (1986). Wegner and Ree evaluated fifteen different composites according to several criteria (e.g., validity for predicting school grades, sex and race effects). The resulting, current AFQT is a measure of verbal, reasoning, numeric, and reading ability (2 VE + MK + AR in standard score form).

Use of the AFQT allows comparison among the Services of the overall aptitude of enlisted accessions and serves as an index of trainability (Schratz & Ree, 1989; Welsh et al., 1990). The AFQT percentile score scale is divided into categories, as cited in Table 5. Categories I and II include scores at or above the 92nd and 65th percentile and represent the highest applicant ability categories. Recruits that perform at about average are placed in Categories IIIa and IIIb. Those individuals in Category IV, falling between the 10th and 30th percentiles, are limited as to the number of slots that are made available for them each year. Federal statutes prohibit the enlistment of Category V applicants (Welsh et al., 1990). All of the Services use an AFQT cut-score at the 31st percentile and typically require an additional qualification for potential acceptance.

Factor Structure of the ASVAB

The factor structure of the ASVAB has been examined by a number of researchers over the years. The three most important findings are: (1) the general factor (psychometric *g*) accounts for approximately 60 percent of the total variance (Kass, Mitchell, Grafton, & Wing, 1983; Welsh, Watson, & Ree, 1990), (2) four factors have been identified and replicated across studies (Kass et al., 1983; Welsh et al., 1990), and (3) the four factors have been replicated for males, females, Blacks, Whites, and Hispanic subgroups separately (Kass et al., 1983). The four factors and ASVAB subtests that have substantial loadings are:

(1) Verbal (WK and PC)
(2) Speed (CS and NO)
(3) Quantitative (AR and MK)
(4) Technical (AS, MC, and EL)

GS has loaded on the Verbal factor (Ree, Mullins, Mathews, & Massey, 1982) and has yielded split-loadings on the Verbal and Technical factors (Kass et al., 1983). Otherwise this factor solution is relatively straight forward and is highly replicable. Even so, over half of the variance in ASVAB scores is accounted for by the general factor (Welsh, et al. 1990).

Sex and Racial/Ethnic Group Differences

The ASVAB has been evaluated over the years by both military and civilian test experts. According to Eitelberg et al. (1984), the ASVAB predicts success in technical training for males and females and for racial/ethnic groups equitably. The ASVAB typically overpredicts training performance for minority groups (Welsh et al., 1990). In short, the ASVAB meets accepted guidelines for fairness.

Table 4

Content of Service ASVAB Composites

Army	Air Force	Marine Corps	Navy	ASVAB Subtests
General Technical (GT)	General (G)	—	General Technical (GT)	AR + WK + PC
—	—	General Technical (GT)	—	AR + WK + PC + MC
Electronics (EL)	Electronics (E)	Electronics Repair (EL)	Electronics (EL)	GS + AR + MK + EI
Clerical (CL)	—	—	—	WK + PC + AR + MK
—	Administrative (A)	—	Clerical (CL)	NO + CS + WK + PC
—	—	Clerical (CL)	Business and Clerical (BC)	MK + CS + WK + PC
Motor Maintenance (MM)	—	—	—	NO + AS + MC + EI
—	—	—	Mechanical (ME)	AS + MC + WK + PC
—	—	Motor Maintenance (MM)	—	AR + AS + MC + EI
—	Mechanical (M)	—	—	GS + 2AS + MC
Combat (CO)	—	—	—	AR + CS + AS + MC
Field Artillery (FA)	—	—	—	AR + CS + MK + MC
Operators/Foods (OF)	—	—	—	NO + AS + MC + WK + PC
Surveillance/Communi-cations (SC)	—	—	—	AR + AS + MC + WK + PC
—	—	—	Basic Electricity/ Electronics (E)	GS + AR + 2MK
Skilled Technical (ST)	—	—	—	GS + MK + MC + WK + PC
—	—	—	Boilerman/Enginemen/ Machinist Mate (EG)	AS + MK
General Maintenance (GM)	—	—	—	GS + AS + MK + EI
—	—	—	Machinery Repairman (MR)	AR + AS + MC
—	—	—	Submarine (ST)	AR + MC + WK + PC
—	—	—	Communications Technician (CT)	AR + NO + CS + WK + PC
—	—	—	Hospitalman (HM)	GS + MK + WK + PC

Source: Blozom (1992).

Table 5 AFQT Percentile Score Scale	
AFQT Category	Percentile Range
I	93-99
II	65-92
IIIc	50-64
IIIb	31-49
IV	10-30
V	01-09

Note. From "Armed Services Vocational Aptitude Battery (ASVAB): Integrative review of validity studies" (AFHRL-TR-90-22) by J. R. Welsh, Jr., S. K. Kucinkas, and L. T. Curran, 1990, Brooks Air Force Base, TX: U. S. Air Force Human Resources Laboratory.

Adverse impact is defined as a substantially different rate of selection or classification for population subgroups. Although adverse impact by itself is not an indicator of unfairness, it does seem *prima facie* unfair. Adverse impact certainly raises questions of fairness. Adverse impact results when there are large mean differences in test scores. Such differences have dramatic effects on the proportions of people scoring in the tails of the distribution, and, thus selection. Burnett (1986) illustrated the potential effect of a half of a standard deviation difference in mean scores of males and females, favoring males. If "500 men and 500 women apply for approximately 213 openings in architectural school, ... and a spatial ability test in which men and women differ by half a standard deviation is used as a part of the selection battery, then approximately 142 (28.43%) men but only 71 (14.23%) women would be admitted--twice as many men as women" (p. 1013). Fewer openings, higher cut scores, and/or a greater sex difference would magnify the effect.

Sex Differences

Some ASVAB subtests yield substantial sex effects. The effect size is the difference between males' and females' mean scores expressed in standard deviation units. Table 6 provides ASVAB subtest effect size estimates based on the 1980 youth population from the *Profile of American Youth* (DoD, 1982). As can be seen in Table 6, males tend to perform better on ASVAB measures of mathematics and technical abilities, whereas females perform better on the speeded tests and on the measure of reading comprehension. The largest differences, which favor males, are on the subtests of Auto and Shop Information, where there is more than one standard deviation difference, Electronics Information and Mechanical Comprehension. The differences are large, but they are also consistent with those reported elsewhere for mechanical ability (Bennett & Cruikshank, 1974; Bennett, Seashore, & Wesman, 1974).

Racial/Ethnic Group Differences

As shown in Table 6, mean scores for Whites are higher than mean scores for Blacks and for Hispanics on all of the subtests. The largest differences between Whites and Blacks occur on General Science, Word Knowledge, Auto & Shop Information, and Electronics Information subtests. The smallest difference in performance is reported on the Mathematics Knowledge test. The largest differences between Whites and Hispanics are on the General Science, Word Knowledge, and Electronics Information subtests. The smallest difference is found on the Coding Speed subtest. It is possible that the observed effect sizes in Table 6 distort the magnitude of any "true score" difference between the races because ASVAB subtests are not perfectly reliable. Regardless, the observed differences are of interest because they provide an estimate of adverse impact that will result from an imperfect measure used operationally. Also, as with sex differences, these race/ethnic differences observed for ASVAB subtests are consistent with those reported in reviews of the cognitive abilities literature (Jensen, 1980).

Table 6

Effect Size Values* for ASVAB Subtest Scores by Sex and Racial/Ethnic Group

Subtest	Male-Female	White-Black	White-Hispanic
General Science (GS)	0.36	1.24	1.00
Arithmetic Reasoning (AR)	0.28	1.16	0.85
Word Knowledge (WK)	-0.01	1.29	1.00
Paragraph Comprehension (PC)	-0.19	1.07	0.89
Numerical Operations (NO)	-0.19	0.94	0.70
Coding Speed (CS)	-0.42	0.96	0.60
Auto & Shop Information (AS)	1.25	1.23	0.82
Mathematics Knowledge (MK)	0.14	0.88	0.73
Mechanical Comprehension (MC)	0.83	1.20	0.83
Electronics Information (EI)	0.78	1.22	0.92

* The effect size is the standardized mean difference between two subgroups' mean scores [$d=(MN_{group1} - MN_{group2})/S_p$, where S_p is the pooled standard deviation]. A positive Male/Female effect size indicates superior performance by males, and a negative effect size indicates superior performance by females. A positive White/Black effect size indicates superior performance by Whites, and a negative effect size indicates superior performance by Blacks. Similarly, a positive White/Hispanic effect size indicates that the White mean score was higher than the Hispanic mean score, and a negative score indicates that the Hispanic mean score was higher than the White mean score.

Adapted from: "Profile of American Youth: 1980 Nationwide Administration of the Armed Services Vocational Aptitude Battery" by the Department of Defense, Office of the Assistant Secretary of Defense (Manpower, Reserve Affairs, and Logistics), 1982, Washington, D.C.: Author.

ASVAB Validity

ASVAB Subtests

The validity of ASVAB composites, not the subtests, is usually the focus of validity studies, and thus subtest validity is not always reported. Welsh et al. (1990) meta-analyzed available subtest validities for ASVAB forms that are currently in use. As shown in Table 7, corrected-for-range-restriction validities ranged from .44 for CS to .63 for WK and MK, and .64 for GS, AR and PC for predicting school grades. Standard deviations were relatively large for PC, EI, and GS, suggesting differences across studies, Services, and/or jobs in absolute levels of validity.

Ree and Earles (1992a) recently reported average corrected-for-range-restriction ASVAB subtest validities for predicting final school grades in 150 Air Force jobs. The validities resembled those reported by Welsh et al. (1990). CS yielded the lowest validity, .47; the highest validities were for AR (.68), GS and WK (both .66), and MK (.65). AR, PC, and WK yielded the widest range of validities across jobs.

The ASVAB is usually validated with school grades as criteria. Maier and Mayberry (1989) reported ASVAB subtest validities (corrected-for-range-restriction) for predicting hands-on performance in the infantry rifleman job. The profile of correlations differed substantially from those reported by Ree and Earles (1992) and Welsh et al. (1990). EI (.52), MC (.51), and GS and AS (both .50) were the best predictors. It is tempting to conclude that the changes in the pattern of correlations across studies was due to either substantive criterion differences or differences in criterion reliability. Of course, Ree and Earles (1992) and Welsh et al. (1990) were multi-job studies; that also could account for the difference in patterns.

The Air Force Composites

The four Air Force job clusters--Mechanical, Administrative, General, and Electrical (MAGE)--have been used in one form or another since the mid-1950s (Weeks, Mullins, & Vitola, 1975). They evolved through expert judgment coupled with empirical evidence about the relationships between ASVAB subtests and performance in Air Force training. MAGE has been shown to be "remarkably robust considering the myriad of changes that have taken place since the system was first established" (Alley, Treat, & Black, 1988, p. 10). Alley et al. computed regression equations for predicting performance in 211 training schools, and clustered the individual equations on the basis of their regression weights. After forming clusters, they computed composite regression equations for each cluster. Six clusters were defined, four of which were equivalent to the existing M, A, G, and E clusters in terms of job content and profiles of regression weights. The remaining two clusters contained jobs that either (a) were not well-predicted by the ASVAB subtests or (b) required ability across the full spectrum of the ASVAB.

Table 7 ASVAB Subtest Validity Coefficients*					
	Final School Grades				Hands-on Job Performance
	Welsh et al. (1990)[1]		Ree & Earles (1992)[2]		Maier & Mayberry (1989)[3]
ASVAB Subtest	r	SD_r	r	range	r
General Science (GS)	.64	.35	.66	.17-.84	.50
Arithmetic Reasoning (AR)	.64	.25	.68	.03-.85	.44
Word Knowledge (WK)	.63	.29	.66	.06-.82	.46
Paragraph Comprehension (PC)	.64	.40	.62	-.01-.77	.40
Numerical Operations (NO)	.49	.26	.51	.13-.68	.29
Coding Speed (CS)	.44	.17	.47	.08-.66	.26
Auto & Shop Information (AS)	.49	.19	.52	.04-.70	.50
Mathematics Knowledge (MK)	.63	.25	.65	.11-.84	.38
Mechanical Comprehension (MC)	.58	.25	.59	.01-.73	.51
Electronics Information (EI)	.60	.38	.61	.06-.76	.52
AFQT	.44	.09	.73	.06-.91	---

[1]Based on meta-analysis of validities across Services. Validities are individually corrected for restriction in range and averaged. Total N is greater than 52,000.

[2]Based on 88,724 subjects in 150 Air Force jobs. Mean validities are corrected for range restriction and weighted by sample size for each job.

[3]Based on infantry rifleman data. Sample size is not reported. Validities are corrected for range restriction.

Ree and Earles (1992) investigated the validity of the current MAGE composites. They computed correlations between final school grades and the 10 ASVAB and the MAGE composites. They organized the correlations by job family (e.g., one group of correlations for Mechanical jobs) and computed average corrected-for-range restriction correlations. Some of the Mechanical jobs were better predicted by the Electronics composite, while Administrative jobs included in the study were generally not well-predicted by the Administrative composite. General and Electronics jobs were well predicted by their appropriate composites. The Air Force is currently conducting research to restructure their test composites.

The Army Composites

The Army began using aptitude area (AA) composites along with the Army Classification Battery operationally in 1949 (Maier & Fuchs, 1969). In the four decades since then, the test battery (now the ASVAB), the Army's occupational structure, and the AA composites have changed. The Army has used nine AA composites resembling those in use today since 1973. The two latest generations of AA composites were formed by Maier and Grafton (1981) and McLaughlin, Rossmeissl, Wise, Brandt and Wang (1984). Maier and Grafton used Skill Qualification Test (SQT) scores as criteria for developing composite formulas; they did not investigate alternative groupings of jobs. Usually, when new jobs are assigned AA composites based on rational judgments.

McLaughlin et al. (1984) examined ASVAB validities corrected for range restriction against SQT and training scores for 98 jobs. Almost all jobs were best predicted by their assigned AA composite. McLaughlin et al. also developed an alternative set of composites; they were four-fold--Clerical, Skilled and Technical, Operations, and Combat. However, validity of the four-composite set was not significantly different from that of the nine composite set.

The Marine Corps Composites

The Marine Corps developed the latest version of its composites in 1985. Maier and Truss (1985) computed regression equations for predicting training school grades in 34 job groups. They computed composites that were linear combinations of the tests that group to form the four factors that are traditionally identified for the ASVAB (i.e., not factor scores). The ASVAB factor-based composites were used in the regression rather than subtests to enhance the stability of the results.[2] The mathematical factor had a high weight for all samples, and the authors concluded that all composites should include at least one math subtest. Similarly, the technical factor had high weights for all specialties except clerical jobs, the speed factor had high weights for clerical and field artillery jobs, and the verbal factor had high weights for general technical and clerical jobs. The authors constructed the new composites accordingly.

[2]The four factors traditionally identified for the ASVAB subtests are: (1) Verbal (WK and PC), (2) Mathematical (AR and MK), (3) Technical (AS, MC, and EI), and (4) Speed (NO and CS).

The Navy Composites

The Navy currently uses 11 ASVAB composites and has, over the last few years, investigated ways of reducing the number of composites. Peterson, Gialluca, Borman, Carter, and Rosse (1990) gathered school performance data on more than 20,000 students attending 22 Navy Class "A" schools. They applied several rational and empirical strategies for ASVAB composite development including: the 11 composites currently used operationally, alternative rational composites suggested by Navy Personnel Research and Development Center (NPRDC) researchers, rationally-derived composites used in other Services, and three strategies for empirically identifying alternative composites. They computed corrected-for-range-restriction validities associated with each method. The current composites demonstrated good validity. Alternative rational composites were slightly more valid, on average. The empirical strategies produced somewhat higher validities but appeared to capitalize on chance to some extent.

ASVAB Factors

Validity of *g*. As mentioned earlier, about 60 percent of the variance in ASVAB subtest scores can be accounted for by the first principal factor (or component). Psychometric *g* computed on ASVAB scores has been shown to predict job performance (Ree et al., 1992; Ree & Earles, 1992b) and training school grades (Ree & Earles, 1991b).

The Four ASVAB Factors. The Army's Project A and its follow-on project, Career Force, suggest that the ASVAB is a good predictor of on-the-job technical performance criteria. In short, Project A/Career Force involved:

(1) development of a battery of non-cognitive, spatial, perceptual, and psychomotor measures (Peterson, 1987; Peterson, Hough, Dunnette, Rosse, Houston, Toquam, & Wing, 1990),

(2) development of hands-on, job knowledge, and peer/supervisor/self assessment materials (Campbell, Ford, Rumsey, Pulakos, Borman, Felker, DeVera, & Riegelhaupt, 1990),

(3) concurrent validation of the predictor battery with a sample of more than 9,000 first tour enlisted personnel in 19 jobs (McHenry, Hough, Toquam, Hanson & Ashworth, 1990),

(4) administration of the predictor battery to approximately 45,000 Army recruits (Peterson, Russell, Hallam, Hough, Owens-Kurtz, Gialluca, & Kerwin, 1990), and

(5) collection of end-of-training, first-tour, and second-tour performance criteria for those recruits who remained in the Army. First-tour longitudinal validation data have been analyzed (Oppler, Peterson, & Russell, in press); second-tour data have been collected and will be analyzed this fall and next year.

Criteria were organized to form five composites that can be grouped into two broader categories, "can-do" and "will-do" criteria. Can-do criteria--Core Technical Proficiency (CTP) and General Soldiering Proficiency (GSP)--subsumed job knowledge and hands-on task proficiency measures. Will-do criteria--Effort and Leadership (ELS), Personal Discipline (MPD) and Physical Fitness and Military Bearing (PFB)--generally included a variety of supervisor, self, and peer assessments. Validities were computed using the four ASVAB composites (based on the ASVAB factors): Verbal, Mathematical, Technical, and Speed. Multiple *R*s for ASVAB factors, corrected for range restriction and adjusted for shrinkage, were quite high for the can-do criteria--.63 for CTP in both the Longitudinal Validation (LV) and Concurrent Validation (CV) samples and .65 (CV) and .67 (LV) for GSP. Validities were lower for will-do job performance criteria--.31 (CV) and .39 (LV) for ELS, .16 (CV) and .22 (LV) for MPD and .20 (CV) and .21 (LV) for PFB.

CAT-ASVAB

Research on the development of a Computerized Adaptive Testing (CAT) version of the ASVAB (CAT-ASVAB) began in about 1979 and is on-going. CAT technology relies on Item Response Theory, IRT, to tailor the test to each individual (Schratz & Ree, 1989). Simply stated, CAT narrows in on a specific ability estimate for each individual by selecting and administering items that will provide the best information about the individual's ability. When an IRT-based test begins, an item of moderate difficulty is presented as the test taker is assumed to be of average ability. If the examinee answers the question correctly, a more difficult item is chosen. This continues until the individual fails to respond correctly. The computer then presents questions that fall between the difficulty level of the last question answered correctly and the one to which an incorrect response was just given. In sum, the program continuously selects an item, scores the response, updates the examinee's ability estimate, and identifies the next best item (Sands, 1987). Administration time for the battery is greatly reduced as the program is able to arrive at an accurate estimate of the examinee's ability using fewer items than would a traditional paper-and-pencil test (Palmer, Haywood, & Curran, 1989).[3]

Two subtests in the CAT-ASVAB are modifications of those presently in the paper-and-pencil battery (Palmer et al., 1989). First, Paragraph Comprehension is altered in that only one item is presented for each paragraph on each screen. As such, examinees do not have to scroll back and forth to read a lengthy paragraph and/or multiple items - as they would have to do with the paper-and-pencil version. Second, the Auto and Shop Information subtest is divided into two subtests on the CAT-ASVAB--Auto Information and Shop Information.

Construct and criterion-related validity evidence suggest that the CAT-ASVAB is comparable to its paper-and-pencil counterpart. Factor analyses suggest that CAT subtests measure the same factors as the paper-and-pencil ASVAB (Moreno, Wetzel, McBride, & Weiss, 1984). The predictive validity for the CAT-ASVAB is similar to that of the paper-and-pencil battery regarding final school grades in technical training, using ASVAB subtest and Air Force

[3]Only power tests are administered according to IRT.

composite scores (Palmer et al., 1989). The same classification to training school results regardless of the method by which the ASVAB is administered (Sympson, Weiss, & Ree, 1982).

There also appear to be some differences. Segall (1991) reported that mean performance on the paper-and-pencil battery was better than that on the CAT-ASVAB for both CS and AS. Additionally, there was a sex by administration method interaction for two tests, AS and MC. Males performed at about the same level, regardless of the method of administration. Females performed better on paper-and-pencil version than on the CAT-ASVAB.

The feasibility of operationalizing the CAT-ASVAB, its correspondence with the paper-and-pencil battery, and other associated research is currently being assessed through a joint-Service effort--with NPRDC serving as the lead laboratory (Schratz & Ree, 1989). The equivalence of CAT and printed ASVAB has been pretty well established. CAT-ASVAB has been equated to the printed version, and the accuracy of the equating has been checked in separate follow-on study. CAT-ASVAB is now in the process of a formal Operational Test & Evaluation in several MEPS, where it is being used operationally for personnel selection (J. R. McBride, Personal Communication, 30 August 1992).

Aptitude Measures used to Select Officer Candidates

The Services use several aptitude tests to select officer candidates. The academies use the Scholastic Aptitude Test (SAT) or the American College Test (ACT) in conjunction with high school class rank. Reserve Officer Training Corps (ROTC) programs primarily use SAT and ACT scores to determine eligibility, but some programs require additional tests. For example, the Air Force ROTC program requires that candidates for the Professional Officer Course (a college junior and senior level course) take the Air Force Officer Qualifying Test (AFOQT). The Army requires that applicants to its ROTC non-scholarship programs take the Officer Selection Battery (OSB). Officer Candidate/Training School (OCS) programs require Service-specific tests. The Army uses the OSB and the General Technical (GT) composite of the ASVAB; the Navy uses the Officer Aptitude Rating (OAR), the Academic Qualification Test (AQT), and the Flight Aptitude Rating (FAR), all of which are composites from the Aviation Selection Test Battery (ASTB); and the Air Force uses the AFOQT. The Marine Corps requires that applicants to all of its precommissioning programs obtain a qualifying score on the SAT, the ACT, or the Electronics Repair (EL) composite of the ASVAB. In addition, aviation applicants in the Marine Corps are required to achieve passing scores on the AQT-FAR. The Army also has a special test for pilot selection, the Alternate Flight Aptitude Selection Test (AFAST) and, the Multi-Track Test Battery is used for classification into different rotary wing tracks.

Both the SAT and the ACT play an important role in the selection of officer candidates in college programs. College programs involve a substantial monetary investment in the candidate's education. Consequently, the selection of individuals who will succeed in college is critical at this stage, and the SAT and ACT have been shown to predict college grades. The tests used in selecting officer candidates for OCS and Officer Training School (OTS) programs are intended to predict officer performance--since nearly all applicants, as college graduates, have already demonstrated a level of academic success. Many ROTC non-scholarship programs, geared mainly for college juniors and seniors, likewise use tests that are aimed at predicting success in the military--since most of the upper-class students are expected to complete college.

In this section, we summarize information about several tests developed and administered by the Services--AFOQT, OSB, ASTB, BAT, AFAST and Multi-Track.

The Air Force Officer Qualification Test (AFOQT)

The Air Force developed the AFOQT in the early 1950s as a tool for selecting civilian applicants for officer precommissioning training programs and for classifying commissionees into aircrew job specialties (Rogers, Roach, & Short, 1986; Skinner & Ree, 1987). It has 16 subtests that tap verbal, quantitative, spatial, and mechanical aptitudes. Each subtest is independently timed, and administration time for the entire battery is about 4.5 hours. Table 8 provides names and reliabilities of AFOQT subtests.

The AFOQT measures attributes similar to those measured by the ASVAB. Sperl, Ree, and Steuck (1990, 1992) administered the verbal and quantitative components of the AFOQT to 516 airmen in Basic Military Training (BMT) who had taken the ASVAB prior to enlistment. Correlations between AFOQT and ASVAB subtests of the same attributes were high, and similar attributes from different tests loaded together in a factor solution. Sperl et al. also demonstrated that the ASVAB could be used to predict AFOQT composite scores.

Scores on the 16 AFOQT subtests form five composites: pilot, navigator-technical, academic aptitude, verbal, and quantitative (Sperl & Ree, 1990). The academic aptitude composite combines the verbal and quantitative composites and is roughly analogous to sections of the SAT (Rogers et al., 1986). The pilot and navigator-technical composites are used for classification into Undergraduate Pilot Training (UPT) and Undergraduate Navigator Training (UNT), respectively. The composites have factor-analytic support, and composite reliabilities (alpha coefficients) have been consistently high, ranging from .93 for the verbal composite to .97 for the navigator-technical composite in two samples (Sperl & Ree, 1990).

Sex and race/ethnic score differences on the AFOQT composites are similar in magnitude to those reported elsewhere for scores on cognitive tests (cf. Jensen, 1980; Willerman, 1979). Male means are higher than female means on all five composites. Sperl and Ree (1990) reported means and standard deviations (SD) of scores for two samples (male N=1285, female N=320; male N=1201, female N=208). The average effect sizes across the two samples were .76 SD for the pilot composite, .67 SD for navigator-technical, .30 SD for academic aptitude, .16 SD for verbal, and .45 SD for quantitative, all favoring males. Similarly, White means are higher than Black means on all five composites. The average differences across two samples ranged from 1.29 SD for the verbal composite to 1.78 SD for the pilot composite (Sperl & Ree, 1990). However, the differences varied considerably across the two samples and were probably not very stable given the small samples of Blacks; each sample included fewer than 200 Blacks and more than 1100 Whites.

Table 8 Content and Reliability of AFOQT Subtests				
Subtest	Number of Items	Test Time (Minutes)	Internal Consistency Reliability (Alpha)	
			Form P1	Form P2
Verbal Analogies	25	08	.83	.84
Arithmetic Reasoning	25	29	.90	.90
Reading Comprehension	25	18	.87	.89
Data Interpretation	25	24	.83	.86
Word Knowledge	25	05	.90	.90
Math Knowledge	25	22	.92	.92
Mechanical Comprehension	20	22	.78	.79
Electrical Maze	20	10	---	---
Scale Reading	40	15	---	---
Instrument Comprehension	20	06	---	---
Block Counting	20	03	---	---
Table Reading	40	07	---	---
Aviation Information	20	08	.83	.83
Rotated Blocks	15	13	.76	.79
General Science	20	10	.82	.80
Hidden Figures	15	08	.80	.78
Total/Average	380	208	.84	.85

Note. From "Air Force Officer Qualifying Test (AFOQT): Forms P pre-implementation analyses and equating" (AFHRL-TP-88-6) by K. W. Steuck, T. W. Watson, and J. Skinner, 1988, Brooks Air Force Base, TX: U. S. Air Force Human Resources Laboratory.
Alpha coefficients are not reported for speeded tests.

The AFOQT predicts training success. Initially, AFOQT composites were shown to be valid only for aircrew positions (Hunter & Thompson, 1978; Valentine, 1977). Recent work has focused on predicting training for non-aircrew jobs. Arth and Skinner (1986) analyzed scores of 1025 active duty officers assigned to eight Air Force Specialties (AFS) that required entry-level technical training. AFOQT subtests and composites correlated significantly with performance in non-rated technical training courses. Arth (1986), and Finegold and Rogers (1985) reported similar findings. In a study of the Air Force Reserve Officer Training Corps (AFROTC) selection system, Cowan, Barrett, and Wegner (1989) found that the academic aptitude composite predicted instructor's ratings of performance in the Professional Officer Course (N=5249) as well as training school grades (N=1645) and supervisors' ratings of potential (N=1080). Hartke and Short (1988) conducted a meta-analysis of the validity of the academic aptitude composite for predicting training school grades. Academic aptitude was consistently predictive for intelligence and security police specialties. However, academic aptitude validities were not sufficiently homogeneous within the other occupational groups studied, suggesting that validities vary across jobs within occupations.

The Officer Selection Battery (OSB)

The Army, like the Air Force, initiated a testing program for selecting officers shortly after World War II. Since 1986, the Army has used Forms 3 and 4 of the OSB to select candidates for Reserve Officer Training Corps (ROTC) non-scholarship programs.[4] Table 9 lists the types of items on the OSB. The verbal, quantitative, general information, and spatial items are much like those on other tests designed to predict success in school (e.g., the SAT). The other items, labeled "Problem Solving," are situational judgment items that present problem scenarios along with choices of solutions. The situational judgment items are designed to tap initiative, assertiveness, and interpersonal abilities--abilities that were identified job-analytically as important for officer performance.

As shown in Table 9, internal consistency estimates for the OSB are relatively high, even though the content of the OSB is intended to be diverse. The OSB yields one score; there are no composites. Males' scores are higher than females' scores by about .17 SD, and Whites score higher than Blacks.

The OSB has demonstrated validity for predicting school performance (i.e., faculty ratings and school grades) (Fischl, Edwards, Claudy, & Rumsey, 1986). The two test forms were administered to two samples of senior ROTC cadets (N=577 and N=788); faculty ratings of leadership characteristics and officer potential served as criteria. Uncorrected correlations were .26 (Form 3) and .28 (Form 4) between faculty ratings and OSB scores. Regression analyses for each form separately by sex and ethnic subgroup as well the total group suggested no differential validity, though subgroup samples were quite small. One of the test forms (Form 3) was also administered to 577 Second Lieutenants in their first assignments (at Officer Basic Courses).

[4]Forms 1 and 2 of the OSB are actually the Cadet Evaluation Battery (CEB) renamed. Forms 1 and 2 of the OSB are substantively different tests from forms 3 and 4.

Uncorrected correlations between OSB scores and Officer Basic grades ranged from .45 to .77 across seven courses, with a mean of .52.

Table 9

Content of Officer Selection Battery

	Form 3 (N = 2836)	Form 4 (N = 2446)
	Number of Items	Number of Items
Verbal	32	32
Quantitative	25	25
General Information	20	20
Problem Solving		
General Information	18	15
Assertiveness Problems	02	02
Initiative Problems	02	03
Managerial Problems	02	03
Social Problems	01	02
Spatial		
Folding/Unfolding Geometric Forms	00	04
Map Reading		
Three-Dimensional Figures	04	04
	04	00
Mean	74.82	74.40
Standard Deviation	14.27	15.39
Coefficient Alpha	.92	.94
Male/Female Effect Size	.17	.17
Black/White Effect Size	1.73	1.78

The effect size is the standardized mean difference between two subgroups' mean scores [$d=(MN_{group1} - MN_{group2})/S_p$, where S_p is the pooled standard deviation]. A positive Male/Female effect size indicates superior performance by males, and a negative effect size indicates superior performance by females. Similarly, a positive White/Black effect size indicates that the White mean score was higher than the Black mean score, and a negative score indicates that the Black mean score was higher than the White mean score.

Note. From "Development of Officer Selection Battery Forms 3 and 4" (Technical Report 603) by M. A. Fischl, D. S. Edwards, J. G. Claudy, and M. G. Rumsey, 1986, Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

The Aviation Selection Test Battery (ASTB)

The Navy and Marine Corps use the Aviation Selection Test Battery to select individuals for Officer Candidate School (OCS) programs (Brown, 1989). The ASTB grew out of a World War II research effort, the Pensacola 1000 Aviator Study, which examined over 60 psychological, psychomotor, and physical tests (North & Griffin, 1977). Three tests recommended for aviator selection on the basis of the study were the Wonderlic Personnel Test (a test of general intelligence), the Bennett Mechanical Comprehension Test (a test of mechanical interest and ability), and the Purdue Biographical Inventory (a measure of morale, interest, and attitude). The current version of the ASTB closely resembles the original; it has four tests: the Academic Qualification Test (AQT), the Mechanical Comprehension Test (MCT), the Spatial Apperception Test, and a Biographical Inventory (BI). The Academic Qualification Test contains subtests that measure quantitative ability, verbal ability, practical judgment, and perceptual speed. Like its counterparts from the other Services, AQT resembles the Scholastic Aptitude Test. The MCT, Spatial Apperception Test, and BI scores form the Flight Aptitude Rating (FAR) composite which has been shown to predict performance in pilot training (Gibb, 1987, 1990; North & Griffin, 1977). Another composite, the Officer Aptitude Rating (OAR), is used in the selection of nonaviation applicants; it combines the AQT and MCT scores.

The Basic Attributes Test (BAT)

The Air Force developed the Basic Attributes Test (BAT) to supplement the AFOQT for pilot selection (Carretta, 1987a, 1987b, 1987c). The BAT is not currently in use, but the Air Force plans to administer the BAT at ROTC field encampments and selected colleges in the near future. It will be used, initially, during ROTC as a guidance tool, not as a screening device (MAJ. D. Perry, personal communication, 24 April 1992).

The BAT is a battery of tests designed to measure cognitive, perceptual, and psychomotor aptitudes as well as personality and attitudinal characteristics (Carretta, 1987a, 1987b, 1987c, 1991, 1992). Several of the BAT subtests are descendants of the classic Army Air Force work and later work by Fleishman and his colleagues (e.g, Fleishman & Hempel, 1956). Other tests are based on more recent information-processing research. Descriptions of BAT subtests appear in Table 10.

Some BAT subtests have proven to be effective predictors (Bordelon & Kantor, 1986; Carretta 1987a, 1987b, 1987c, 1990, 1991, 1992; Stoker, Hunter, Kantor, Quebe, & Siem, 1987). The psychomotor abilities tests on the BAT have demonstrated strong relationships with success in Undergraduate Pilot Training (UPT), advanced training assignment, and in-flight performance scores. The cognitive/perceptual tests have not predicted training outcomes, although they have shown a relationship to in-flight performance measures. Initially, research using scores from the personality and interest portions of the BAT yielded little or no relationship with training outcomes or assignments, however research in this arena has recently intensified.

Alternate Flight Aptitude Selection Tests (AFAST)

The Army developed the original FAST in response to unacceptably high attrition rates in the flight training program (Kaplan, 1965). At that time, FAST was two separate batteries, one for commissioned officers and one for warrant officers. Eastman and McMullen (1978) revised the FAST to form one shorter battery, the RFAST, which has been shown to predict rotary-wing training performance (Lockwood & Shipley, 1984). The current AFAST is a modified version of the RFAST. It has alternate forms and better graphics than the RFAST, and some RFAST items with poor psychometric properties have been removed. The AFAST is a paper-and-pencil test with six subtests that are described in Table 11 (Department of the Army, 1987). A new battery, the NFAST, is in development but has not been used operationally.

Table 10

Basic Attributes Test (BAT) Battery Summary

Test Name	Length (mins)	Attribute Measured	Types of Scores	Cronbach Alpha	Guttman Split-Half
Two-Hand Coordination (rotary pursuit)	10	Tracking and Time-Sharing Ability in Pursuit	Tracking error x axis Tracking error y axis	.94 .95	.58 .65
Complex Coordination (stick and rudder)	10	Compensatory Tracking Involving Multiple Axes	Tracking error x axis Tracking error y axis Tracking error z axis	.95 .99 .94	.62 .56 .41
Encoding Speed	20	Verbal Classification	Response time Response accuracy	.96 .71	.65 .40
Mental Rotation	25	Spatial Transformation and Classification	Response time Response accuracy	.97 .90	.79 .71
Item Recognition	20	Short-Term Memory, Storage, Search and Comparison	Response time Response accuracy	.95 .54	.79 .55
Time-Sharing	30	Higher-Order Tracking Ability, Learning Rate and Time-Sharing	Tracking difficulty Response time Dual-task performance	.96	.80
Self-Crediting Word Knowledge	10	Self-Assessment Ability, Self-Confidence	Response time Response accuracy	.89 .65	.72 .86
Activities Interest Inventory	10	Survival Attitudes	Response time Number of high-risk choices	.95 .86	.70 .86

Source: Carretta (1991, 1992)

Table 11. Alternate Flight Aptitude Selection Test (AFAST) Subtests

Subtest	Description	Number of Items	Time Limit (min)
Background Information	Contains questions about the examinee's background.	25	10
Instrument Comprehension	Requires that the examinee identify which one of five airplane drawings has a position and direction consistent with given instrument readings.	15	5
Complex Movements	Requires the examinees to judge distance and visualize motion based on a given set of symbols.	30	5
Helicopter Knowledge	Measures the degree of general and technical knowledge of helicopter operation and aerodynamics.	20	10
Cyclic Orientation	Measures the ability to identify the cyclic movement required to produce a specific change in the orientation of a helicopter.	15	5
Mechanical Functions	Measures general mechanical aptitude through pictures illustrating various mechanical principles.	20	10

Multi-Track Test Battery

In 1988, the Army implemented the Multi-Track Test Battery for assigning flight students into four helicopter tracks (Intano & Howse, 1991; Intano, Howse, & Lofaro, 1987, 1991a, 1991b). The Multi-Track is actually an assembly of test batteries developed by the Army, Navy, Air Force, and National Aeronautics and Space Administration (NASA). As shown in Table 12, it includes: (1) five subtests from the Complex Cognitive Assessment Battery (CCAB) which was developed by ARI, (2) two tests from the Air Force's Basic Attributes Test (BAT), (3) a questionnaire designed for NASA to assess attitudes and leadership potential (i.e., the Cockpit Management Attitude Questionnaire), and (4) the Complex Coordination/Multi-Tasking Battery (CCMB) which was developed by the Naval Aeromedical Research Laboratory (NAMRL). The CCMB contains seven computer assisted subtests, in increasing difficulty. It begins with a relatively simple psychomotor task, then a dichotic listening task. Subsequent tasks require various combinations of psychomotor tasks, along with dichotic listening. The current platform for the Multi-Track Battery is a 15 MHz, 80286 processor-based personal computer with a custom interface card.

Tests of Specific Aptitudes

The Defense Language Aptitude Battery (DLAB)

The Defense Language Institute (DLI) maintains the Defense Language Aptitude Battery (DLAB). The DLAB is a 90 minute, 119-item multiple choice test that is used to select candidates for foreign language training (Petersen & Al-Haik, 1976; Silva, White, & Rumsey, 1991; White, Hanser, & Park, 1988). Different cut scores are applied to the DLAB for different languages. Foreign languages are divided into four difficulty levels. For example, Spanish, one of the easier languages for English speaking people to learn, is in the lowest difficulty category.

The DLAB requires examinees to learn and use an artificial language. The items on the DLAB came from two tests: Horne's Assessment of Basic Linguistic Abilities (HABLA) and the Al-Haik Foreign Language Auditory Aptitude Test (AFLAAT). The HABLA items require subjects to form language concepts from pictures. Pictures captioned with text (in an artificial language) are shown at the top of the page. At the bottom of the page, the subject must match pictures with appropriate text. Sections of the AFLAAT that appear on the DLAB involve processing auditory information, recognizing phonetic patterns, and applying new grammatical rules to English text. The DLAB items measure three factors underlying HABLA and AFLAAT items. Even so, the DLAB yields only one score.

There is some evidence that language aptitude measured by the DLAB is related to quantitative ability. White et al. (1988) correlated DLAB scores with ASVAB subtest scores using data from 5010 Army enlisted personnel. Correlations ranged from .11 for Auto Shop to .50 for Math Knowledge, with a median of .35. Unfortuantely, White et al. did not correct for range restriction due to preselection on the ASVAB. However, Silva et al. (1991) did compute corrected-for-range-restriction correlations between DLAB scores and four ASVAB composites: Verbal (GS + .5 WK + .5 PC), Quantitative (AR + MK), Technical (AS + MC + .5 EI), and

Speed (NO + CS). Correlations with DLAB scores (N=5671) were .75 with Quantitative, .70 with Verbal, . 59 with Speed, and .53 with Technical.

Table 12. Multi-Track Test Battery Origins and Subtests

Original Battery	Subtests/Scales	References
Complex Cognitive Assessment Battery (CCAB)	Word Anagrams	Samet, Gerselman, Zajaczknowski, & Marshall-Mils (1986)
	Tower Puzzle	
	Mark Numbers	
	Numbers and Words	
	Information Purchase	
Basic Attributes Test (BAT)	Word Knowledge	Siem & Carretta (1986)
	Manikin	
Cockpit Management Attitude Questionnaire (CMAQ)	2 Performance Composites	Helmreich (1987)
	Cockpit Procedure and Atmosphere	
	Leadership	
	Vulnerability	
	5 Item Cluster Composites	
Complex Coordination/Multi-Tasking Battery (CCMB)	Psychomotor (PMT) - Stick Only	Griffin & McBride (1986)
	Dichotic Listening Task (DLT)	
	Dual PMT and DLT	
	Psychomotor (Stick and Rudder)	
	Triple (Stick, Rudder, and DLT)	
	Triple (Stick, Throttle, and DLT)	
	Psychomotor (Stick, Rudder, and Throttle)	

The DLAB predicts success in language training (Petersen & Al-Haik, 1976; Silva et al., 1991). Peterson and Al-Haik (1976) validated the DLAB on a sample of 879 graduates from 12 language courses. The zero-order correlation of the DLAB total score with course grades was .43. Silva et al. showed that the DLAB improved the prediction of end-of-training language proficiency over using the ASVAB alone, with gains ranging from .02 to .14. Verbal and Quantitative ASVAB composites were not as consistent in predicting training outcomes as the DLAB. DLI is currently in the process of developing a new version of the DLAB.

Tests of Aptitudes Relevant to Intelligence Jobs

The Air Force and the Army both have tests of aptitudes relevant to intelligence jobs, and both tests are called RCAT. However the two tests are very different from each other. The Air Force test is a paper-and-pencil cognitive measure--the Radio Communication Aptitude Test--and has demonstrated incremental validity over the ASVAB for predicting training success. The Army's Radio Code Aptitude Test (RCAT), formerly the Army Radio Code (ARC) test, is an auditory perception test designed to identify individuals likely to perform Morse Code tasks adeptly. Although initial validation results for the ARC were very promising (Fleishman, 1955), recent Army validation work suggests that ASVAB composites are better predictors of both training proficiency and attrition than the Army RCAT (J. M. Silva, personal communication, 23 July 1992). Consequently, the Army is in the process of developing a new battery named Superdit (a nickname for those who perform Morse Code tasks well). Superdit contains 10 subtests that measure reaction time to auditory and visual stimuli, sound memory, response initiation time, learning rate, and the ability to hold sound patterns in memory as auditory stimuli continue. Superdit is in the pilot testing stages.

Discussion

The ASVAB

The ASVAB is a highly useful general purpose predictor. ASVAB subtests, composites, and the ASVAB general factor are valid predictors of job and training performance. The ASVAB predicts training success in a host of schools, for a variety of jobs, and in all the Services. Job performance validity information is limited but what is available indicates that the ASVAB predicts performance of the technical aspects of jobs (e.g., hands-on tasks). Efforts to improve the ASVAB need to focus on two major areas: (1) broadening its coverage of cognitive constructs and (2) reducing adverse impact.

Broadening the Coverage of Cognitive Constructs

Horn's (1989) taxonomy provides a heuristic organization for the ASVAB subtests and other tests discussed in this chapter. Recall that Horn's factors are not necessarily based on factor-analytic evidence; Horn drew on physiological and cognitive studies in preparing his framework. Indeed, factor-analytic evidence would show that all cognitive factors are correlated with each other--that one general factor underlies cognitive test scores (e.g., Jensen, 1986). The organization of tests into Horn's factors, shown in Table 13, represents our judgment, not factor-analytic results. We intend to use the framework simply as an organization tool.

Within this taxonomy, NO and CS are speeded tests subsumed by G_s, *Broad Speediness*, and MK and AR are tests of G_q, *Quantitative Thinking* or G_f, (Kyllonen and Christal, 1991, use AR and MK as measures of G_f). The remaining subtests are primarily measures of *Knowledge* or *Crystallized Intelligence*, G_c. They are achievement-oriented. Indeed, five of the 10 ASVAB subtests (WK, PC, NO, AR, MK) can be construed as the "basic skills" core of a standardized achievement test battery. Three others (GS, AS, EI) are plainly measures of attained knowledge, and MC contains a lot of physics knowledge (J. R. McBride, Personal Communication, 30 August 1992). As mentioned in Chapter I, the extent to which a test measures G_c versus G_f depends in part upon the content of the test in relation to the knowledge base of the examinees. If the items require application of rules or principles that are equally familiar to all examinees, the test is a

G_f measure. Since almost all military applicants are high school graduates, in general they should be knowledgeable in areas that are a part of a standard curriculum, such as math and English. Applicants are more likely to vary in their knowledge of areas that are not a core part of such a curriculum (e.g., AS, MC, and EI). Moreover, within the context of Horn's (1989) framework, G_c, G_s, and G_q are covered by the ASVAB. G_v, *Broad Visualization*, G_f, *Fluid Intelligence*, SAR, *Short Term Acquisition and Retrieval*, TSR, *Long Term Storage and Retrieval*, and G_a, *Auditory Intelligence* are not covered.

G_c measures are very useful selection and classification tools (Humphreys, 1986), but the ASVAB's coverage of G_c appears to lack balance. The three technical tests correlate about .75 with each other and load together in factor solutions. All three may not be necessary. Also, there are no complimentary G_c measures for other aspects of curricula likely to be relevant to enlisted jobs (e.g., business information, accounting, computer knowledge). This is not an argument for more G_c measures; indeed some authors have suggested that the ASVAB is too achievement-oriented (McBride, 1991).

The choice between achievement-oriented and ability measures is more a matter of policy than psychometrics. Achievement-oriented tests are very useful selection and classification tools (Humphreys, 1986), and training costs are likely to decrease if experienced individuals are hired. However, because they tap prior knowledge and experience, achievement measures are more susceptible to opportunity bias. Some individuals may not have had (or undertaken) opportunity to acquire specific knowledges. Therefore, if the Services choose to emphasize achievement, high school students interested in joining the Services should be informed about the kinds of course work likely to be most useful to them in preparing for a military job. There is probably some loss of predictive efficiency to the extend that the Services choose to emphasize ability. However, such measures may be perceived as more fair.

Comparisons of the factor structure of the ASVAB with other published tests have provided empirical evidence that the ASVAB lacks a G_v measure. Wise and McDaniel (1991) used confirmatory factor analysis to compare the ASVAB and the General Aptitude Test Battery (GATB). They concluded that the ASVAB is weak in the area of spatial and perceptual abilities. In a similar comparison of the ASVAB and the Differential Aptitude Battery (DAT), McBride (1991) reported mixed results regarding a spatial construct. Analyses suggested that the spatial construct contributed a small amount of unique reliable variance to the ASVAB, and that the spatial construct was highly correlated with the ASVAB quantitative construct. In all, there is some empirical evidence that a spatial measure would complement the current ASVAB subtests, although the magnitude of the incremental validity of a spatial test beyond that of the ASVAB is likely to be small (Carey, 1992).

Assuming that broadening the coverage of cognitive constructs measured by the ASVAB is a worthwhile goal, future supplements should focus on G_f, G_v, SAR, TSR, and perhaps G_a constructs. The Services recognized these deficiencies and included G_v and G_f measures on the ECAT. Chapters III and V discuss some cognitive measures that are being considered for inclusion in the ASVAB and other measures that are in developmental stages.

Table 13

Individual Differences Attributes and Constructs and Selected Military Measures

Broad Attributes		Related Constructs	Selected Measures Developed by the Services
Cognitive			
G_c -	Knowledge or Crystallized Intelligence	Knowledge of general information Word knowledge	ASVAB [GS, PC, WK, AS, MC, EI] OSB, AFOQT
G_f -	Broad Reasoning or Fluid Intelligence	Inductive reasoning Conjunctive reasoning Deductive reasoning	AFOQT
G_v -	Broad Visual Intelligence	Spatial visualization Spatial orientation	BAT, AFOQT, OSB
SAR -	Short Term Acquisition and Retrieval	Recency memory Word span	BAT
TSR -	Long Term Storage and Retrieval	Associational fluency Expressional fluency Ideational fluency	
G_s -	Broad Speediness	Visual scanning Visual matching	ASVAB [CS, NO] BAT, AFOQT
G_a -	Auditory Intelligence	Discrimination among sound patterns Auditory cognition of relations	DLAB, ARC, Superdit
G_q -	Quantitative Thinking	Computational fluency Numerical computation	ASVAB [AR, MK] OSB, AFOQT
Eng -	English Adeptness	Word parsing Phonetic decoding	
Psychomotor			
Dexterity		Finger dexterity Manual dexterity	
Basic Movement Speed and Accuracy		Reaction time Control precision Speed of arm movement	
Perceptual-Motor Movement Control		Multi-limb coordination Rate control	BAT

(Continued)

Table 13 Individual Differences Attributes and Constructs and Selected Military Measures (Continued)		
Broad Attributes	Related Constructs	Selected Measures Developed by the Services
Physical		
Muscular Strength	Muscular tension Muscular power Muscular endurance	
Cardiovascular Endurance	Cardiovascular endurance	
Movement Quality	Flexibility Balance Coordination	
Personality		
Extraversion	Sociable, Gregarious Ambitious, Achievement-Oriented	OSB
Emotional Stability	Emotional, Anxious, Depressed	
Agreeableness	Good-natured, Cooperative	
Conscientiousness	Dependable, Responsible	
Intellectance	Curious, Broad-minded	
Interest		
Realistic	Practical, likes hand-on work	BAT
Investigative	Curious, likes academic endeavors	
Artistic	Creative, likes self-expression	
Social	Friendly, likes people	
Enterprising	Ambitious, likes managing & directing	
Conventional	Concrete, likes exactness in work	

Source: Cognitive (Horn, 1989); Psychomotor (Fleishman, 1967; Imhoff & Levine, 1981; McHenry, 1987); Physical (Hogan, 1991a); Personality (Barrick & Mount, 1991; Digman, 1990; Tett, Jackson, & Rothstein, 1991); Interests (Holland, 1983).

Reducing Adverse Impact

As noted before, the ASVAB does not typically result in predictive bias. There is, however, adverse impact in selection and classification. Adverse impact results not only from the nature of the ASVAB itself but also from policy. For example, the Air Force requires applicants to meet minimum standards on MAGE, which does yield a gender difference, while the other Services use AFQT which results in a smaller gender difference (Russell et al., 1992). In light of this difference, the General Accounting Office recommended that the Air Force review its selection policy (GAO, 1991). Moreover, reducing adverse impact against women will require changes in policy--in the way test scores are used--as well as changes in the ASVAB itself.

Sex and race differences in ASVAB scores are not trivial, particularly on the technical subtests. AS, MC, and EI yield the largest sex and race differences. Sex differences range from .80 SD for MC to 1.18 SD for AS. Black-White differences are greater than 1.25 SD for each test. When the three tests are unit weighted to form a "technical" score, the sex difference is 1.06 SD and the Black-White difference is 1.45 SD (Peterson, Russell et al., 1990).

It is unrealistic to expect to completely eradicate sex and race differences with new tests. Hopes of finding culture fair cognitive tests with no differences, popular in the 1960s and 1970s, were dashed when tests designed to be culturally fair often yielded results favoring whites (Jensen, 1980). Even so, there are two ways to reduce the impact. First, there is evidence that some tests yield differences that are smaller than those from other tests of the same broad construct (Linn & Petersen, 1985); therefore decision- makers have some leeway. The ASVAB Technical Review (ART) committee has, for example, included adverse impact as one of its evaluation criteria for the ECAT measures. The other way of reducing overall adverse impact is to use non-cognitive, particularly personality, measures that traditionally yield no differences or differences favoring minority groups.

Officer Measures

Efforts to improve officer measures should focus on three areas: (1) development of a Joint-Service Test Bank, (2) continued expansion of the individual differences domain, and (3) reduction of adverse impact.

The objectives of a Joint-Service Test Bank would be to enhance the accessibility of test information and to encourage experimentation with tests across Service boundaries. Currently, information about officer and other tests used by the Services is not easy to collect. Information is spotty; for example, race and sex differences are often not reported. What information is available is inconsistent in format and difficult to cumulate. Finally, there is no central resource where test information is available. Researchers planning to develop new tests or a battery of tests must do considerable "leg-work" (phone-calls, literature searches) to find out whether another Service is undertaking a similar effort or has such tests on hand. A Joint-Service Test Bank would maintain a data base of descriptive and psychometric test information for military research purposes. It is important to note here that the Services have taken some steps in this direction with the development of joint-service Training and Personnel Systems Technology

Evaluation and Management (TAPSTEM) committees that oversee selection and classification research.

With regard to measurement of individual differences, officer measures tap a wider range of constructs than does the ASVAB (see Table 11). The AFOQT, for example, includes a number of spatial tests, and if used in conjunction with the BAT and a personality measure, taps a large portion of the full domain of constructs. Similarly, the OSB and the ASTB include interpersonal and/or attitudinal measures. Such efforts appear promising, particularly to the extent that non-cognitive measures add incremental validity to cognitive measures and reduce adverse impact.

Information about sex and race differences on tests is often not reported, but information that is available suggests that differences on the cognitive portions of tests are large enough to be concerned about. Efforts to reduce race and sex effects should continue.

III. COGNITIVE, PSYCHOMOTOR, AND PHYSICAL ATTRIBUTE MEASURES

Teresa L. Russell, Felicity A. Tagliareni, and Linda Batley

Although the ASVAB has been shown to be a valid predictor in a variety of military settings, there are several individual differences domains not measured directly by the ASVAB. Recognizing this, the Services are working to select predictors for inclusion in the ASVAB. In 1989, a Technical Advisory Selection Panel (TASP) was established to recommend tests to supplement or enhance the ASVAB. The TASP solicited suggestions for supplemental predictors from the military testing community. After reviewing information about tests, TASP recommended nine tests: three spatial tests (G_v), two working memory capacity tests (G_f), one figural reasoning test (G_f), one perceptual speed test (G_s), and two psychomotor tests. These tests form the Enhanced Computer Assisted Test (ECAT) battery. Table 14 provides descriptive information about the ECAT measures. Six of the tests originated in the Army's Project A (Peterson, Hough et al., 1990; Peterson, Russell et al., 1990; Walker, 1989). One spatial test and the two working memory capacity tests were drawn from Navy projects (Alderton, 1989a, 1989b; Larson, 1989). The Navy has overseen the preparation of the battery and is currently collecting and analyzing data on some of the ECAT measures (Martin, 1992; Sands, 1990). Similarly, both the Army and the Marine Corps have recently collected data on some of the ECAT measures (Carey, 1992; Mayberry & Hiatt, 1990; Oppler et al., 1992).

Currently the ASVAB Review Technical Committee (ART) is assembling information to make decisions about changes in ASVAB content, and all Services are contributing ideas and analyses. With this in mind, we summarize research relevant to cognitive, psychomotor, and physical abilities constructs likely to supplement the ASVAB in this chapter, highlighting, where appropriate and available, results on the ECAT and other measures likely to be of interest to the Services.

Cognitive Attributes

Cognitive Attribute Definitions

In this section of Chapter 3, we begin by reviewing definitions of cognitive constructs and placing specific tests within that framework. Then we discuss sex and race differences, practice effects, and validity evidence for each construct.

As described in Chapter I, Horn (1989) integrated information processing research with traditional factor-analytic results and evidence from physiological studies of brain injury and other impairments to identify narrow and broad cognitive factors. Narrow (or primary) factors are ones for which the intercorrelations among the sub-factors are large; broad factors (second-order) are defined by tests that are not as highly intercorrelated. As previously discussed, he defines six broad cognitive attributes--G_c, G_f, G_v, SAR, TSR, G_s, and G_a--and two other factors that are important in specific settings, G_q and Eng. In this chapter, we focus on factors that are

Table 14

Psychometric Properties of Enhanced Computer Assisted Test (ECAT) Measures

Test[1]	Number of Items	Time Limit	Sample[2]	Mean	SD	Internal Consistency[3]	Test-Retest	Uniqueness[4]	Average Difficulty
ECAT Spatial Tests									
Assembling Objects	32 items	16 min	A	23.3	6.7	.91	.70	.65	
	36 items	18 min	B	23.6	7.2	.88			.65
	36 items	18 min	D	22.3	7.2	.88	.57		
	28 items	*	F	17.7	5.6	.88			.63
	32 items	12 min	I	24.4	7.9	.87	.83	—	—
	32 items	12 min	J	—	—	.91	.75	.43	—
Orientation Test	24	10 min.	A	11.0	6.2	.89	.70	.60	
			B	12.3	6.2	.89			.51
			F	11.9	5.7	.86			.49
			I	12.8	6.0	.89	.75	—	—
			J	—	—	.92	.73	.42	—
Integrating Details	40		C	1.08	.21	.98	.65	.90	
Integrate Time			C	.45	.14	.93	.65	.92	
Decision Time			C	.69	.13	.78	.66	.45	
			I	.78	.13	.79	.79	—	—
Proportion Correct			J	—	—	.75	.79	.23	—
ECAT Reasoning Tests (Figural)									
Figural Reasoning Test	30	12 min.	A	19.1	5.7	.87	.65	.54	
			B	19.5	5.4	.85			.65
			D	18.9	5.8	.85	.58		
			F	18.8	5.9	.85			.63
			I	20.9	6.0	.87	.75	—	—
			J	—	—	.93	.80	.35	—
ECAT Reasoning Tests (Working Memory)									
Mental Counters	40		E	.67	.19	.90	.64	.68	
			I	.79	.16	.89	.79	—	—
Sequential Memory	40		H	.67	.15	.90		.65	
			I	.71	.14	.88	.81	—	—
			J	—	—	.90	.74	.48	—
ECAT Perceptual Speed Tests									
Target Identification	36	7 min.							
Decision Time			A	1.94	.63	.97	.78	.83	
			B	1.79	.60	.97		.82	
			G	1.54	.54	.97			
			I	1.67	.55	.97	[illegible]	—	—
			J	—	—	.98	[illegible]	.76	—
Proportion Correct			A	.91	.07	.62	.4[illegible]	.59	
			B	.90	.08	.66			
			G	.90	.10	.77			
			J	—	—	.64	.52	.59	—

(Continued)

Test[1]	Number of Items	Time Limit	Sample[2]	Mean	SD	Internal Consistency[3]	Test-Retest	Uniqueness[4]	Average Difficulty[5]
Table 14									
Psychometric Properties of Enhanced Computer Assisted Test (ECAT) Measures (Continued)									
ECAT Psychomotor Tests									
Target Tracking 1	18	none	A	2.98	.49	.98	.74	.82	
			B	2.89	.46	.98		.80	
			G	2.73	.33	.97			
			I	2.93	.38	.97	.84	—	—
			J	—	—	.95	.76	—	—
Target Tracking 2	18	none	A	3.70	.51	.98	.85	.79	
			B	3.55	.52	.98		.76	
			G	3.58	.43	.97			
			I	3.89	.44	.97	.91	—	—
			J	—	—	.96	.73	—	—

[1]Assembling Objects, Orientation Test, and Figural Reasoning are available in both computer-administered and paper-and-pencil forms. Unless otherwise noted under "Sample", data reported here are for paper-and-pencil versions of these tests. Regardless of the mode of administration, the score on these tests is "number correct." The other tests are computer-administered. Integrating Details produces three scores: integrate time and decision time (in log units), and proportion correct. Target Identification produces two scores: decision time (in seconds) and proportion correct. Mental Counters and Sequential Memory are scored on proportion correct. The score on the two tracking tests is a measure of the extent to which the subject was off-track (i.e., mean log(distance + 1)).

[2]Sample A N=9332-9345 first tour Army personnel in 18 jobs; test-retest correlations are based a sample of 460-479 with a two-week interval (Peterson, Hough et al., 1990).
Sample B N=6754-6950 Army recruits (Peterson, Russell et al., 1990).
Sample C N=460 Navy recruits, 427 Navy recruits, and 542 high school students; test-retest correlations are the average correlations from two samples (N=127 and N=445) both with a 4 to 5 week between-testing interval (Alderton, 1989).
Sample D N=197 first tour Marines; test-retest correlations are based on a 7 to 10 day interval (Mayberry & Hiatt, 1990).
Sample E N=1267 Navy recruits; test-retest correlations are based on 220 Navy recruits (Larson, 1989).
Sample F N=202-205 new Army recruits; computer-administered versions of spatial tests were used. Means presented for Assembling Objects are based on 28 items that were common across multiple versions of the test. The time limit for Assembling Objects was based on the full version of the test, not the 28 common items (Oppler et al., 1992).
Sample G N=800 Army recruits which were divided into four groups of approximately 200. Statistics reported here are the means across the four groups of 200 (Oppler et al., 1992).
Sample H N=411 Navy recruits (Larson, 1989).
Sample I N=313 high school and junior college students (Larson & Alderton, 1992); the test-retest interval was four to five weeks. Computer-administered versions of the spatial tests were used. The Assembling Objects test was a computerized version of the test administered to Sample A.
Sample J N=1141 first tour Marine Corps personnel in two specialties; test retest correlations are based on a 10-14 day interval with 130 examinees (Carey, 1992). Carey corrected all reliability estimates for range restriction. Computer-administered versions of the spatial tests were used. The Assembling Objects test was a computerized version of the test administered to Sample A.

[3]Internal consistency reliabilities for the spatial tests are alpha coefficients. Split-half correlations are reported for Integrating Details, Target Identification, and the psychomotor tests.

[4]Internal consistency reliability minus the ASVAB adjusted R^2.

[5]Average item difficulty is the mean of the item proportion corrects.

THIS
PAGE
IS
MISSING
IN
ORIGINAL
DOCUMENT

p. 42

Figural Reasoning, ETS Syllogistic Reasoning) are indicators of G_f.[2] ECAT Mental Counters requires subjects to make rapid mental adjustments to the values of three "mental counters." The initial counter values are zero. As the test proceeds, stimuli that change the value of the counter appear. Subjects must keep track of the current value of each counter. ECAT Sequential Memory requires subjects to manipulate, in memory, the order of sets of numbers. The Army's Number Memory was modeled after a number memory test developed by Dr. Christal as part of the LAMP program at the Air Force Armstrong Laboratory (Peterson, 1987). It requires subjects to perform numeric operations progressively as instructions appear on the screen (e.g., "add 9" or "subtract 6"). The ECAT Figural Reasoning Test is a series completion test. The subject must identify the pattern or relationship among four figures and select the figure (from five alternatives) that best represents the next step in the series. Because ECAT Figural Reasoning is spatial in content it could also be classified as a measure of G_v.

Broad Visualization (G_v)

Researchers first identified a spatial factor, distinct from verbal ability, during the 1920s and 1930s. This factor underlying spatial tests (e.g., pattern perception, mazes) was called perceptual ability (Brown & Stephenson, 1933), practical ability, or simply "k" Smith (1934) reported in Smith, 1948). "Space" was a label applied by Thurstone (1938). He administered 56 tests, designed to tap a wide range of abilities, to 218 subjects. He extracted 13 factors but could only label nine: Perceptual Speed, Number, Verbal Relations, Word Fluency, Memory, Induction, Reasoning, Deduction, and Space. Five tests with the highest loadings on the Space factor were Flags, Lozenges B, Cubes, Pursuit, and Surface Development--all of which require the ability to imagine the transformation of an object or figure in space. In a separate study of eighth grade children, Thurstone and Thurstone (1941) identified seven factors: Perceptual Speed, Number, Verbal Comprehension, Word Fluency, Memory, Inductive Reasoning, and Space, with three tests (Flags, Figures, and Cards) loading on the Space factor.

In the 50 years since Thurstone's initial work, most spatial abilities research has focused on defining the number and structure of spatial subabilities rather than the existence of a broad spatial construct. Numerous studies have yielded at least one spatial factor. Three spatial factors have strong support--Visualization, Spatial Orientation, and Speeded Rotation--and several other factors have some support (Ekstrom et al., 1979; Guilford & Lacey, 1947; Hoffman, Guilford, Hoepfner, & Doherty, 1968; Lohman, 1979, 1988; McGee, 1979; Michael, Guilford, Fruchter, & Zimmerman, 1957).[3]

---

[2]Kyllonen and Christal (1990) used the ASVAB AR and MK subtests as reasoning measures. They can be classified as measures of G_f or G_q.

[3]Unfortunately, authors have not labelled factors consistently. For example, McGee (1979) refers to the factor defined by Thurstone's Flags, Figures, and Cards as Visualization, whereas Guilford and Lacey (1947) named it Spatial Relations. Lohman (1988) refers to it as Speeded Rotation and others (Ekstrom et al., 1979) have used the name Spatial Orientation. It is, therefore, very important to consider the marker tests as well as the label and definition researchers apply in defining factors. Lohman (1988) appears to have used labels that are most true to prior research efforts. In this report, his labels are used for groups of marker tests that tend to load together on factors.

Visualization, Vz, is the "ability to manipulate or transform the image of spatial patterns into other visual arrangements" (Ekstrom et al., 1979, p. 41). Visualization underlies complex spatial tasks that are relatively unspeeded, such as paper-folding, paper form board, surface development, block design, mechanical principles, and three-dimensional rotation tests. In this framework, ECAT Assembling Objects, ECAT Integrating Details, and some of the tests on the AFOQT and BAT are Visualization tests. Assembling Objects, for example, has two types of items; both types require the subject to figure out what an object will look like when its parts are put together. Half of the items are form board items, like puzzle pieces; the other half are geometric figures (e.g., squares, circles) that must be assembled in a specific way. Lohman (1988) and others (Guilford & Lacey, 1947) have noted that figural reasoning tests often load on this factor. Consistent with this observation, the ECAT Assembling Objects and ECAT Figural Reasoning tests are correlated about .55 and load together in factor solutions (Peterson, Russell et al., 1990).

Spatial Orientation (SO) involves reorienting an imagined self; that is, "subjects must imagine they are reoriented in space and then make some judgment about the situation" (Lohman, 1979, p. 188). Marker tests for Spatial Orientation include Aerial Orientation (Guilford & Lacey, 1947), ECAT Orientation, and the Project A Map tests (Peterson, Hough et al., 1990). For example, each item on the Aerial Orientation test shows a cockpit view of a shoreline. Pictures of an airplane at different altitudes are also presented. Subjects must identify the picture of the airplane that would produce the cockpit view provided. In the Map test, subjects are given a map. With each new item the subject is dropped to a new location on the map and instructed to reach a specific objective. Subjects must indicate the appropriate direction (e.g., NW, SW) to reach the objective. The ECAT Orientation test involves reorienting a picture to match a frame. This task, and most other orientation tasks like it, can be accomplished by mentally rotating parts of the object, rather than reorienting oneself. Lohman (1979, 1988) suggests that most orientation tests can also be solved with a rotation strategy.

Speeded Rotation (SR) is defined by tests such as Flags, Figures, and Cards (Thurstone & Thurstone, 1941) that involve rapidly rotating a stimulus (in the picture plane). The Project A Object Rotation test is a test of Speeded Rotation. More difficult rotation tests, involving three dimensions or rotation in the depth plane, often load with more complex tests on Visualization (Lohman, 1988). Speeded Rotation, sometimes called Spatial Relations, is probably one of the most consistently and cleanly identified spatial factors; it emerges in virtually all studies where "two dimensional" rotation tests are used. Compared to Visualization and Spatial Orientation, it is a narrow factor measuring a fairly specific ability. Within Horn's (1989) framework, Speeded Rotation, measures are more relevant to G_s than G_v, because G_v measures rely more heavily on power than speed.

At least four other spatial constructs have some factor-analytic support: Flexibility of Closure (Cf), Speed of Closure (Cs), Spatial Scanning (Ss), and Visual Memory (Vm) (Ekstrom et al., 1979; Lohman, 1988). Flexibility of Closure involves breaking one gestalt to form another, (to locate concealed figures in a distracting environment, for example). Hidden Patterns published by the Educational Testing Service (ETS) is a marker test. Speed of Closure, which sometimes combines with Cf in factor solutions, requires the ability to unify an apparently disparate perceptual field into a single percept (Ekstrom et al., 1979); ETS's Gestalt Completion is an example marker test. Spatial Scanning is marked by maze-tracing or path-finding tests and involves the ability to find an appropriate path (Ekstrom et al., 1979; Lohman, 1988). The Project A Maze Test is, for example, a Spatial Scanning test (Peterson, Hough et al., 1990). Visual memory tests require "the examinee to recognize a previously presented picture or geometric form" (Lohman, 1988, p. 188). ETS's Orientation Memory test, involves recalling the locations of buildings on a previously studied map (Ekstrom et al., 1979).

Broad Speediness (G_s)

Broad Speediness, G_s, underlies performance on all types of speeded measures including clerical or perceptual speed and visual matching tasks. According to Horn (1989), almost any task can be made into a measure of G_s by increasing speediness and decreasing knowledge and reasoning requirements. Horn also issues a caveat about the interpretation of G_s measures because physiological, emotional, or even strategical (i.e., carefulness) differences may influence performance on speeded tests more than on other cognitive measures.

In literature reviews, this factor is often called Perceptual Speed. It is sometimes grouped with spatial constructs (e.g., Lohman, 1979, 1988), sometimes placed in a domain of its own (e.g., Toquam et al., 1989), or sometimes described along with psychomotor abilities (e.g., Siegel, Federman, & Welsand, 1980). Perceptual Speed involves matching stimuli rapidly. The ECAT computerized Target Identification, the Army's Perceptual Speed and Accuracy Test (Peterson, Hough et al., 1990), the Navy's computerized Perceptual Speed test (Alderton, 1990, 1991), ETS's Identical Pictures, and the ASVAB CS and NO subtests are Perceptual Speed tests. They vary in content. ECAT Target Identification, for example, presents a target object and three stimulus objects. The objects are pictures of military vehicles or aircraft. The subject must decide which of the stimulus objects is the same as the target object. The target object is sometimes presented at an angle or on a smaller scale than the stimulus objects. The other perceptual speed tests are not figural; they involve matching alphanumeric characters.

Short Term Acquisition and Retrieval (SAR)

Short-Term Acquisition and Retrieval, SAR, is derived from information processing research. It encompasses tasks that involve sequential processing of information in short term memory. Recency memory, for example, requires recalling the most recently presented stimuli out of a string of stimuli presented in temporal order. SAR and WMC are related but not unitary constructs. Cantor, Engle, and Hamilton (1991) distinguish between short-term memory and working memory. In their view, short-term memory is a temporary storage buffer. Working memory, on the other hand, consists of a more complex and flexible space for processing information and storing the processing outcomes. While both short-term and working memory are described as limited in their capacity, working memory is more expansive than short-term memory. The argument for the separateness of these two functions is based on the assumption that if memory were a single process, having an individual focus on remembering specific information (e.g., a series of numbers) would restrict that individual's ability to carry out other mental functions simultaneously. Research demonstrating that individuals can retain information on a short-term basis without disrupting other cognitive abilities has supported the hypothesis of separate cognitive functions. Tasks that measure SAR appear to focus on recall of information, whereas WMC tasks are more complex and may require transformation or reorganization of information in short-term memory.

None of the ECAT tests are SAR measures. The Army's Project A did include one SAR measure modelled after a search task developed by Dr. R. Sternberg (Peterson, 1987). This test, Short Term Memory, displays one to five stimuli (letters or symbols). The screen clears and after a delay period a probe item is presented. The subject must decide whether the probe item was a member of the original stimulus set.

Long Term Storage and Retrieval (TSR)

Long-Term Storage and Retrieval, TSR, constructs refer to the organization of information or concepts in long-term memory and the fluency of retrieval. TSR is measured by unspeeded fluency tasks that require the individual to produce (retrieve) ideas, expressions, or words given a stimulus or given tasks that require recitation of previously learned material. Fluency measures have not been used frequently in personnel research (Toquam et al., 1989). There are no fluency measures on the ASVAB or on any of the other test batteries we reviewed.

Subgroup Differences in Cognitive Abilities

Sex Differences

Mean test score differences between males and females on general intelligence measures are generally small (Toquam et al., 1989). Differences arise on specific ability measures. Small differences favor females on reading comprehension and memory measures and favor males on numerical ability and reasoning tests (Toquam et al., 1989). Larger differences arise for perceptual speed measures (favoring females by about .40 to .50 SD) and spatial measures (favoring males by about .34 to .92 SD).

A sex difference on tests of spatial ability is a prevalent finding (Anastasi, 1958; Maccoby & Jacklin, 1974; McGee, 1979; Tyler, 1965). For example, the precursor to the current ASVAB included a Space Perception test. Kettner (1977) reported test scores for 10th, 11th, and 12th grade males and females. In total, 656 males and 576 females were included in the sample. Male means consistently exceeded female means on Space Perception; effect sizes were .32 for 10th graders, .51 for 11th graders, and .34 for 12th graders.

No doubt one of the most important recent findings is that the magnitude of the sex difference varies considerably with the type of test (Linn & Petersen, 1985; Sevy, 1983). Linn and Petersen performed a meta-analysis of standardized mean differences (effect sizes) between males' and females' scores. They grouped spatial tests into three categories: (1) spatial perception tests which included measures that correspond with the definition of Orientation above, (2) mental rotation tests which included both two- and three-dimensional rotation tests, and (3) spatial visualization tests, which included measures that correspond with the definition of Visualization given above. The spatial perception effect sizes were not sufficiently homogeneous for meta-analysis and were, therefore, partitioned by age[4]. Studies of subjects over 18 years of age produced an effect size of .64 SD favoring males, whereas the effect size for subjects under 18 years of age was .37 SD favoring males. The mental rotation effect sizes were also not homogeneous, but this time the effective partitioning variable was two- versus three-dimensional rotation tests. The effect size for two-dimensional tests was .26 SD favoring males, while the effect size for three-dimensional tasks was nearly a full standard deviation favoring males (i.e., .94 SD). The effect sizes for spatial visualization (the largest category of studies) were homogeneous. The average effect size was .13 SD; no changes in sex differences in spatial visualization were detected across age groups. A separate meta-analysis by Sevy (1983) yielded essentially the same results; three-dimensional rotation tasks produced the largest effect size, and paper form board and paper folding tasks yielded the smallest effect.

Table 15 provides effect sizes on the ECAT cognitive tests and other tests that the Services have developed recently.[5] Sex differences are generally small (under .10 SD) for the ECAT Assembling Objects, which is a Visualization test. This is consistent with Linn and Peterson's (1985) finding of about .13 difference between male and female means for Visualization tests. Differences on the other spatial tests (i.e., Orientation and Speeded Rotation tests) are in the area of one-third to one-fourth of a standard deviation difference favoring males. Differences on all tests are somewhat larger for Sample G which was considerably smaller than the other samples and therefore is likely to yield unstable results.

ECAT Figural Reasoning yields essentially no sex difference, or in the case of job incumbents (not recruits) a difference favoring females. To date, there is little information about the ECAT working memory capacity tests. Data from one small sample study yielded no sex difference on ECAT Sequential Memory and .28 SD difference favoring males on ECAT Mental Counters (Larson & Alderton, 1992). The Project A Number Memory test, also a WMC test, yields a moderate difference (.13 to .18 SD favoring males). Mean differences on Project A Short Term Memory (an SAR measure) were small, but consistently favored females.

Available information about the military's perceptual speed tests suggests that females outperform males on accuracy but that males respond more quickly to perceptual speed items. For example, the ECAT Target Identification test produces two scores: accuracy (percent correct) and decision time. Females outperform males on accuracy, but the accuracy score has little variance and is not as reliable as the time score. Males outperform females by about .50 SD on decision time (Peterson, Russell et al., 1990). The Project A Perceptual Speed and Accuracy Test

[4]"Hedges (1982) reports a statistical test for homogeneity of effect size within groups and a strategy for fitting a model to effect sizes divided into a priori classes. Hedges's homogeneity test assesses whether studies in the sample can be viewed as replicates of each other. Thus findings that studies are nearly homogeneous imply that they come close to being replications. Since studeis entering meta-analysis do differ on many dimensions, near homogeneity may be appropriate" (Linn & Petersen, 1985, p. 1481).

[5]Most of the samples shown in the table are samples of new recruits who had been selected on AFQT. To the extent that ECAT measures are correlated with AFQT, indirect range restriction will attenuate the standard deviations and inflate the means associated with both subgroups. Correcting for indirect range restriction would probably increase the standard deviation (thus, decreasing the effect size) and increase the difference between means (thus, increasing the effect size). Moreover, the net impact of indirect range restriction on the effect size is uncertain.

yields a much smaller sex differences on decision time and a difference favoring females on accuracy.

Race and Ethnic Differences

Both verbal and non-verbal I.Q. test data frequently yield well over one standard deviation difference between White and Black means and about eight to nine tenths of a standard deviation difference between White and Hispanic means. Results for Oriental and American Indian subgroups are less consistent, but indicate larger differences in verbal I.Q. than nonverbal I.Q. (Jensen, 1980).

Available data on ECAT and other tests developed by the military are summarized in Table 15, along with the sex difference effect size. The White/Black difference was larger, ranging from .60 SD to 1.08 SD, and more consistent for G_v measures than for other measures. Even so, virtually all the differences were smaller than those for the ASVAB subtests (see Chapter II). Race differences on all the G_f measures are relatively small, particularly for the working memory capacity tests. However, data for Mental Counters and Sequential Memory are based on very small samples which are likely to yield unstable results.

The two G_s measures vary greatly in content. Target Identification involves figural stimuli while the Perceptual Speed and Accuracy Test requires matching numbers and letters. Like the other tests that are figural in content, the Target Identification test yields .65 to .71 SD difference. Race differences on the Perceptual Speed and Accuracy test are negligible.

Practice Effects

Performance on spatial ability tests is to some degree malleable; test scores improve with practice (Lohman, 1988; McGee, 1979; Mittleholtz & Lohman, 1986). However, gains also occur for tests of other aptitudes. For example GATB researchers conducted 11 practice effects studies, with a total of 2783 subjects (Department of Labor, 1970). The average gain in group mean scores from the first testing to the second testing, in standard deviation units, was .55 for spatial aptitude tests and .50 for form perception tests. Average gains for other GATB aptitudes were .43 for general aptitude, .31 for verbal aptitude, .35 for numeric aptitude, and .55 for clerical perception. There is also some evidence that gains from practice are larger for speeded tests than for power tests (Lohman, 1988; Dunnette, Corpe, & Toquam, 1987) and that individuals throughout the ability range benefit from practice (e.g., good performers improve as much as poor performers) (Oppler et al., 1992).

Table 15

Sex, Race, and Retest Effects on Cognitive Tests

Major Constructs	Military Marker Tests	Sample[1]	Effect Size[2] Male/Female	White/Black	Retest
G_f - Broad Reasoning	**ECAT Figural Reasoning**				
		A	-.17	.77	.15
		B	-.01	.77	.27
		C	.00	.76	—
		G	.13	—	.22
	ECAT Mental Counters				
		E	—	.59	—
		G	.28	—	.08
	ECAT Sequential Memory				
		F	—	.22	—
		G	-.01	—	.38
	Project A Number Memory				
		B	.18	.50	.44
		C	.13	.47	—
G_v - Broad Visualization	**ECAT Assembling Objects**				
		A	-.02	.78	.28
		B	.06	.84	.08
		C	.08	.83	—
		G	.17	—	.06
	ECAT Integrating Details				
	Proportion Correct	D	—	.86	—
		G	.48	—	.08
	ECAT Orientation				
		A	.23	.82	.11
		B	.35	.91	.27
		C	.33	.77	—
		G	.48	—	.38
	Project A Map Test				
		A	.28	1.03	.08
		B	.30	1.08	.08
		C	.29	.98	—
	Project A Object Rotation				
		A	.31	.71	.57
		B	.21	.69	.33
		C	.28	.60	—
G_s - General Speediness	**ECAT Target Identification**				
	Decision Time	A	.52	.71	.46
		B	.47	.66	.32
		C	.47	.65	—
		G	.64	—	.54
	Proportion Correct	A	-.12	.13	-.19
		B	-.01	.23	.03
		C	.00	.14	—

(Continued)

Table 15					
Sex, Race and Retest Effects on Cognitive Tests (Continued)					
			Effect Size[2]		
Major Constructs	Military Marker Tests	Sample[1]	Male/Female	White/Black	Retest
G_s - General Speediness (Cont.)	Project A Perceptual Speed and Accuracy				
	Decision Time	A	.09	.02	.35
		B	.19	.04	.08
		C	.18	.03	---
	Proportion Correct	A	-.37	.24	.05
		B	-.33	.13	.11
		C	-.36	.12	---
SAR - Short Term Acquisition & Retrieval	Project A Short Term Memory				
		B	-.05	.19	.15
		C	-.11	.21	---

[1]Sample A- Total N = 9332 - 9345 first tour Army enlisted personnel in 18 jobs; the sample included more than 8100 males and 770 females, 2300 Blacks and 5900 Whites (Peterson, Hough et al., 1990). Retest effects are based on a sample of 100 with a two-week interval between testing sessions (Peterson, 1987).

Sample B- Total N = 6754 - 6950 Army recruits (Peterson, Russell et al., 1990); the sample included more than 5900 males and 800 females, 1650 Blacks and 4740 Whites. Retest effects are based on a sample of 473 with a one-month interval between testing sessions (Toquam, Peterson, Rosse, Ashworth, Hanson, & Hallam, 1986).

Sample C- Total N = 6435 Army recruits (Peterson, Russell et al., 1990); the sample included more than 5900 males and 770 females, 1677 Blacks and 4670 Whites.

Sample D- Total N = 1254 Navy recruits (Alderton, 1989); the sample included 826 Whites and 242 Blacks.

Sample E- Total N = 1267 Navy recruits; 240 Blacks and 829 Whites (Larson, 1989).

Sample F- Total N = 377 Navy recruits 283 Whites and 40 Blacks (Larson, 1989).

Sample G- Total N = 300 High School Students; 86 females and 205 males (Larson & Alderton, 1992). Retest effects are based on four- to five-week intervals between testing sessions.

[2] The effect size is the standardized mean difference between two subgroups' mean scores [$d = (MN_{group1} - MN_{group2})/S_p$, where S_p is the pooled standard deviation]. A positive Male/Female effect size indicates superior performance by males, and a negative effect size indicates superior performance by females. Similarly, a positive White/Black effect size indicates that White mean score was higher than the Black mean score, and a negative score indicates that the Black mean score was higher than the White mean score. A positive retest effect indicates a gain with practice on the test.

Even though test performance improves with practice, attempts to train spatial ability have produced modest gains at best (Brinkman, 1966; Kyllonen, Lohman, & Snow, 1984; McGee, 1979). Training gains, when realized, usually generalize to tasks closely related to the training intervention, not necessarily to other spatial tasks (Levine, Brahlek, Eisner, & Fleishman, 1979; Levine, Schulman, Brahlek, & Fleishman, 1980). However, Embretson (1987) observed a significant difference between pre- and post- training scores on the Differential Aptitude Test (DAT Forms S and T) after training in text editing. Post-training DAT scores were more internally consistent and yielded increased predictive validity over pre-training scores. In most studies, it is difficult to ascertain whether performance improvement reflects alteration of ability, greater familiarity with instructions and item types, or development of individual strategies for dealing with spatial problems.

To what extent are the ECAT cognitive tests susceptible to practice effects? There are some gains on most ECAT cognitive tests at retest, even without a practice intervention (see Table 15). The gains reported for each test vary considerably because (a) gains for samples B and G in Table 15 were based on a one-month interval between testing sessions while there were two weeks between sessions for sample A and (b) samples B and G were considerably larger than sample A. The mean gain score is not dependent on the sample size, but it is less stable when N is small than when N is large. Even so, a few findings are consistent. A gain of about one quarter of a standard deviation after a one-month interval has been observed for ECAT Figural Reasoning in two samples. Slightly larger gains, .27 SD and .38 SD, were reported for the Orientation test in the two studies. The largest gains for any of the ECAT cognitive tests, were .32 SD and .54 SD for Target Identification decision time after a one-month interval. One test is anomalous. ECAT Assembling Objects yielded gains of only .08 SD and .06 SD with a one-month between test interval.

A few recent studies found that practice and/or coaching alters test scores. Oppler et al. (1992) administered Target Identification repeatedly, five times, with a one minute break between administrations to examine the immediate effect of extreme practice. Decision time scores improved dramatically, 1.57 SD, and proportion correct increased by .12 SD. It is possible that subjects were memorizing the items and responses, rather than learning how to solve Target Identification problems because subjects received the same items in each replication. Most of the gain was achieved over the course of the first two administrations of the test.

Busciglio and Palmer (1992) studied the effects of practice and coaching on ECAT Assembling Objects, Figural Reasoning, and Orientation test scores. The subjects, 1914 Army receptees, were assigned to one of five treatments: (1) specific coaching with practice, (2) specific coaching without practice, (3) general coaching with practice, (4) general coaching without practice, and (5) practice. Subjects in specific coaching treatments were told about specific strategies that could be used to perform a certain test more effectively. Subjects in general coaching conditions received broad instructions for improving performance on multiple choice tests. Subjects in sessions with practice practiced by taking the test twice. Practice effects were significant for all three tests. General coaching was ineffective for all groups. There was a significant specific coaching-practice interaction for the Assembling Objects and Figural Reasoning tests; the effect of coaching was much less pronounced when subjects had practice. Coaching did not add much to practice. For the Orientation test, however, the coaching and practice effects were additive. Specific coaching was highly effective for this test, suggesting that subjects can adopt a simple strategy that alters test performance.

In sum, it is a good idea to include practice items on spatial tests since spatial items are probably less familiar to people than verbal and numerical items. There are some gains on most ECAT cognitive tests at retest, even without a practice intervention; however gains for a few tests (e.g., Assembling Objects) are negligible. ECAT Assembling Objects and Figural Reasoning, though affected by practice, do not appear to be coachable; ECAT Orientation does appear to be coachable.

Validation Results for Cognitive Measures

Cognitive measures are valid predictors for virtually all jobs (Ghiselli, 1973; Hunter, 1986). Toquam et al. (1989) conducted a meta-analysis of validities for cognitive predictors to identify measures likely to supplement the ASVAB. Virtually all available studies published by 1983 were reviewed, and studies involving young children or college students were excluded.

Toquam et al. (1989) arranged validities according to the type of criterion as well as the type of predictor and type of job, as shown in Table 16. Jobs were organized into a taxonomy derived from the Dictionary of Occupational Titles scheme (Department of Labor, 1977). The major categories were: (1) professional, technical, and managerial jobs including military officers and aircrew as well as civilian managers and professionals; (2) clerical, including military and civilian office clerk and administrative jobs; (3) protective services, subsuming jobs like military police, infantryman, corrections officer; (4) service, comprising food and medical service jobs; (5) mechanical/structural maintenance, covering all mechanical and maintenance jobs; (6) electronics, including electricians, radio operators, radar and sonar technicians; and (7) industrial, covering jobs such as machine operator and coal miner.

Criteria were organized into four categories: (1) education, including course grades and instructor evaluations; (2) training, composed of exam scores, course grades, instructor ratings, work sample and hands-on measures; (3) job proficiency, including supervisor ratings, job knowledge measures, and archival measures, and (4) adjustment, referring to measures of delinquency such as disciplinary actions (e.g. Article 15) and discharge conditions.

Definitions of the predictor constructs were: (a) spatial ability, including measures of spatial visualization, two- and three-dimensional rotation, and spatial scanning; (b) perceptual speed and accuracy, including measures that involve performing simple processing tasks quickly and accurately; (c) verbal, subsuming word, verbal, and reading comprehension measures; (d) reasoning, containing tests of induction, deduction, analogical reasoning, figural reasoning, and word problems; (e) number facility, including both simple and complex arithmetic and mathematics tests; (f) memory, containing measures of recall, memory span, and visual memory; (g) perception, including speed and flexibility of closure measures; and (h) fluency, subsuming measures of associational, expressional, ideational, and word fluency.

Two predictor categories contained tests that are primarily spatial in nature: spatial and perception. Spatial measures were effective predictors of training criteria in virtually all jobs, but particularly for electronics jobs (median $r = .49$). Perception measures best predicted training outcomes. The relationships between education and training criteria and cognitive measures overall was higher than the relationship between job proficiency criteria and cognitive variables.

Perceptual speed and accuracy tests predicted education and training criteria, and perception measures best predicted training outcomes. Validities for verbal, reasoning, and number facility measures were relatively uniform across all job types.

Memory and Fluency measures are notable in that they have been used less frequently than other measures in validity studies. Validity data that are available suggest that Fluency measures might be better predictors for professional, technical, and managerial jobs than other jobs. Memory tests, on the other hand, have been useful predictors for service jobs.

Please note that the validities reported in Table 16 are not corrected for artifacts. Range restriction and measurement error create artifactual variation across studies (Hunter & Schmidt, 1990). The results presented here, therefore, can be misleading. Application of meta-analytic methods to the Toquam et al. (1989) data would be a highly useful project.

Table 18
Uncorrected Median Validities for Cognitive Ability Constructs

Cognitive Abilities & Criterion Categories	Professional Technical & Managerial		Clerical		Protective Services		Service		Mechanical Structural Maintenance		Electronics		Industrial		Miscellaneous		All Job Types	
Spatial																		
Education	.24	(16)	.26	(3)	–	–	.35	(4)	.15	(26)	.15	(5)	.60	(1)	.24	(1)	.21	(56)
Training	.34	(34)	.35	(4)	.04	(15)	.26	(14)	.24	(32)	.49	(6)	.24	(5)	–	–	.26	(110)
Job Proficiency	.18	(12)	.08	(40)	-.04	(16)	.25	(25)	.17	(71)	.15	(77)	.29	(41)	.17	(39)	.16	(321)
Adjustment	–	–	–	–	–	–	–	–	.04	(3)	–	–	–	–	–	–	.04	(3)
Overall	.26	(62)	.11	(47)	.00	(31)	.26	(43)	.16	(132)	.16	(88)	.29	(47)	.17	(40)	.19	(490)
Perception																		
Education	.17	(16)	.36	(1)	–	–	.38	(1)	.16	(6)	.03	(3)	.41	(1)	.16	(1)	.17	(29)
Training	.31	(33)	.14	(15)	.08	(12)	.23	(6)	.23	(13)	.36	(13)	.36	(3)	–	–	.25	(97)
Job Proficiency	.12	(11)	.11	(22)	.10	(19)	.24	(13)	.13	(35)	.17	(16)	.28	(43)	–	–	.18	(159)
Adjustment	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Overall	.26	(60)	.13	(38)	.09	(31)	.24	(22)	.16	(54)	.18	(32)	.29	(47)	.16	(1)	.20	(285)
Perceptual Speed and Accuracy																		
Education	.16	(14)	.34	(3)	–	–	.36	(4)	.18	[illegible]	.19	(5)	.36	(1)	.30	(1)	.20	(52)
Training	.18	(26)	.16	(97)	.20	(47)	.21	(39)	.15	(113)	.11	(79)	.31	(5)	–	–	.16	(406)
Job Proficiency	.19	(12)	.16	(194)	.08	(23)	.13	(36)	.14	(56)	.05	(104)	.24	(44)	.10	(2)	.13	(470)
Adjustment	–	–	–	–	–	–	–	–	.20	(1)	–	–	–	–	–	–	.20	(1)
Overall	.18	(52)	.16	(294)	.15	(70)	.18	(79)	.15	(194)	.08	(188)	.25	(50)	.17	(3)	.15	(929)
Verbal																		
Education	.26	(16)	.38	(3)	–	–	.29	(4)	.20	(29)	.27	(10)	.24	(2)	.01	(1)	.24	(65)
Training	.16	(9)	.30	(52)	.24	(28)	.34	(14)	.33	(94)	.35	(70)	.35	(5)	–	–	.31	(272)
Job Proficiency	.31	(18)	.19	(153)	.15	(6)	.21	(25)	.18	(48)	.24	(76)	.20	(66)	.21	(19)	.21	(409)
Adjustment	–	–	–	–	–	–	–	–	.02	(1)	–	–	–	–	–	–	.02	(1)
Overall	.26	(43)	.22	(208)	.22	(34)	.26	(43)	.27	(170)	.29	(156)	.21	(73)	.20	(20)	.25	(747)
Reasoning																		
Education	.30	(30)	.30	(2)	–	–	.32	(3)	.31	(11)	.23	(2)	–	–	–	–	.30	(48)
Training	.33	(28)	.28	(79)	.32	(31)	.30	(51)	.22	(109)	.32	(101)	–	–	.14	(8)	.28	(407)
Job Proficiency	.38	(19)	.14	(63)	.16	(8)	.12	(19)	.09	(36)	.11	(107)	.22	(18)	.18	(1)	.14	(281)
Adjustment	–	–	–	–	–	–	–	–	.01	(1)	–	–	–	–	–	–	.01	(1)
Overall	.35	(77)	.21	(164)	.29	(39)	.25	(73)	.20	(157)	.21	(210)	.22	(18)	.14	(9)	.23	(737)
Number Facility																		
Education	.45	(16)	.54	(1)	–	–	.22	(1)	.34	(12)	.35	(9)	.35	(2)	.29	(1)	.38	(42)
Training	.33	(19)	.31	(15)	.19	(12)	.32	(17)	.32	(22)	.22	(11)	.38	(5)	.14	(4)	.29	(105)
Job Proficiency	.23	(13)	.21	(158)	.25	(1)	.25	(20)	.18	(40)	.16	(43)	.27	(65)	.09	(1)	.21	(341)
Adjustment	–	–	–	–	–	–	–	–	-.02	(1)	–	–	–	–	–	–	-.02	(1)
Overall	.34	(48)	.22	(174)	.19	(13)	.24	(38)	.24	(75)	.20	(63)	.28	(72)	.16	(6)	.34	(489)
Memory																		
Education	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Training	.11	(2)	.28	(2)	.21	(8)	.21	(1)	.11	(1)	–	–	–	–	–	–	.20	(14)
Job Proficiency	–	–	.12	(25)	.04	(9)	.20	(7)	.08	(10)	.07	(38)	.24	(3)	–	–	.10	(92)
Adjustment	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Overall	.11	(2)	.13	(27)	.12	(17)	.20	(8)	.07	(11)	.07	(38)	.24	(3)	–	–	.16	(106)
Fluency																		
Education	.21	(6)	–	–	–	–	–	–	.02	(2)	–	–	–	–	–	–	.16	(8)
Training	.86	(1)	–	–	–	–	–	–	–	–	.20	(2)	–	–	–	–	.42	(3)
Job Proficiency	42	(1)	.05	(19)	–	–	.03	(6)	.06	(12)	.04	(34)	.19	(1)	–	–	.05	(73)
Adjustment	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Overall	.31	(8)	.05	(19)	–	–	.03	(6)	.05	(14)	.05	(36)	.19	(1)	–	–	.07	(84)

Note. From "Literature review: Cognitive abilities - theory, and validity" (ARI Research Note 91-28) by J. L. Toquam, V. A. Corpe, and M. D. Dunnette, 1989, Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

The number of validity coefficients used to compute the median validity is provided in parentheses.

Validity Evidence Relevant to Working Memory Capacity Tests

Although numerous studies have investigated the validity of reasoning tests for predicting job or training success, few have focused on WMC, probably because WMC is a relatively new construct. Researchers at the Air Force Armstrong Laboratory's Learning Abilities Measurement Program (LAMP) have pioneered measurement in this arena, but LAMP typically conducts basic, not applied, research. In one study, however, Christal (1991) examined the incremental validity of WMC, processing speed, and processing accuracy over the ASVAB for predicting learning performance. Subjects received computer-assisted instruction in solving logic problems. Criteria were latency and accuracy scores on blocks of problem-solving trials, during and after instruction. Conceptually, criteria measured two aspects of learning performance: the acquisition of declarative knowledge (measured by accuracy in solving problems) and the development of procedural skill (measured by the time required to solve problems). Analyses indicated that the LAMP tests added about 20% unique valid variance to the ASVAB tests in predicting the learning criteria.

Validity Evidence on ECAT Measures

What validity evidence exists for the ECAT measures? Alderton (1989a) found ECAT Integrating Details to be an effective predictor of job performance in a small sample study. The uncorrected correlation between the Integrating Details accuracy score and a hands-on performance criterion for electronics technicians was r=.22 (N=94). The uncorrected correlations between the accuracy score and two work samples for aviation sonar technicians produced rs of .38 and .46 (N=30). These uncorrected correlations are likely to be underestimates due to indirect range restriction, since participants had be selected on the ASVAB.

ECAT cognitive measures that were also part of the Army's Project A/Career Forces Project have demonstrated incremental validity over the ASVAB for predicting technical and hands-on job performance. In Project A, McHenry et al. (1990) combined six Project A[6] figural tests (including ECAT Assembling Objects, ECAT Orientation, and ECAT Figural Reasoning) to form one composite score.[7] Mean validity coefficients[8] were .56 for the core technical proficiency criterion and .63 for the general soldiering proficiency criterion, which subsume job knowledge and hands-on task proficiency measures. Mean validities for the remaining criterion constructs were: .25 for effort and leadership, .12 for personal discipline, and .10 for physical fitness and military bearing. The highest incremental validity of the spatial composite (beyond

---

[6]Six paper-and-pencil spatial tests, six computerized perceptual tests, and four psychomotor tests were administered to more than 9,000 first tour Army enlisted personnel in the Project A concurrent validation study. Criteria included hands-on task proficiency measures, job knowledge tests, archival data, and a variety of peer and supervisory rating materials.

[7]As mentioned, before Assembling Objects and Figural Reasoning tests were good markers for the general spatial factor in the Project A/Career Forces spatial tests (Peterson, Russell et al., 1990).

[8]Mean validities were computed across 18 Army enlisted jobs. Validities were corrected for range restriction and were adjusted for shrinkage.

that afforded by the ASVAB) was .03 for general soldiering proficiency. In the longitudinal sister study to Project A, the Career Force Project, Oppler, Peterson, and Russell (in press) report results that concur with the earlier Project A work (McHenry et al., 1990). As in Project A, the six Project A figural tests (including ECAT Assembling Objects, ECAT Orientation, and ECAT Figural Reasoning) were combined to form one composite score. The spatial composite added .01 and .02 incremental validity over the ASVAB for predicting technical aspects of job performance.

Mayberry and Hiatt (1990) administered the ASVAB Form 6 Space Perception, ECAT Figural Reasoning, ECAT Assembling Objects, a video firing test, and the Armed Services Applicant Profile (ASAP) to more than 1300 first tour Marines in four jobs. Criteria included a hands-on performance test, a job knowledge test, proficiency marks, and training school grades. ECAT Assembling Objects was the best new predictor of the job knowledge criterion; corrected incremental validities were .02 for all four jobs. The video firing test and the ASAP provided the best incremental validity for the remaining criteria.

Carey (1992) examined incremental validities (over the ASVAB) for several of the ECAT tests. Examinees were 698 first-term Marine Corps automotive mechanics and 443 helicopter mechanics who were tested as part of the Job Performance Measurement project. ECAT Assembling Objects added the most incremental validity to the ASVAB for predicting the hands-on performance criterion in both the automotive and helicopter mechanic samples.

Wolfe, Alderton, and Larson (1992) conducted a study involving 4989 Navy recruits assigned to nine technical training schools. Memory and spatial predictors (including ECAT Integrating Details, ECAT Figural Reasoning, ECAT Mental Counters, ECAT Sequential Memory, and ASVAB Form 6 Space Perception) were validated against final school grades and scores on laboratory exercises. Fully corrected mean validities were .49 for Integrating Details, .48 for Figural Reasoning, .41 for Space Perception, .43 for Mental Counters, and .39 for Sequential Memory. Fully corrected incremental validities (over the ASVAB) were significant for four, three, and two technical schools for Integrating Details, Space Perception, and Figural Reasoning respectively. Mental Counters added to prediction for two schools, and Sequential Memory supplemented the ASVAB significantly for prediction of criteria in three schools.

In sum, cognitive measures predict performance in technical aspects of job and training performance. ECAT measures have been effective predictors of technical proficiency criteria (e.g., hands-on scores, training test scores) in several studies. Unfortunately, none of the previous studies have included all of the ECAT measures in one battery, making comparisons among tests difficult. Additional validity data are currently being collected and analyzed for all the ECAT measures in a Joint-Service research project. Meta-analyses of ECAT and cognitive test validities are needed to enlighten future predictor development efforts.

Psychomotor Attributes

Psychomotor Attribute Definitions

Psychomotor abilities involve the execution of motor responses such as manipulative, repetitive, and precise limb movements (Imhoff & Levine, 1981). As with spatial ability, much of what we know today about psychomotor abilities stems from research on aircrew performance. During World War II, the Army Air Force (AAF) studies made great strides toward defining and measuring psychomotor abilities (Guilford & Lacey, 1947). Fleishman and his colleagues continued psychomotor abilities research in the 1950s and early 1960s (e.g., Fleishman, 1967, 1972; Fleishman & Hempel, 1954a, 1954b, 1955, 1956). Fleishman performed a series of factor analytic studies with military airmen and airmen trainees to identify the basic structure of the psychomotor domain (Fleishman, 1954; Fleishman & Ellison, 1962; Fleishman & Hempel, 1954a, 1955, 1956). Definitions of the 11 psychomotor abilities Fleishman and his colleagues identified appear in Table 17. ECAT Target Tracking 1 is a measure of Control Precision, and ECAT Target Tracking 2 is a measure of Multilimb Coordination.

Siegel et al. (1980) reviewed psychomotor and perceptual-motor abilities literature encompassing studies of children's motor skills as well as aircrew measurement research and identified 61 abilities. They rated the abilities against several criteria (e.g., scalability, reliability, validity) and concluded that 13 perceptual/psychomotor abilities had strong support. Five abilities were psychomotor in nature and were subsumed by Fleishman's 11 constructs: Control Precision, Manual Dexterity, Finger Dexterity, Multilimb Coordination, and Rate Control (tracking). The perceptual abilities were: Visual Speed and Accuracy, Position Memory, Auditory Discrimination, Auditory Memory, Clerical Perception, Perception of Size and Form, and Depth Perception.

The authors of more recent psychomotor abilities meta-analyses and reviews have concluded that Fleishman's original structure has remained salient over the years (Bosshardt, 1987; McHenry & Rose, 1988). Even so, the authors suggest that Reaction Time and Response Orientation involve little motor skill and should be included with perceptual and cognitive abilities. Moreover, there appears to be consensus on five to nine of Fleishman's original 11 abilities.

Recent works have focused on hierarchical models of psychomotor abilities--models that are compatible with Fleishman's taxonomy. In an extensive review of the psychomotor, perceptual, and cognitive ability literature, Imhoff and Levine (1981) proposed two higher-order dimensions of Fleishman's psychomotor ability factors: (1) Basic Movement Speed and Accuracy and (2) Perceptual-Motor Movement Control. Basic Movement Speed and Accuracy includes Fleishman's Control Precision, Speed of Arm Movement, and Reaction Time abilities--abilities that are highly structured and require speed and accuracy with little processing. Fleishman's Multilimb Coordination, Response Orientation, and Rate Control are subsumed by Perceptual-Motor Movement Control. This category of abilities requires continuously or periodically adjusting movements in response to sensory or perceptual feedback. McHenry (1987) extended Imhoff and Levine's (1981) work, adding a third second-order dimension, Dexterity, to include manual and finger dexterity. He also posited a general factor underlying all tests of psychomotor ability.

Table 17 Psychomotor Abilities
Multilimb Coordination--The ability to coordinate the movements of a number of limbs simultaneously, and is best measured by devices involving multiple controls (e.g., two-hand coordination tests).
Rate Control--This ability involves the timing of continuous anticipatory motor adjustments relative to changes in speed and direction of a continuously moving target or object.
Control Precision--The ability to make rapid, precise, highly controlled, but not overcontrolled, movements necessary to adjust or position a machine control mechanism (e.g., rudder controls). Control precision involves the use of larger muscle groups, including arm-hand and leg movements.
Speed of Arm Movement--The ability to make gross, discreet arm movements quickly in tasks that do not require accuracy.
Manual Dexterity--This ability involves skillful, well-directed arm-hand movements in manipulating fairly large objects under speeded conditions.
Finger Dexterity--The ability to make skillful, controlled manipulations of tiny objects involving, primarily, the fingers.
Arm-Hand Steadiness--The ability to make precise arm-hand positioning movements where strength and speed are minimized; the critical feature is the steadiness with which movements must be made.
Wrist, Finger Speed (also called tapping)--This ability is very narrow. It involves making rapid discrete movements of the fingers, hands, and wrists, such as in tapping a pencil on paper.
Aiming (also called eye-hand coordination)--This ability is very narrow. It involves making precise movements under highly speeded conditions such as in placing a dot in the middle of a circle, repeatedly, for a page of circles.
Response Orientation--The ability to select the correct movement in relation to the correct stimulus, especially under highly speeded conditions (e.g., Choice Reaction Time tests).
Reaction Time--The ability to respond to a stimulus rapidly.

Note. From "Performance assessment based on an empirically derived task taxonomy" by E. A. Fleishman, 1967, Human Factors, 9.

Subgroup Differences

Sex Differences

The magnitude and direction of sex differences varies considerably across psychomotor factors. Synk (1984) reported sex differences for more than 12,000 males and 13,000 females on three GATB psychomotor aptitudes: motor coordination (Fleishman's Wrist, Finger Speed), finger dexterity, and manual dexterity. Females outperformed males on motor coordination (.46 SD) and finger dexterity (.22 SD). Males and females did not differ on manual dexterity. In contrast, McHenry and Rose (1988) report that sex differences are typically very large (i.e., one standard deviation) and usually favor males for tests of multilimb coordination.

Both ECAT psychomotor tests yield large consistent sex differences. Three large samples, each including more than 5900 males and 770 females, data have yielded 1.00 to 1.28 SD difference between means, with males scoring higher than females (Peterson, Russell et al., 1990). Larson and Alderton (1992) reported similar effects for a sample of 205 male and 86 female high school students. Differences were 1.25 SD for Tracking 1 and 1.52 SD for Tracking 2, favoring males.

Race/Ethnic Differences

White to non-White differences are typically smaller and less consistent for psychomotor abilities than are such differences for cognitive abilities (McHenry & Rose, 1988). Differences in means range from two-tenths to one-half of a standard deviation (with Whites' means higher than Blacks' means) on tests of finger dexterity, manual dexterity, wrist-finger speed, and multilimb coordination.

The Army's Project A and Career Forces large-scale data collections provided information on race and ethnic differences on the ECAT psychomotor measures (Peterson, Russell et al., 1990). In all samples, Whites' means were higher than Blacks' and Hispanics' means. Black/White differences in means have ranged from two-thirds to three-quarters of a standard deviation on ECAT Target Tracking 1 and from eight-tenths to nine-tenths of a standard deviation on ECAT Target Tracking 2. Hispanic/White differences on both tests were about one-quarter of a standard deviation.

Practice Effects

Improvement with practice on psychomotor measures is a common finding (McHenry & Rose, 1988). For example, GATB researchers conducted 11 practice effects studies, with a total of 2783 subjects (Department of Labor, 1970). The average gain in group mean scores from the first testing to the second testing, in standard deviation units, was .81 for finger dexterity, .91 for manual dexterity, and .45 for motor coordination, compared to effect sizes ranging from .31 to .55 for the cognitive GATB aptitudes.

To what extent are the ECAT psychomotor tests susceptible to practice effects? McHenry, Toquam, Rosse, Peterson, & McGue (1987) conducted a practice effects study. Pre- and post-practice testing occurred two weeks apart; practice included retesting on new items and occurred about one week after the initial test. A control group also took the pre- and post-tests. Gains in standard deviation units for the practice group (N=74) were .33 SD for Target Tracking 1 and .21 SD for Target Tracking 2. The control group (N=113) improved slightly on Target Tracking 1 (.07 SD), but performance deteriorated on Target Tracking 2 (-.09 SD).

Toquam et al. (1986) retested 473 subjects after a one-month interval (without practice). They reported gains of .27 SD on Target Tracking 1 and .24 SD on Target Tracking 2. Test-retest reliabilities were .74 and .85 for Target Tracking 1 and Target Tracking 2 respectively. Oppler et al. (1992) administered Target Tracking 2 repeatedly, five times, with a one minute break between administrations to examine the immediate effect of extreme practice. Scores improved dramatically--1.00 standard deviation. Although the items were exactly the same across all five trials, the effect cannot mean that subjects simply learn the correct response because there are no "correct" or "incorrect" responses on these tests. Most of the gain was achieved over the course of the first two administrations of the test, and individuals throughout the ability range benefited from practice (i.e., good performers improve as much as poor performers).

Validity of Psychomotor Measures

McHenry & Rose (1989) conducted a meta-analysis of psychomotor predictors. They organized validities for tests according to Fleishman's classification scheme, type of criterion, and type of job.[9] The results appear in Table 18.[10] The bulk of the validation studies were conducted using GATB subtests, as evidenced by the large number of validities reported for GATB aptitudes: Finger Dexterity, Manual Dexterity, and Wrist-Finger Speed. Conversely, measures of Control Precision, Rate Control, Aiming, Arm-Hand Steadiness, and Speed of Arm Movement have rarely been used.

As might be expected, measures of Multilimb Coordination have been effective predictors of criteria for professional, technical, and managerial jobs (which subsume pilots and aircrew) and protective service jobs (which include infantry and military police jobs). In contrast, Finger Dexterity, Manual Dexterity, and Wrist-Finger Speed predictors were most relevant to performance in industrial jobs (e.g., assembler, bench worker, machine operator).

ECAT Psychomotor Test Validity

There is evidence that the ECAT psychomotor tests predict proficiency in military enlisted jobs. McHenry et al. (1990) formed six composites of Project A[11] psychomotor and perceptual test scores (including ECAT Tracking 1 and ECAT Tracking 2). Mean validity coefficients[12] for the combination of six composites were .53 for the core technical proficiency criterion and

[9]McHenry and Rose used the criterion and type of job definitions that are provided previously under "Cognitive Attributes" for the review by Toquam et al. (1989).

[10]Please note that the validities reported in Table 18 are not corrected for artifacts. Range restriction and measurement error create artifactual variation across studies (Hunter & Schmidt, 1990). The results presented here, therefore, can be misleading.

[11]Six paper-and-pencil spatial tests, six computerized perceptual tests, and four psychomotor tests were administered to more that 9000 first tour Army enlisted personnel in the Project A concurrent validation study. Criteria included hands-on task proficiency measures, job knowledge tests, archival data, and a variety of peer and supervisory rating materials.

[12]Mean validities were computed across 18 Army enlisted jobs. Validities were corrected for range restriction and were adjusted for shrinkage.

Table 18

Uncorrected Median Validities for Psychomotor Constructs

Psychomotor Abilities and Criterion Categories	Professional Technical & Managerial	Clerical	Sales	Protective Services	Service	Mechanical/Structural Maintenance	Electronics	Industrial	Miscellaneous	All Job Types
Multilimb Coordination										
Education	.24 (1)	—	—	—	—	—	—	—	—	.24 (1)
Training	.20 (240)	—	—	.20 (38)	—	—	—	—	.18 (4)	.20 (282)
Job Proficiency	—	—	—	.15 (11)	.08 (6)	—	—	—	—	.14 (17)
Job Involvement	—	—	—	—	—	—	—	—	—	—
Overall	.21 (241)	—	—	.20 (49)	.08 (6)	—	—	—	.18 (4)	.20 (300)
Control Precision										
Education	—	—	—	—	—	—	—	—	—	—
Training	.19 (24)	—	—	.14 (7)	—	—	—	—	.02 (1)	.17 (32)
Job Proficiency	—	—	—	—	—	—	—	—	—	—
Job Involvement	—	—	—	—	—	—	—	—	—	—
Overall	.19 (24)	—	—	.14 (7)	—	—	—	—	.02 (1)	.17 (32)
Rate Control										
Education	—	—	—	—	—	—	—	—	—	—
Training	.08 (17)	—	—	.05 (4)	—	—	—	—	.02 (1)	.08 (22)
Job Proficiency	—	—	—	—	-.02 (6)	—	—	—	—	-.02 (6)
Job Involvement	—	—	—	—	—	—	—	—	—	—
Overall	.08 (17)	—	—	.05 (4)	-.02 (6)	—	—	—	.02 (1)	.08 (28)
Finger Dexterity										
Education	.11 (57)	.27 (2)	—	—	.22 (12)	.18 (18)	.17 (3)	.28 (11)	.16 (18)	.16 (121)
Training	.07 (59)	.08 (1)	—	.09 (16)	.08 (6)	.25 (6)	—	.11 (7)	.28 (2)	.08 (97)
Job Proficiency	.13 (47)	.12 (44)	.08 (5)	.07 (6)	.13 (33)	.21 (39)	.15 (7)	.24 (254)	.20 (6)	.19 (443)
Job Involvement	—	—	—	—	—	—	—	.14 (1)	—	.14 (1)
Overall	.10 (163)	.13 (47)	.08 (5)	.09 (22)	.14 (51)	.20 (63)	.16 (10)	.24 (273)	.18 (26)	.16 (662)
Manual Dexterity										
Education	.07 (57)	.32 (2)	—	—	.17 (12)	.20 (16)	.04 (3)	.28 (14)	-.02 (18)	.10 (124)
Training	.20 (3)	.13 (1)	—	—	.29 (2)	.20 (4)	—	.17 (7)	.38 (1)	.19 (18)
Job Proficiency	.16 (47)	.12 (44)	.08 (5)	.15 (5)	.18 (33)	.21 (40)	.12 (7)	.24 (258)	.24 (9)	.22 (448)
Job Involvement	—	—	—	—	—	—	—	.02 (3)	—	.02 (3)
Overall	.12 (107)	.13 (47)	.08 (5)	.15 (5)	.18 (47)	.20 (62)	.08 (10)	.24 (282)	.10 (28)	.19 (593)
Wrist-Finger Speed										
Education	.10 (100)	.32 (2)	—	—	.22 (12)	.18 (18)	.12 (3)	.12 (8)	.16 (18)	.13 (161)
Training	.09 (3)	.25 (1)	—	—	.08 (4)	.08 (3)	—	.16 (4)	.34 (1)	.12 (16)
Job Proficiency	.14 (48)	.14 (46)	.14 (5)	.23 (5)	.20 (33)	.16 (38)	.16 (7)	.22 (244)	.30 (9)	.18 (435)
Job Involvement	—	—	—	—	—	—	—	—	—	—
Overall	.11 (151)	.14 (49)	.14 (5)	.23 (5)	.20 (49)	.16 (59)	.14 (10)	.21 (256)	.22 (28)	.17 (612)
Aiming										
Education	—	—	—	—	—	—	—	—	—	—
Training	.13 (7)	—	—	—	—	—	—	—	—	.13 (7)
Job Proficiency	—	—	—	—	—	—	—	—	—	—
Job Involvement	—	—	—	—	—	—	—	—	—	—
Overall	.13 (7)	—	—	—	—	—	—	—	—	.13 (7)
Arm-Hand Steadiness										
Education	—	—	—	—	—	—	—	—	—	—
Training	.07 (25)	—	—	.08 (3)	—	—	—	—	—	.08 (28)
Job Proficiency	—	—	—	—	—	—	—	—	—	—
Job Involvement	—	—	—	—	—	—	—	—	—	—
Overall	.07 (25)	—	—	.08 (3)	—	—	—	—	—	.08 (28)
Speed of Arm Movement										
Education	—	—	—	—	—	—	—	—	—	—
Training	.10 (1)	—	—	—	—	—	—	—	—	.10 (1)
Job Proficiency	—	—	—	—	—	—	—	—	—	—
Job Involvement	—	—	—	—	—	—	—	—	—	—
Overall	.10 (1)	—	—	—	—	—	—	—	—	.10 (1)

Note. From "Literature review: Validity and potential usefulness of psychomotor ability tests for personnel selection and classification" (ARI Research Note 88-13) by J. J. McHenry and S. R. Rose, 1988, Alexandria, VA: U. S. Army Research Institute for [illegible]oral and Social Sciences. Each entry in this table consists of two numbers. The decimal number is the median validity coefficient for that particular psychomotor ability - criterion category - job type combination. The [illegible]entheses which follows is the number of validity coefficients tabulated in computing the median. The number of validity coefficients used to compute the median is provided in parentheses.

.57 for the general soldiering proficiency criterion, which subsume job knowledge and hands-on task proficiency measures. Mean validities for the remaining criterion constructs were: .26 for effort and leadership, .12 for personal discipline, and .11 for physical fitness and military bearing. The highest incremental validity (beyond that afforded by the ASVAB) was .02 for each of two criteria--general soldiering proficiency and physical fitness and military bearing.

Oppler et al. (in press) report validities of the Project A predictors for a large longitudinal sample. The profile of validities for the psychomotor/perceptual composites replicated the pattern reported above for Project A; however, the magnitude of the longitudinal sample validities was higher for three criteria: .34 for effort and leadership, .15 for maintaining personal discipline, and .17 for physical fitness and military bearing. The psychomotor/perceptual predictors added .02 over the ASVAB to the prediction of general soldiering proficiency.

Physical Abilities

Although the Services have undertaken some physical abilities research (e.g., Kroemer, 1970; Myers, Gebhardt, Crump, & Fleishman, 1984; Robertson, 1982), physical abilities have not been the focus of much research, compared to cognitive and psychomotor abilities. Also in comparison with other types of employment testing (e.g., cognitive ability appraisal), the research history regarding the assessment of physical abilities lacks breadth. Hogan (in press) points out two reasons for this deficiency. First, the implementation of physical abilities tests for the selection of individuals for physically demanding jobs has occurred fairly recently. Second, given that the assessment of physical abilities pulls on input from a variety of disciplines (e.g, biomechanics, physiology, and ergonomics), it is difficult to reach a mutual understanding as to what should be included in a taxonomy that will adequately represent the abilities in question. However, Hogan contends that the Services should attend more closely to physical abilities measurement to ensure that individuals are fully qualified for physically demanding jobs.

Physical Attribute Definitions

Fleishman (1964, 1972) conducted the first research in the physical abilities arena that resulted in a taxonomy of physical performance. Nine abilities were identified and were incorporated into the Ability Requirements Scales that are still used today for job analysis purposes (see Table 19).

Hogan (1991a) adapted and revised Fleishman's dimensions to better reflect physiological functioning and work performance. Her categories are seven-fold:
(1) Muscular Tension, (2) Muscular Power, (3) Muscular Endurance, (4) Cardiovascular Endurance, (5) Flexibility, (6) Balance, and (7) Coordination. In Hogan's model, Muscular Tension, Muscular Power, and Muscular Endurance are organized into a broader Muscular Strength construct. Similarly, Flexibility, Balance, and Coordination are included in a broader

Movement Quality construct. Cardiovascular Endurance has no higher-order counterpart (Hogan, 1984; Hogan, 1991a, 1991b; Hogan, in press). Hogan's factors are defined in Table 20.[13]

Table 19 Physical Abilities
Static Strength: Maximum muscle force that can be exerted for a brief time against external objects (lift, push, pull, or carry).
Explosive Strength: Short bursts of muscular effort to propel oneself or an object (e.g., sprints, jumps).
Dynamic Strength: Muscular endurance in exerting force continuously or repeatedly over time while resisting fatigue (hold up, support, or move body weight and/or objects over time).
Trunk Strength: Use of limited dynamic strength specific to one's stomach and lower back muscles (e.g., leg lifts, sit-ups).
Extent Flexibility: Bend, stretch, twist or reach out with arms, legs, or body (e.g., twist and touch test).
Dynamic Flexibility: Bend, stretch, twist, or reach out with arms, legs, or body quickly and repeatedly (e.g., rapid, repeated bending over and touching the floor).
Gross Body Coordination: Coordinated movement of the arms, legs, and torso for activities where the whole body is in motion (e.g., cable jump).
Gross Body Equilibrium: Ability to keep or regain body's balance or stay upright when in an unstable position, while moving or while standing motionless (e.g., rail walk test).
Stamina: Ability of lungs and circulatory systems of the body to perform efficiently over long periods without getting out of breath (e.g., 600 yard run-walk).

Source: Fleishman (1972); Fleishman & Mumford (1988); and Hogan (1992)

Muscular Strength consists of three specific constructs that account for the force generated by the body (Hogan, 1984). Muscular Tension is the broadest of these. It is defined as the capacity to exert tension against some form of resistance, using isometric and isotonic muscular contractions. Job tasks that require muscular tension involve pushing, pulling, lifting, or carrying heavy objects. Grip strength and leg dynamometer tests tap Muscular Tension. Muscular Power is the second specific construct subsumed under Muscular Strength. Muscular Power requires the quick exertion of muscular force against resistance; it adds the requirement of speed to Muscular Tension. Job tasks that require the use of hand tools (e.g., wrenches, hammers) require this construct. Power is often measured using commercial or custom-designed ergometers, and scores are reported in foot-pounds or Watt conversions (Hogan, 1991a). Muscular Endurance, the third Muscular Strength construct, represents the capacity to perform tasks that require consistent, localized muscular work while managing to postpone the onset of fatigue. Exemplary job tasks include repeated, prolonged use of hand tools or continuously loading materials onto

[13]Hogan's factors are based on physiological parameters as well as factor-analytic work.

pallets. Muscular Endurance tests involve continual applications of Muscular Tension, as in performing cranking motions with one's arm.

Table 20 Physical Ability Construct Definitions
Muscular Strength
Muscular Tension: Exertion of muscular force against resistance, using isometric and isotonic muscular contractions. It is used to either push, pull, lift, lower, or carry objects or materials.
Muscular Power: Quick exertion of muscular force against resistance, using isometric and isotonic muscular contractions.
Muscular Endurance: Exertion of localized muscular force continuously while resisting fatigue.
Cardiovascular Endurance
Cardiovascular Endurance: Sustenance of physical activity that requires increased heart rate. It is dependent on the individual's aerobic capacity and on his/her general level of fitness.
Movement Quality
Flexibility: Flexion or extension of body limbs in order to work in awkward, contorted, or extended positions.
Balance: Maintenance of body stability when the base of support is reduced or changing or both.
Neuromuscular Integration/Coordination: Sequenced movement of the arms, legs, and/or body to result in skilled physical action given the temporal and/or spatial constraints placed upon the individual.

Note. From "Theoretical and applied developments in models of individual differences: Physical abilities" by J. A. Hogan, in press, in Proceedings of the Army Research Institute Conference on Selection and Classification. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

The second major category of physical abilities is Cardiovascular Endurance (Hogan, 1984). This construct is considered to be the most important component required for successful physical performance given that it allows for sustained activity. Cardiovascular Endurance refers to the capability to maintain overall muscular activity for prolonged amounts of time. It is dependent on the individual's aerobic capacity and on his/her general level of fitness. Hogan (1984) points out that Cardiovascular Endurance is different from Muscular Endurance in that Cardiovascular Endurance involves gross body muscular activities while Muscular Endurance focuses on a small group of muscles and joints. This construct is obviously of particular importance when individuals have to sustain muscular work over long durations or for high work loads (e.g., pole climbers, warehouse personnel) (Hogan, 1984). An example of a test of Cardiovascular Endurance might assess how long an individual can continue stepping up and down on a 20-inch high bench in time to a metronome set at one beat per half second (Reilly, Zedeck, & Tenopyr, 1979).

The third major category is made up of factors affecting Movement Quality. Hogan (1984) defines three aspects of Movement Quality: Flexibility, Balance, and Neuromuscular

Integration. Flexibility involves the degree to which multiple segments of the body can be displaced or rotated easily in accomplishing work while in awkward, contorted, or extended postures. Exemplary tasks that require Flexibility include installing light fixtures or, for a professional photographer, stooping to take a picture. Flexibility tests measure range of motion of limbs. The second construct, Balance, involves maintaining one's stability when the base of support is reduced or changing or both (Hogan, 1984, 1991a). Tasks like walking on a tilted roof or on a narrow plank require balance. A test measuring Balance might require subjects to stand on one foot on a three-quarter inch wide balance beam. Balance can be required under both static and dynamic conditions as both are required for physical performance. Static conditions require that individuals maintain their bodies stability while holding a single posture, although forces are working against them. Dynamic conditions require that individuals adjust their center of gravity progressively given that the base of support is continually changing and they are actively moving. The third factor, Neuromuscular Integration or coordination, involves organizing one's movements in sequence within required temporal and spatial constraints in response to either internal or external stimuli (Hogan, 1984). Coordination is the ability to make an accurate and effective physical response given the temporal and/or spatial constraints placed upon the individual (Hogan, 1991a). Tasks such as intercepting a swinging rope require coordination.

Subgroup Differences

Sex Differences

Table 21 provides effect sizes for mean score differences between male and female performance on physical abilities tests. The largest differences arise on Muscular Strength tests, particularly those requiring upper body strength. Male means exceeded female means by about four standard deviations on some of the lifting tests. Measures of Cardiovascular Endurance that are based on the amount of body fat favor men more than those based on performance on stair climbing tests. In the former case, women are at a disadvantage given the denser fat composition in their bodies. Female means exceed male means on tests of Flexibility, but males perform better on other measures of Movement Quality (i.e., Balance and Neuromuscular Integration). It should be pointed out, however, that individuals with short arms and long legs or individuals with long arms and short legs tend to be at a disadvantage when given Flexibility measures (Hogan, 1991a).

Validity of Physical Abilities Tests

Hogan (1991a) reviewed fourteen validation studies that allowed assessment of validity coefficients using the seven physical ability constructs. Criteria were organized into three groups: objective (e.g., training completion time), subjective (e.g., supervisor's ratings of performance), and work sample (i.e., performance on a work sample task. Physical abilities measures were good predictors of all types of criteria. As expected, validities of physical abilities tests for predicting work sample criteria were often quite high.

Table 21

Sex Differences in Physical Ability Test Performance

Predictors	Males			Females			Estimated Mean Effect Size
	Mean	SD	n	Mean	SD	n	
Muscular Tension							
Lift 72 kg	56.70	10.50	969	25.60	4.70	986	4.19
Lift 60 kg	60.60	10.70	969	29.80	5.40	986	3.64
Arm Pull	147.50	26.10	350	79.40	17.60	269	2.99
Arm Lift	104.80	17.50	350	60.90	11.40	269	2.90
Upright Pull	124.80	21.20	974	77.10	13.50	1000	2.69
Handgrip	50.60	7.50	694	33.20	5.20	507.50	2.63
Cable Pull	112.50	26.20	266.50	70.60	17.30	81.50	1.72
Push (Isometric)	285.20	65.70	425.50	201.45	44.30	101.50	1.35
Muscular Power							
Ergometer	58.40	9.40	350	35	8.50	268	2.59
Medicine Ball Put	39.20	8.40	851	19.10	5	101.50	2.48
Muscular Endurance							
Push-Ups	20.40	7.90	138	4.80	4.50	51	2.18
Pull-Ups	6.50	3.80	168	0.50	0.80	81	1.90
Dynamic Arm Strength (Ergometer)	131.50	28.97	71.33	93.83	19.87	40.67	1.45
Arm Ergometer	196.40	53.20	425.50	123.50	40.40	101.50	1.43
Sit-Ups	15.40	4.20	132	10.10	3.90	78	1.30
Leg Lifts	14.70	2.70	138	11	3.30	51	1.29
Squat Thrusts	16.60	3.80	168	12.10	4.10	81	1.15
Dynamic Leg Strength (Ergometer)	82	20.70	83	61.70	14.70	54	1.09
Cardiovascular							
Lean Body Mass	60.70	6.80	980	43.70	4.20	1003	3.02
Maximum VO_2	46.80	7.30	715	36.50	6.80	659	1.46
Body Density	106.60	1.70	107.50	105	1.30	60.50	1.02
Step-Up Time	132.80	66	107.50	81.50	41.60	61.50	0.88
Step Test	40.20	22.70	168	25.50	11.70	81	0.74
Harvard Step Index	32	19.10	83	22.80	12.20	45	0.54
Skinfold	25.60	10.20	425.50	49.80	12.90	101.50	-2.25

(Continued)

Table 21

Sex Differences in Physical Ability Test Performance (Continued)

Predictors	Males			Females			Estimated Mean Effect Size
	Mean	SD	n	Mean	SD	n	
Flexibility							
Twist and Touch	15.80	7.30	127.70	16.10	6.40	68	-0.04
Sit & Reach	39.40	12.50	318.70	45.90	10.40	73.70	-0.54
Flexibility Course Time	86.50	23.20	41	105.20	22.30	22	-0.82
Flexibility Course	19.10	5.20	126	24.00	6.30	25	-0.91
Balance							
Balance Against Resistance	350.20	79.50	41	198.60	74.10	22	1.95
Static Rail Balance (3/4 inch beam)	11.80	12.10	266.50	9.30	8.10	80.50	0.22
Static Rail Balance (1 inch beam)	.05	.06	168	.04	.01	81	0.20
Neuromuscular Integration							
Minnesota Rate Manipulation	78.30	12.20	233.30	76.60	9.30	60.80	0.15

Note. From "Theoretical and applied developments in models of individual differences: Physical abilities" by J. A. Hogan, 1992, paper presented at the Army Research Institute Conference on Selection and Classification for the U. S. Army, Alexandria, Virginia.

Hogan (1991a) notes that the most successful predictors are those that are the most sensitive to individual differences. Such measures are, however, those that also have the greatest potential to result in adverse impact against females. The most successful predictors also are those that are the most highly correlated with work sample criteria. It is important to note that work sample criteria may lack external validity because they typically sample such a small portion of the total domain of work behavior.

Discussion

Cognitive Attribute Assessment

As mentioned, the ASVAB is weak in the measurement of some cognitive attributes. The ECAT battery offers measures that fill some of the gaps (Table 20), particularly for the measurement of G_f and G_v. To date, data on working memory capacity tests are insufficient for determining their usefulness as a part of the ASVAB. Future analyses should report sex and race differences in means and variances in addition to traditional validity analyses to allow estimation of levels of adverse impact.

With regard to G_v, the available data suggest that ECAT Assembling Objects is a test the Services will want to examine closely when supplementing the ASVAB. It has yielded small sex differences (relative to other spatial measures) in three large (but pre-selected) samples, appears to be less susceptible to coaching and practice than are the other tests, and has been a useful predictor in studies conducted by the Marine Corps as well as the Army.

Cumulating evidence suggests that Assembling Objects "behaves" empirically like a G_f measure. Peterson, Russell et al. (1990) showed that ECAT Assembling Objects and ECAT Figural Reasoning are good marker tests for a general spatial factor.[14] A second-order general spatial factor explained most of the variance in scores on six spatial tests; loadings on first-order factors were modest. Assembling Objects and Figural Reasoning tests loaded more highly than the other tests on the second-order factor and loaded essentially zero on the specific factor for which they were intended. The other four tests had smaller loadings on the second order factor and modest loadings on the specific factors for which they were intended.

TSR and SAR are the only two major types of constructs that are not measured by either the ASVAB or the ECAT. It is possible that TSR measures might be useful for predicting officer performance; Toquam et al. (1989) reported that fluency measures showed some predictive validity for professional, technical, and managerial jobs. There does not appear to be a lot of data regarding predictive validity of SAR tests, such as the Army's Short Term Memory. Perhaps future Army research will inform other Services about the potential usefulness of SAR.

Finally, there is good reason to continue studying a number of measures developed by the Services for possible inclusion in future ASVABs. The recent findings regarding working memory capacity may add significantly to the measurement of individual differences and, in turn, to prediction of training or job performance. Further research is needed to determine the utility of these and other Service developed measures.

Psychomotor Attribute Assessment

Addition of ECAT Tracking tests to the ASVAB would represent measurement of a new domain, and there is reason to expect these psychomotor tests would supplement the validity of the ASVAB. However, both tests are probably not necessary. ECAT Tracking 1 and 2 have virtually identical items and are highly correlated with each other (Peterson, Russell et al., 1990). Also, before implementing these psychomotor tests, the Services will need to decide how to deal with the large practice effects associated with them. Perhaps testing practice stations could be set up in the MEPS or in recruiting stations where applicants would be encouraged to try out practice items on tests. Alternatively, perhaps a number of practice items could be added to the tests.

[14] As mentioned before, the Figural Reasoning test is a series completion test that best fits the definition of a measure of G_f.

Table 22

Individual Differences Attributes and Constructs and Selected Military Measures

Broad Attributes		Related Constructs	Selected Measures Developed by the Services
Cognitive			
G_c -	Knowledge or Crystallized Intelligence	Knowledge of general information Word knowledge	ASVAB [GS, WK, AS, MC, EI] OSB, AFOQT
G_f -	Broad Reasoning or Fluid Intelligence	Inductive reasoning Conjunctive reasoning Deductive reasoning	AFOQT ECAT Mental Counters ECAT Sequential Memory ECAT Figural Reasoning
G_v -	Broad Visual Intelligence	Spatial visualization Spatial orientation	BAT, AFOQT, OSB ECAT Assembling Objects ECAT Orientation Test ECAT Integrating Details
SAR -	Short Term Acquisition and Retrieval	Recency memory Word span	BAT
TSR -	Long Term Storage and Retrieval	Associational fluency Expressional fluency Ideational fluency	
G_s -	Broad Speediness	Visual scanning Visual matching	ASVAB [CS, NO] BAT, AFOQT ECAT Target Identification
G_a -	Auditory Intelligence	Discrimination among sound patterns Auditory cognition of relations	DLAB, ARC, Superdit
G_q -	Quantitative Thinking	Computational fluency Numerical computation	ASVAB [AR, MK] OSB, AFOQT
Eng -	English Adeptness	Word parsing Phonetic decoding	
Psychomotor			
Dexterity		Finger dexterity Manual dexterity	
Basic Movement Speed and Accuracy		Reaction time Control precision Speed of arm movement	
Perceptual-Motor Movement Control		Multi-limb coordination Rate control	BAT ECAT Tracking 1 ECAT Tracking 2

(Continued)

Table 22

Individual Differences Attributes and Constructs and Selected Military Measures (Continued)

Broad Attributes	Related Constructs	Selected Measures Developed by the Services
Physical		
Muscular Strength	Muscular tension Muscular power Muscular endurance	Air Force Strength Factor
Cardiovascular Endurance	Cardiovascular endurance	
Movement Quality	Flexibility Balance Coordination	
Personality		
Extraversion	Sociable, Gregarious Ambitious, Achievement-Oriented	OSB
Emotional Stability	Emotional, Anxious, Depressed	
Agreeableness	Good-natured, Cooperative	
Conscientiousness	Dependable, Responsible	
Intellectance	Curious, Broad-minded	
Interest		
Realistic	Practical, likes hand-on work	BAT
Investigative	Curious, likes academic endeavors	
Artistic	Creative, likes self-expression	
Social	Friendly, likes people	
Enterprising	Ambitious, likes managing & directing	
Conventional	Concrete, likes exactness in work	

Source: Cognitive (Horn, 1989); Psychomotor (Fleishman, 1967; Imhoff & Levine, 1981; McHenry, 1987); Physical (Hogan, 1988); Personality (Barrick & Mount, 1991; Digman, 1990; Tett, Jackson, & Rothstein, 1991); Interests (Holland, 1983).

Sex differences on the ECAT tracking measures are large. As long as these tests are used for selection and classification into combat jobs and combat jobs remain off-limits for women, however, this is a moot point. If, however, combat exclusion policies and laws are removed in the future, a number of issues will arise. First, perhaps it will be more important to use psychomotor measures to make classification decisions because a wider range of individuals may be considered for combat jobs. Second, because the sex differences are so large, it will be necessary to show that psychomotor tests, if used, are based on real job requirements identified in job analysis. Otherwise, it could be alleged that the Services adopted such tests as a surrogate for combat exclusion policies/laws, since psychomotor measures would exclude women from jobs.

Physical Attribute Measurement

Until recently, both the Army and the Air Force administered strength tests at the MEPSs. The Army's test--the Military Entrance Physical Strength Capability Test (MEPSCAT)--measured the amount of weight that an individual can lift, and is no longer in use. The Air Force is the only Service that administers a strength test. This measure, the X factor, is comparable to the MEPSCAT.

Hogan (in press) argues persuasively for use of physical abilities measures in military settings. She contends that physical abilities are almost inherent to successful performance within many military jobs (e.g., infantry positions). It is reasonable to expect that physical abilities measures would supplement the ASVAB for the prediction of performance in physically demanding jobs. Also, taxonomies of physical abilities are now available and can facilitate generalizability of validation results from civilian jobs to the domain of military jobs, making research less costly and more efficient.

The issues involved in implementing physical abilities and psychomotor tests are similar. Specialized job analysis information would be needed to determine the physical and psychomotor requirements of the jobs. Both types of tests will yield some, if not a great deal of, adverse impact. The Services may also want to consider job redesign to reduce physical demands for some jobs. The issues of if, how, and where to appropriately set cut-off scores for the tests utilized would need to be addressed.[15] Another consideration would be the cost of acquiring special equipment to conduct physical abilities and psychomotor testing. For physical abilities testing, test administrators would also have to be hired and/or trained to validly and reliably measure individuals. In addition, there may well be a space problem to deal with should such testing be implemented at Military Entrance Processing Stations (MEPS). Rooms for testing and space for equipment storage would be needed. Despite these concerns, assessing the capacity of military applicants to handle physical tasks would appear to be fundamental to selecting individuals to perform in certain fields.

[15]When we interviewed selection and classification experts in earlier phases of this project, experts voiced some concern that cut scores on physical tests (as well as other physical restrictions on height, for example) lack job analytic support.

IV. PERSONALITY, INTEREST, AND BIOGRAPHICAL ATTRIBUTE MEASURES

Douglas H. Reynolds

In this chapter, personality, interest, and biographical constructs are examined that may hold promise for supplementing the cognitive measures that have traditionally been used by the Services to make selection and classification decisions. In each of the following sections, constructs are outlined that are typically assessed by current measures and issues are disscussed that have surfaced in the literature regarding these constructs. For each set of constructs, evidence is provided of the quality of the measurement (i.e., reliability, validity, and incremental validity) as well as information relating to moderators of validity (such as fakability and socially-desirable responding). Sub-group differences are also discussed. Because non-cognitive predictors may relate to different criteria than cognitive predictors, we have specified the types of criteria that have been shown to be best predicted by the measures discussed. Throughout the chapter, we highlight the major findings regarding instruments that have been developed by the Services, and our focus is on recently developed measures.

Personality

Some aspects of military job performance involve behavior that may be accounted for by variables other than cognitive abilities. Job performance reflects both the individual's ability and willingness to do the job over a substantial timeframe. Army enlisted personnel, for example, demonstrate the willingness to exert effort or persevere under adverse conditions, and maintain personal discipline, professional bearing, and physical fitness (Campbell, 1986). Such performance criteria may imply a greater range of job success components than can be predicted by cognitive measures alone. Thus, it is important to consider the contribution of personality variables when attempting to account for a range of job performance criteria.

Over the past decade, researchers have begun to reach agreement on the latent structure of personality (cf. Digman, 1990). Although different researchers have used different terms to describe the factors resulting from multifaceted examinations of personality, the number of higher-order factors has often centered on five (e.g., Norman, 1963; Tupes & Christal, 1961). In a recent review, Digman (1990) referred to these factors as Extraversion, Emotional Stability, Agreeableness, Conscientiousness, and Openness to Experience. Although five factors have been accepted by many, some researchers have argued for one or two more factors, often differentiating within the Openness or Extraversion factors (e.g., Hogan, 1982).

The Extraversion dimension, or "Surgency" as it is labeled by some researchers (e.g., Norman, 1963), includes traits such as sociability, gregariousness, and assertiveness (Barrick & Mount, 1991). Hogan (1982, 1986) developed a six factor personality structure by separating Extraversion into two components: Ambition, a motivational component that includes traits such as initiative and ambition, and Sociability, a social component that includes exhibitionism and expressiveness. However, most current discussions of the higher order factors have primarily concentrated on the "big five" (e.g., Barrick & Mount, 1991; Cortina, Doherty, Schmitt, Kaugman, & Smith, 1992; Tett, Jackson, & Rothstein, 1991).

The other four factors have also been given a variety of labels. The second factor, Emotional Stability, has been referred to as "Adjustment" (Kamp & Hough, 1986) or "Neuroticism" (Eysenck & Eysenck, 1969). This factor is associated with traits such as anxiety, depression, anger, and emotionality. Agreeableness, or sometimes "Likeability" (Kamp & Hough, 1986; Norman, 1963), makes up the third factor and includes traits such as kindness, trust, warmth, and sympathy. The fourth factor, Conscientiousness, has also been called "Dependability" (Kamp & Hough, 1986) and "Prudence" (Hogan, 1986) among other labels. This factor includes the traits organization, thoroughness, and reliability. The labels "Openness to Experience," (McCrae & Costa, 1985), Intellectance (Kamp & Hough, 1986), or Culture (Norman, 1963) have been used to describe the fifth and final factor. It is associated with such traits as imagination, intelligence, perceptiveness, creativity, artistic sensitivity, originality, and broad-mindedness.

The consensus concerning the conceptualization of personality as a hierarchical amalgamation of traits that results in approximately five latent components has helped researchers investigate the relationships between personality and various performance criteria. Past reviews of the personality-performance relationship generally indicated that personality measures were poor predictors of performance (e.g., Ghiselli, 1973; Guion & Gottier, 1965; Schmitt, Gooding, Noe, & Kirsch, 1984); however, none of these efforts used a common taxonomy of traits when condensing results from several studies. Unlike the cognitive domain, personality characteristics are not expected to covary greatly across the major trait dimensions. As such, grouping personality characteristics together into one domain, and investigating the predictive ability of "personality" as a construct unto itself, may have confused our understanding of the relationships between distinct aspects of personality and performance.

In recent reviews, researchers have been more specific about the nature of the relationships between the various personality constructs and performance, and they have also been more enthusiastic about the practical value of the personality domain. For example, in a meta-analysis conducted by Barrick and Mount (1991), it was concluded that each of the personality dimensions is valid (i.e., has a true-score validity coefficient greater than zero) for some criteria and some occupational groupings, and one dimension, Conscientiousness, is valid for predicting all criteria on all types of jobs. Similar results were found in Army research using a Dependability composite that includes Conscientiousness and Non-Delinquency scales (Campbell & Johnson, in press; McHenry, Hough, Toquam, Hanson, & Ashworth, 1990). Tett, Jackson, and Rothstein (1991) were equally encouraged by the validity of personality dimensions; however, these researchers found that Agreeableness is the most valid predictor of the five. Tett et al. (1991) concluded that the personality dimensions can be a powerful predictor of performance, especially when jobs were analyzed based on personality components and a confirmatory strategy was undertaken for identifying appropriate predictors.

Given the results of the meta-analyses of personality constructs, measures of these constructs could make a valuable contribution to the prediction of performance in the military. A recent study of the utility of personality measures for predicting performance in a military setting has shown that these measure can have substantial incremental validities over cognitive measures for criteria such as effort and leadership (McHenry et al., 1990). As Goldberg (in press) pointed out, personality measures are likely to be more valuable to the Services now because the downsizing of the military may lead to a more favorable selection ratio. Thus, as fewer positions are available, there may be more room for improvement when deciding who to accept into the Services.

Measures of Personality

Despite the apparent simplicity of the latent structure of personality, measures of personality traits tend to be complex and theory-bound. This is not surprising given the large number of traits that have been suggested. For example, based on the notion that most individual differences in personality have been identified and incorporated into the world's languages, Allport and Odbert (1936) cataloged over 17,000 trait descriptors.

In their review of current personality inventories, Kamp and Hough (1986) located 146 trait scales among 12 different inventories. The inventories included in the review were: the California Psychological Inventory (CPI; Gough, 1975), the Comrey Personality Scales (Comrey, 1970), the Edwards Personal Preference Schedule (Edwards, 1959), the Eysenck Personality Questionnaire (Eysenck & Eysenck, 1975), the Gordon Personal Profile-Inventory (Gordon, 1978), the Guilford-Zimmerman Temperament Survey (Guilford, Zimmerman, & Guilford, 1976), the Jackson Personality Inventory (Jackson, 1976), the Minnesota Multiphasic Personality Inventory (MMPI; Dahlstrom, Welch, & Dahlstrom, 1975), the Omnibus Personality Inventory (Heist & Yonge, 1968), the Personality Research Form (Jackson, 1967), and the Sixteen Personality Factor Questionnaire (Cattell, Eber, & Tatsuoka, 1970). These inventories were devised using a variety of different development strategies in order to serve a number of diverse purposes. Nonetheless, it is possible to map the trait scales of these inventories onto the major latent personality dimensions. Kamp and Hough (1986) rationally grouped 117 of the 146 scales into six higher-order content categories, the remainder of the scales were grouped into a "miscellaneous" category. (The researchers were using a six factor model similar to that proposed by Hogan [1982; 1986], that splits Extraversion into both a social and motivational component.) Based on prior research, the pattern of average correlations between the scales was then examined. Without exception, the scales grouped in the same higher-order category had higher average intercorrelations with each other than did scales grouped in different categories (i.e., the intercorrelation matrix had a convergent-discriminate structure). This finding provides evidence that a very diverse set of current personality assessment measures can be classified according to the personality taxonomy discussed above. It is important to note that the taxonomy does not suggest that there are only five traits or elements of personality; rather, the implication is that all traits can be organized under this conceptualization. The findings presented by Kamp and Hough (1986) tend to confirm this notion.

Recent efforts in the military personnel research community to measure personality have focused both on the development and validation of new personality measures.

A personality inventory, the "Assessment of Background and Life Experiences" (ABLE), was developed as a part of the Army's Project A to measure some of the constructs identified in the Kamp and Hough (1986) review. Developers of the ABLE created scales corresponding to constructs that had the most useful relationships with various criterion measures[1] (Hough, Eaton, Dunnette, Kamp, & McCloy, 1990). The set of scales developed for the ABLE included the "big five" constructs of Surgency (a component of Extraversion), Adjustment (Emotional Stability), Agreeableness, and Dependability (Conscientiousness), and two "miscellaneous" scales that also showed high criterion-related validities: Achievement and Locus of Control. The ABLE contains 10 content scales that relate to these constructs; an eleventh was later included to assess Physical Condition. The scales that measure each construct, the number of items on each scale, and scale reliabilities are shown in Table 23. An additional four scales were included to detect response distortions and careless responders.

The ABLE was developed to be administered in a paper-and-pencil format; however, some research has examined a computerized version of the test (Oppler et al., 1992). Comparisons between two versions indicated that the computerized version yielded higher score variances, higher internal-consistency scale reliabilities, higher scale intercorrelations, and the mean scale scores only different on one dimension (Traditional Values). The Oppler et al. (1992) findings provide some evidence that the ABLE may be computer-administered without adverse consequence.

Personnel researchers at the Air Force's Armstrong Laboratory have developed a personality measure to be used for the prediction of pilot performance (cf. Siem, 1990). The Automated Aircrew Personality Profiler (AAPP) is a computer-administered instrument that includes 94 items from the MMPI. The MMPI items included in the AAPP can be scored along five factors: Sociability, Emotional Stability, Extraversion, Competency, and Cynicism. The AAPP is administered as a component of the BAT (Carretta, 1990). The computerized administration allows for the computation of response latencies for each of the MMPI items. Response latency, or the time it takes an individual to respond to an item, has been hypothesized to be related to that individual's standing on the trait being measured by the item (e.g., Popham & Holden, 1990). Specifically, individuals who are high on a trait are expected to endorse items that are representative of that trait quickly, but expected to reject items that are reverse scored more slowly. The opposite pattern is shown for people who are low on a trait.

Siem (1991) investigated the relationship between scale scores and response latencies for the five MMPI dimensions assessed by the AAPP. The study found the predicted pattern of correlations between four of the five scales and their respective latencies. That is, the scale scores related negatively with the response latency for endorsed items and related positively for the rejected items. A later factor-analytic study (Siem, 1992) showed that scale scores and response latencies for the same scale often load on different factors. Only two factors, Extraversion and Competency, were defined by the scale scores and their corresponding latencies. This finding indicates that further work is necessary before response latencies can be assumed to be measuring traits and before the construct validity of latencies can be fully explicated. One potential problem with the use of latencies is the presence of aberrant responders in the data.

[1]The most useful relationships are not always the highest relationships, however. For example, in the Kamp and Hough review (1986), scales in the Openness to Experience (Intellectance) category were related positively to education and training criteria as well as to substance abuse criteria. This pattern of relationships makes the implementation of such scales difficult.

Random or careless responders may distort the relationships between latency measures and other characteristics because although they tend to respond quickly, their response is unlikely to be indicative of an underlying trait.

Table 23

ABLE Constructs, Scales, and Scale Characteristics

Construct	Scale	Number of Items	Reliability: Alpha	Reliability: Test-Retest
Surgency	Dominance	12	.80	.79
	Energy Level	21	.82	.78
Adjustment	Emotional Stability	17	.81	.74
Agreeableness	Cooperativeness	18	.81	.76
Dependability	Traditional Values	11	.69	.74
	Nondelinquency	20	.81	.80
	Conscientiousness	15	.72	.74
Achievement	Self-Esteem	12	.74	.78
	Work Orientation	19	.84	.78
Locus of Control	Internal Control	16	.78	.69
Physical Condition	Physical Condition	6	.84	.85
Response Validity	Nonrandom Response	8	---	.30
	Social Desirability	11	.63	.63
	Poor Impression	23	.63	.61
	Self-Knowledge	11	.65	.64

Note: From "Criterion-related validities of personality constructs and the effect of response distortion on those validities" [monograph] by L. M. Hough, N. K. Eaton, M. D. Dunnette, J. D. Kamp, and R. A. McCloy, 1990, Journal of Applied Psychology, 75.

Validity Evidence

Predicting Performance

Criterion-related validity studies with the ABLE indicate that the sub-scales predict a number of performance criteria. In Project A/Career Force research, personality constructs tended to correlate with the criterion measures of Effort and Leadership, Personal Discipline, and

Physical Fitness and Military Bearing, but not with technical task performance criteria (Core Technical Proficiency and General Soldiering). Table 24 lists the criterion-related validities (uncorrected for range restriction and criterion unreliability) for each of the ABLE sub-scales with three categories of performance measures, as they were reported by Hough et al. (1990).

Table 24

ABLE Scales: Criterion-Related Validities

Construct	Scale	Effort & Leadership	Personal Discipline	Physical Fitness & Military Bearing
Surgency	Dominance	.15*	.02	.18*
	Energy Level	.22*	.14*	.25*
Adjustment	Emotional Stability	.17*	.12*	.16*
Agreeableness	Cooperativeness	.15*	.21*	.14*
Dependability	Traditional Values	.13*	.25*	.16*
	Nondelinquency	.12*	.29*	.14*
	Conscientiousness	.18*	.23*	.22*
Achievement	Self-Esteem	.20*	.13*	.20*
	Work Orientation	.23*	.18*	.21*
Locus of Control	Internal Control	.13*	.13*	.13*
Physical Condition	Physical Condition	.09*	-.03*	.29*

Note: From "Criterion-related validities of personality constructs and the effect of response distortion on those validities" [monograph] by L. M. Hough, N. K. Eaton, M. D. Dunnette, J. D. Kamp, and R. A. McCloy, 1990, Journal of Applied Psychology, 75.
* $p < .01$.

Recent meta-analyses have found that the "big five" personality constructs show significant relationships with performance criteria (Barrick & Mount, 1991; Tett et al., 1991). Specifically, Barrick and Mount (1991) found that Conscientiousness predicted performance across occupational groups (population parameter = .20 - .23) and criteria types (population parameter = .20 - .23); other factors proved to be significant predictors for some occupations and some criteria. Tett et al. (1991) reported mean validities (corrected for predictor and criterion unreliability) of .16 for Extraversion, .18 for Conscientiousness, -.22 for Neuroticism, .27 for Openness to Experience, and .33 for Agreeableness.

Personality measures are rarely used by themselves for making personnel decisions; rather, they are typically combined with measures of other human characteristics when predicting future job performance. Thus, the incremental validity of a measure of personality is as important as a demonstration of a significant zero-order correlation with performance indicators. McHenry et al. (1990) reported incremental validities for the ABLE above and beyond the performance variability accounted for by the ASVAB. It was found that "the ABLE added .11 to the validity [of the ASVAB] 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" (p. 347). As such, the ABLE proved to have the greatest incremental validity of any of the supplemental measures that were studied in Project A/Career Force when volitional criteria were considered.

Predicting Attrition

The validity of the Air Force's AAPP has been investigated for predicting training attrition (Siem, 1991, 1992). Two scale scores, Sociability and Cynicism, showed modest relationships (.13 - .14, uncorrected for range restriction) with the dichotomous attrition variable. One response latency, for items endorsed for the Extraversion dimension, evidenced incremental validity (.10) over the scale score validity (Siem, 1991). The Extraversion factor was also found to contribute significant incremental validity after other BAT sub-tests (assessing information processing speed, psychomotor skills, and attitudes) were added to the model (Siem, 1992). Unfortunately, the incremental validity of the latency scores alone, above the BAT sub-tests, was not reported in the study. These findings whet the appetite for more information about the predictive validity of response latencies; one is led to wonder about the relationships between latencies and more sophisticated criteria, such as job effort measures.

Moderators of Personality Test Validities

A number of studies have examined factors that have been hypothesized as moderators of personality test validity. Kamp and Hough (1986) reviewed five types of moderators (nonpurposeful responding, response sets, faking, personality characteristics, and group membership); an updated summary of their findings is presented below.

Nonpurposeful Responding

Nonpurposeful, random, or careless responding, while almost certain to affect test validity, can be readily detected (e.g., Gough, 1975; O'Dell, 1971). A typical method for detecting nonpurposeful responding involves the inclusion of item scales that have extreme endorsement frequencies. For example, the California Psychological Inventory (CPI) includes items such as "I must admit that people sometimes disappoint me" (Gough, 1975) which have predictably high endorsement rates. In studies that have used randomly generated response profiles, the use of these scales allowed for the identification of a high proportion of the bogus protocols (O'Dell, 1971). A nonrandom response scale was used on the ABLE to identify careless responders, and a comparison of the criterion-related validities computed for careful and careless responders indicated that the majority of the validities compared did vary significantly between these groups

(Hough et al., 1990). Although the validities for the careless responders were typically lower in that study, they were still above zero in many cases.

Social Desirability and Faking

Response sets (e.g., the tendency to respond to items in a socially desirable manner) can affect scale validities if respondents' profiles differ due to a particular response tendency rather than by differences in the trait of interest. There is some controversy about whether variance on items that measure constructs such as social desirability represent a response set or more meaningful variance that can be attributed to true differences on a trait that happens to be "socially desirable." After reviewing research on this issue, Kamp and Hough (1986) concluded that response set variability may not be a practical concern because "there is no strong evidence that the criterion-related validity of temperament scales is moderated by response sets" (p. 56). This conclusion was supported by a concurrent validation study on the ABLE that indicated that validities based on subjects who responded in a socially desirable manner differed to a small degree (effect size = .03) from those based on more "accurate" subjects (Hough et al., 1990). This finding was especially true for the scale-criterion correlations that were expected to be strong. Unfortunately, Hough et al. (1990) aggregated data across jobs without regard to differences in sample sizes from different jobs and differences in criterion measures. Further analysis is needed to buttress Hough et al.'s conclusions.

More recent analyses of Project A/Career Force data have yielded less favorable results (Oppler et al., in press). When the validity of the ABLE was examined with a longitudinal research design, it was found that the prediction of the Effort and Leadership factor was lower for this design (.20) compared to the concurrent design (.33). This difference may be in part due to higher levels of social desirability (by about one-half of a standard deviation) found in the longitudinal design sample. In the longitudinal sample, social desirability also showed higher relationships with other ABLE scales (average $r = .29$) than was the case in the concurrent sample (average $r = .20$). That the longitudinal sample showed higher levels of desirable responding may not be surprising, considering these subjects completed the ABLE shortly after enlistment, and may have imbued the test with more administrative significance than those in the concurrent sample, who completed the ABLE during their first tour. Another interesting finding from the study was that social desirability was found to correlate negatively with the Armed Forces Qualification Test (a test of general cognitive ability) in both the longitudinal and concurrent samples. These findings may indicate that response sets, and social desirability in particular, may lower criterion-related validity. The good news is that even in a situation where social desirability appeared to be operating, the validity coefficient was still significant (e.g., .20 in the Project A/Career Force longitudinal validation sample).

Purposeful faking has been examined in a number of studies, however the effects of faking on validity are still unclear. Studies comparing subjects who have been instructed to fake their responses have shown that it is possible to purposefully distort responses in a desired direction (Hough et al., 1990; Schwab, 1971). In a selection situation, it has been shown that applicants' scores on a personality instrument are repeatedly higher than those provided by incumbents (Kleinke, 1992), indicating purposeful distortion on the part of the applicants. However, other research has suggested that the prevalence of faking in actual selection contexts may be low (Dunnette, McCartney, Carlson, & Kirchner, 1962) and that it is possible to detect faked profiles (Tellegen, 1982).

It is important to note that purposeful faking and socially desirable responding may differ in important ways. For example, when subjects have been asked to purposely distort their responses to the ABLE, they raised their scores on each scale about a half a standard deviation; however, subjects responding in a socially desirable manner (i.e., those in the Project A/Career Force longitudinal validation sample) show a much more varied pattern of distortion across the scales (Oppler et al., in press). Future research should be conducted to examine the differences between faking and socially desirable responding in more detail.

Given the fact that the effect of purposeful faking on validity has not been well researched, Kamp and Hough (1986) suggest not eliminating possible faked responses until the effects of those responses on validity have been investigated. This may be especially important given that some researchers have suggested that faking may actually increase validity (Dunnette et al., 1962; Ruch & Ruch, 1967), while others have some evidence that validity may drop (Oppler et al., in press). The issues of faking and socially desirable responding are of special importance in the military, where applicants may be motivated to fake good <u>and</u> bad, and where it is possible that recruiters may assist applicants in order to increase their probability of succeeding in the selection process. More research in this context is needed.

<u>Personality Characteristics</u>

The characteristics of various personality traits themselves have also been hypothesized to affect validity coefficients (e.g., Bem & Allen, 1974). These characteristics include the consistency and observability of individual traits and the general characteristics of social communication skill and introspectiveness. Early findings concerning these effects have been questioned on methodological grounds (Tellegen, Kamp, & Watson, 1982), leading Kamp and Hough (1986) to conclude that practically significant moderation by these variables has yet to be "unequivocally demonstrated." However, in that review, the authors also suggest that introspection or self-knowledge may show the most potential as a moderator. Later research on the ABLE (Hough et al., 1990) included a self-knowledge scale, but no significant moderating effect on the criterion-related scale validities was identified.

<u>Subgroup Differences</u>

The research that has been conducted examining group differences (i.e., sex and race) on personality traits indicates that males and females differ on many traits and are thus normed separately on many tests (cf. Maccoby & Jacklin, 1974). Differences between races, however, are inconsistent and probably negligible in most situations (Kamp & Hough, 1986). Research in this area is not extensive, however the available research indicates that personality measures are likely to have little if any differential impact on protected groups (Goldberg, in press).

Current research with the ABLE tends to confirm these conclusions (Peterson, Russell et al., 1990). Table 25 presents the mean differences (on a standard deviation scale) between males and females and between various minority groups and Whites. These results show females to have meaningfully higher scores on the Nondelinquency, Internal Control, and Self-Knowledge scales, while males showed higher scores on the Physical Condition scale. The effect size differences between races reveal that Blacks tend to score higher than Whites on most scales and that Hispanics tend to be more conscientious, less delinquent, and respond in a less socially desirable manner than Whites. Other minorities show few differences when compared with Whites. Future work on personality measurement needs to examine methods of handling differences between subgroups. Personality tests are often used for diagnostic purposes and separate norms are appropriate under those conditions. The use of separate norms in a selection context will be likely to meet some opposition and thus needs further consideration.

General Conclusions Regarding Personality

This brief review of recent research on personality test yields several conclusions:

- Recent consensus in the area of personality structure have led to agreement on basic factors around which traits may be organized. These factors have helped researchers to be specific about the nature of the criterion relationships that may be expected for personality variables.

- Meta-analyses have shown personality variables to be consistent predictors of a variety of criteria.

- Current research indicates that personality measures are good candidates as supplemental measures to existing and experimental cognitive tests, especially for the prediction of "will-do" criteria such as Effort and Leadership, Personal Discipline, and Physical Fitness and Military Bearing, as well as training attrition.

- Socially desirable responding may moderate personality test validity, however contradictory evidence indicates that the extent to which this is a problem has yet to be fully determined. Purposeful faking also requires further research to determine its prevalence. Also, faking may be detectable, but it is not yet clear how to best deal with faked responses.

- Personality measures appear to show smaller differences between races than do cognitive measures, and the differences that have been shown tend to favor minority respondents. However, sex differences have often required separate norms.

Personality tests are promising candidates as supplements to the cognitive measures traditionally used by the Services. As with any measure, however, some important issues (e.g., the effect of socially desirable responding) require continued research before their effects are fully understood.

Table 25				
ABLE Effect Size Differences by Gender and Race				
ABLE Scale	Male-Female Effect Size[a] (d)	White-Black Effect Size (d)	White-Hispanic Effect Size (d)	White-Other Effect Size (d)
Dominance	.17	-.26	.00	-.02
Energy Level	-.04	-.15	-.07	-.02
Emotional Stability	.18	-.21	-.07	-.03
Cooperativeness	-.11	-.27	-.10	.02
Traditional Values	-.11	-.10	-.10	.07
Nondelinquency	-.33	-.22	-.32	-.07
Conscientiousness	-.18	-.25	-.24	-.14
Self-Esteem	.09	-.23	.02	-.12
Work Orientation	-.14	-.27	-.08	-.05
Internal Control	-.25	.07	.09	.18
Physical Condition	.54	-.23	.00	-.03
Non-Random Response	-.05	-.21	-.79	-.57
Social Desirability	.00	-.21	-.79	-.57
Poor Impression	-.07	.06	.06	.00
Self-Knowledge	-.21	-.29	.03	-.10

Note: From "Analysis of the experimental predictor battery: LV Sample" by N. G. Peterson, T. L. Russell, G. Hallam, L. M. Hough, C. Owens-Kurtz, K. Gialluca, and K. Kerwin, 1990, in J. P. Campbell and L. M. Zook (Eds.), Building and retaining the career force: New procedures for accessing and assigning Army enlisted personnel, Annual Report, 1990 Fiscal Year (ARI FR-PRD-90-6), Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

[a]d is the standardized mean difference between two subgroup scores. All effect sizes in this table are relative to the majority (male, White) subgroup. A positive effect size indicates that Whites (or males) score higher than the minority (or females), and a negative value indicates that Whites (or males) score lower.

Interests

Several comprehensive reviews of research on vocational interests and preferences have been conducted in recent years (Barge & Hough, 1986; Dawis, 1991). This section will summarize the major findings from these reviews and examine current measures used by the Services to examine interests and preferences.

One of the most widely cited conceptualizations of the structure of the interest domain is Holland's categories of occupations and vocational interests (cf. Holland, 1973; 1976). These categories are based on four assumptions: (1) People can be categorized according to a combination of six interest types (realistic, investigative, artistic, social, enterprising, or conventional); (2) Work environments can be grouped into the same six types, and these environments will be dominated by a people of a specific interest type; (3) People will seek out work environments that allow them to use the skills they possess, and to express attitudes and values they hold; (4) Work behavior is affected by the interaction between a person's personality and the characteristics of the environment in which he or she works.

Other factor analytic studies of the interest domain have produced similar factors. For example, Roe (1956) compared the factors that resulted from a number of early studies and found that the factors from each of the studies fit well within an eight category classification. The categories were: service, business contact, organization, technology, outdoor, science, general cultural, and arts and entertainment (as cited in Holland, 1976). A more recent review comparing interests found similar factors (Dawis, 1991). Another approach for identifying basic interests was used with the Strong-Campbell Interest Inventory (Campbell & Hansen, 1981); clusters of intercorrelated items were distinguished and labeled according to their content. Twenty-three basic interest scales were created that represent homogeneous categories (e.g., Agriculture, Science, Art, Sales). These basic interests are at a conceptually lower level than the Holland Themes and have been shown to fit within the six factors (Campbell & Hansen, 1981). The advantage of the basic interest scales is that their high degree of internal consistency lends understanding and face validity to the scales.

Information on vocational preference and work characteristics may be of particular interest to the Services because recruits tend to be career naive and because they are unlikely to make use of professional guidance when making career decisions (Baker, 1985). Thus, several efforts have been undertaken to link recruit preferences to work characteristics in the military. This section highlights instruments that have already been developed (e.g., the VOICE and the AVOICE), although there is also an instrument being developed at the Defence Manpower Data Center that may be available in the future (McBride, J. R., Personnel Communication, August 30, 1992).

The Air Force's Vocational Interest Career Examination (VOICE) was constructed in the late 1970's to provide vocational interest data for making initial recruit classification decisions (cf. Alley & Matthews, 1982). The VOICE includes items that require examinees to indicate whether they like, dislike, or are indifferent about various jobs, work tasks, spare time activities, and desired learning activities. The VOICE can be used with two different types of scoring procedures. First, homogeneous basic interest scales were developed based on a factor analysis of the individual items. The basic interest scales have shown internal consistency reliabilities from the high .80s to mid-.90s (Alley & Matthews, 1982). Second, occupational scales were created as a function of the basic interest scales and regression weights representing the relationship between the basic interest scales and later job satisfaction in each of 20 occupational groups (Alley, Wilbourn, & Berberich, 1976). Thus, the occupational scale scores are actually predictions of how satisfied an individual is likely to be in a given occupational grouping. Subsequent research reduced the number of occupational scales to eight (Watson, Alley, & Southern, 1979). These eight occupational areas are: mechanical, administrative, electronics, security and support services, medical care, medical and dental technician, utilities maintenance, and technical and allied specialties.

A modification of the VOICE, the Army Vocational Interest Career Examination (AVOICE), was used in Project A/Career Force to examine the predictive value of interest composites (Peterson, Hough et al., 1990). The AVOICE uses 22 scales that were based on VOICE scales and were intended to measure all six of the interest constructs in Holland's (1966) model. A principal components factor analysis of the AVOICE scales revealed a six-factor structure; these interest factors were labeled Skilled Technical, Structural/Machines, Combat-related, Audiovisual Arts, Food Service, and Protective Service. A later analysis of the AVOICE split the Skilled Technical factor into Administrative, Interpersonal, and Skilled Technical, and renamed Combat-related as Rugged/Outdoors, for a total of eight factors (Peterson, Hough et al., 1990). The AVOICE scales, the number of items in each, and their internal consistency and test-retest reliabilities are shown in Table 26. The AVOICE requires twenty minutes to administer.

Validity of Interest Measures

In their review of the interest measurement area, Barge and Hough (1986) conclude that the convergence of findings in the area testifies to the robust validity of interest measures. The predictive utility of interest measures has been examined for several types of criteria. Generally, it has been shown that interests can predict occupational membership with reasonable accuracy. For example, Lau and Abrahams (1971) used a version of the Minnesota Vocational Interest Inventory that had been modified for the Navy to examine the relationship between interests and occupational membership over time. It was found that 60 percent of the people in the sample were in occupations that corresponded with one of their two highest scale scores six years after taking the inventory. Similar findings have been shown with other interest measures (cf. Dawis, 1991).

Job satisfaction has also been used as a criterion in studies of the predictive ability of interest measures. In their review, Barge and Hough (1986) found that over 18 validation studies, the median correlation between interest in an occupational field and job satisfaction was .31. However, both Barge and Hough (1986) and Dawis (1991) note that many studies have not found significant validities when job satisfaction is used as the criterion. Some recent research with the AVOICE found only low correlations (i.e., less than .20) with job satisfaction criteria (Carter, 1991). It has been suggested that low validities may be due to restriction of range in satisfaction criteria (Campbell, 1971) or to moderator variables such as self-esteem (Korman, 1967).

Research with the Air Force's VOICE did find significant relationships between interests and job satisfaction for several job groups using a concurrent validation design (cf. Alley & Matthews, 1982). The study found multiple correlations ranging from .25 to .46 between the basic interest scales and job satisfaction. A predictive design used in a later study (also reported in Alley & Matthews, 1982) demonstrated similar relationships (multiple correlations ranged from .22 to .42 using common gender equations). Alley and Matthews (1982) indicate that nearly 100 percent of the recruits with high predicted satisfaction for their assigned career areas actually reported high satisfaction levels; conversely, only about 30 percent of those who were assigned to areas where their predicted level of satisfaction was low reported being satisfied with their jobs.

Table 26 Number of Items, Means, Standard Deviations, and Reliability Estimates for AVOICE Scale Scores			
AVOICE Scale	No. Items	Coefficient Alpha[a]	Test-Retest Reliability[b]
Clerical/Administrative	14	.92	.78
Mechanics	10	.94	.82
Heavy Construction	13	.92	.84
Electronics	12	.94	.81
Combat	10	.90	.73
Medical Services	12	.92	.78
Rugged Individualism	15	.90	.81
Leadership/Guidance	12	.89	.72
Law Enforcement	8	.89	.84
Food Service - Professional	8	.89	.75
Firearms Enthusiast	7	.89	.80
Science/Chemical	6	.85	.74
Drafting	6	.84	.74
Audiographics	5	.83	.75
Aesthetic	5	.79	.73
Computers	4	.90	.77
Food Service-Employee	3	.73	.56
Mathematics	3	.88	.75
Electronic Communication	6	.83	.68
Warehousing/Shipping	2	.61	.54
Fire Protection	2	.76	.67
Vehicle/Equipment Operator	3	.70	.68

Note: From "Analysis of the experimental predictor battery: LV Sample" by N. G. Peterson, T. L. Russell, G. Hallam, L. M. Hough, C. Owens-Kurtz, K. Gialluca, and K. Kerwin, 1990, in J. P. Campbell and L. M. Zook (Eds.), Building and retaining the career force: New procedures for accessing and assigning Army enlisted personnel, Annual Report, 1990 Fiscal Year (ARI FR-PRD-90-6), Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

[a]N = 8,224 to 8,493.
[b]N = 389 to 409 for test-retest correlation.

The relationship between interests and job performance also has been investigated, and interests have been found to have significant, but low correlations with performance ratings (Dawis, 1991). In the Barge and Hough (1986) review, interests showed a median correlation of .20 with performance ratings over 11 studies. More recently, the six interest scales of the AVOICE were related to five different components of job performance in an Army sample (McHenry et al., 1990). This study showed that vocational interests related most highly with "can do" aspects of performance such as technical proficiency (.44) and general soldiering proficiency (.44), and slightly less well with "will do" factors such as Effort and Leadership (.38), Personal Discipline (.35), and Physical Fitness and Military Bearing (.38). Although these relationships are substantial, when interests were included in a predictor composite that included cognitive and personality measures, the validity of the composite did not improve. This finding suggests that vocational interests may hold more promise for predicting job satisfaction and tenure than for performance.

Subgroup Differences

Race and Ethnic Differences

Barge and Hough (1986) found few studies that have shown differences in vocational interests between different races and ethnic groups. An exception was found by Berger and Berger (1977) who reported differences between Blacks and Whites on the VOICE; however, the differences were small. Another study (Whetstone & Hayles, 1975) found that Blacks score higher than Whites on some scales of the Strong Vocational Interest Blank (SVIB), but the differences were not statistically significant. More recently, Project A/Career Force researchers (Peterson, Russell et al., 1990) found larger differences between races, and most of these differences showed minorities scored higher than Whites on most scales (see Table 27). These findings suggest that race differences, at least in military samples, may be more of an issue than has been concluded in the past. In the future, research will need to consider what these differences mean in relation to various criteria.

Sex Differences

As is the case with many personality variables, men and women differ in their vocational interests. Campbell and Hansen (1981) indicated that the sexes differ by at least 16 percentage points on almost half the items on the SVIB. Furthermore, these differences have not abated since they were first shown in the 1930's (Campbell & Hansen, 1981). Sex differences are a problem for interest measurement, some have argued, because the use of biased inventories perpetuates the status quo in occupations that are dominated by one sex (e.g., Tittle & Zytowski, 1978).

Research using the AVOICE during Project A/Career Force (e.g., Peterson, Russell et al., 1990) indicated that males outscored females on 13 of the 22 scales, and women scored higher than men on the Clerical/Administrative, Medical Services, and Aesthetics scales. Both gender and race differences are summarized by the eight AVOICE factors in Table 27.

Several things have been done to address the problem associated with sex differences. Inventories have been modified to use gender-neutral language (e.g., Boyd, 1978), include separate scales and norms for each sex (Campbell & Hansen, 1981), and use sex-balanced items (Rayman, 1976). The appropriateness of these approaches continues to be a matter of controversy (Dawis, 1991). The National Institute of Education has issued guidelines concerning sex fairness in vocational interest inventories that provide direction for addressing the issues raised by sex differences (Diamond, 1975; reprinted in Campbell & Hansen, 1981).

Table 27

AVOICE Composite Score Effect Size Differences by Gender and Race

Composite	Male-Female Effect Size[a] (d)	White-Black Effect Size (d)	White-Hispanic Effect Size (d)	White-Other Effect Size (d)
Rugged/Outdoors	1.13	.67	.31	.27
Audiovisual Arts	-.26	-.35	-.35	-.35
Interpersonal	-.39	-.45	-.25	-.19
Skilled/Technical	.00	-.55	-.56	-.43
Administrative	-.40	-.82	-.36	-.20
Food Service	-.16	-.52	-.11	-.01
Protective Services	.36	.23	.10	.14
Structural/ Machines	.85	-.11	.00	-.01

Note: From "Analysis of the experimental predictor battery: LV Sample" by N. G. Peterson, T. L. Russell, G. Hallam, L. M. Hough, C. Owens-Kurtz, K. Gialluca, and K. Kerwin, 1990, in J. P. Campbell and L. M. Zook (Eds.), Building and retaining the career force: New procedures for accessing and assigning Army enlisted personnel, Annual Report, 1990 Fiscal Year (ARI FR-PRD-90-6), Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

[a]d is the standardized mean difference between two subgroups' scores. All effect sizes in this table are relative to the majority (White, male) subgroup. A positive effect size indicates that Whites (males) score higher than the minority (females), and a negative value indicates that Whites score lower.

Fakability

In their review of the literature, Barge and Hough (1986) concluded that, as with other self-report instruments, it is possible to distort one's responses in a desired direction. The extent to which this is a problem in applied contexts has been questioned, however (Campbell & Hansen, 1981). For example, one study found that students taking the SVIB while applying for a Navy scholarship provided responses that did not differ from those provided when they completed the instrument under more routine conditions (Abrahams, Neumann, & Githens, 1971). In a military context, the evaluation of faking during Project A/Career Force echoed the findings from other contexts (Campbell, 1987). Specifically, it was shown that, when instructed to do so, soldiers could fake their responses in a way that might ensure that they would be placed in a combat situation, as well as in the opposite direction. Project A/Career Force findings also indicated that the incidence of response distortion is not high in a research context. More research is needed comparing levels of distortion in research and operational contexts.

General Conclusions Regarding Interests

Research on vocational interests indicates the value of continued research in the area. A summary of the major findings reviewed here is provided below.

- Similar to the personality domain, interest researchers have defined a small set of factors that define the structure of the interest domain.

- Validation findings indicate that interest measures predict later occupational membership and job satisfaction, however interests do not appear to add much in the prediction of job performance over that accounted for by cognitive and personality predictors. These findings suggest that interest measures may by more useful for classification or vocational guidance counseling purposes, rather than as selection measures.

- Interest measurement must recognize and deal with sex differences in some manner (e.g., use separate norms).

- Recent research with the AVOICE suggests that race differences may be larger than earlier research has suggested, however the pattern of difference does not appear likely to lead to differential impact on protected groups.

The second conclusion is critical and warrants expansion. Several branches of the Services currently use individual job preferences in their classification process (cf. Russell et al., 1992). Because new recruits are likely to have little job experience and know almost nothing about the types of jobs that are available to them, interest inventories are likely to provide valuable information to inform the classification process. It is possible that classification procedures that match recruit and job interest profiles will produce better person-job matches than other methods that rely only on the stated job preferences of the recruit. However, the validity of this hypothesis depends on our ability to identify and measure stable interest patterns among young adults.

Another possibility that deserves consideration is that interest measures could be used to guide job counseling with prospective recruits. For example, recruiters or other job counselors could use interest profiles to help prospective recruits understand job options in light of their own interests. Such a procedure may increase the quality of the preference judgments made by recruits, and those judgments could then be used in the classification process. Additionally, using interest information in job counseling would eliminate some of the concern that may arise if separate sex or race norms were used.

Biographical Information

Biographical information or "biodata" represents another realm of predictor measurement, however, exactly what constructs are measured by this information is still an unanswered question. Although biodata are generally regarded as "non-cognitive" measures, some research has demonstrated relationships between cognitive measures and biodata (e.g., Trent, in press). Based on the adage that "past behavior is the best predictor of future behavior," biodata researchers have typically sought life events that distinguish between groups that differ on some criterion (e.g., sales performance). Biodata has a reputation of being atheoretical due to the fact that inventories were traditionally based on items that were empirically shown to discriminate between criterion groups. It has been only recently in the long history of biodata use that researchers have begun to define the factors that are being assessed.

Biodata instruments have been developed in a variety of forms, however their development often proceeds along a common logic. The prototypical procedure for the development of a weighted information blank was documented in seven steps by England (1971, cited in Barge & Hough, 1986). (1) A criterion of interest is chosen; (2) groups that are high and low on the criterion are identified and subdivided into validation and cross-validation groups; (3) a large pool of potential items is developed; (4) response categories are developed for each item; (5) percentages of people in each validation criterion group falling in each response category for an item are compared, and items are weighted to reflect the degree to which they discriminate between the two groups; (6) a total score on the application blank is calculated for each individual, the correlation between the scores and group membership is computed, and the correlation is cross-validated; (7) a cut-off score is devised such that the number of people correctly predicted on the criterion is optimized.

Although the approach described above is a common method of biodata development, several other procedures have been used. One central difference between procedures is that some use a rational approach to scale development, while others use an empirical approach akin to that described by England (1971). Rational approaches typically begin with the identification of biographical constructs that are hypothesized to relate to the criterion(ia) of interest. Next, items are developed and included in an instrument based on rational judgments about the degree to which the items tap the identified constructs. When rationally developed scales are subjected to empirical keying, it is often found that the rational structure is difficult to maintain. However, some empirical keying approaches may allow for more of a rational structure than others (e.g., the alternating least squares optimal scaling procedure).

It has often been noted that the personality and biodata domains overlap substantially (e.g., Mael, 1991). In fact, in this review the Army's Assessment of Background and Life Experiences (ABLE) was categorized as a personality measure (because it was designed to tap personality constructs), while some others have categorized it as a biodata instrument (e.g., Steinhaus & Waters, 1991). Mael (1991), in discussing the underlying rationale for biodata measurement, indicated that "the only necessary attribute of biodata items is that they be historical" (p. 783). That is, they should pertain to events that have taken place in the past or that continue to take place. According to Mael, biodata may capture two different aspects of individuals: biodata measure underlying dispositions and directly assess the events that shape past and future behavior. Personality measures, in contrast, are oriented toward assessing unitary dispositional constructs. Thus, biodata provide a somewhat broader range of measurement. These distinctions were applied in the categorization of measures for this review.

Past criticism regarding the atheoretical nature of biodata led some researchers to try to identify the underlying dispositional characteristics that are assessed. For example, Owens and his colleagues (e.g., Owens, 1976; Owens & Schoenfeldt, 1979) prepared a biographical information blank that was used over several years to assess University of Georgia undergraduates. The instrument has been factor analyzed and a set of 13 male and 15 female factors have been derived. Research in other contexts has shown the factors to be fairly stable across organizations and over time (Eberhardt & Muchinsky, 1982; Neiner & Owens, 1982). These factors only account for about 40 percent of the variability in the responses, however, suggesting that the instrument may be too heterogeneous to be more adequately accounted for by a clear set of factors.

Biodata have also been used to create subgroups of individuals who are similar on clusters of biographical items (Owens, 1976). Subgrouping recognizes the inherent heterogeneity of biodata by grouping people by the full range of variables that associate them with others with similar characteristics. When examined in this manner, the notion that biodata may tap underlying dispositional constructs becomes irrelevant and the focus is placed on the characteristics that discriminate between groups, regardless of the homogeneity of those characteristics.

The next section describes biodata measures that have been developed or used in a military context.

Military Biodata Instruments

A number of biodata instruments are available or are under development in the Military Services. Several instruments have been recently developed for the Navy. These include the Profile of Experiences and Characteristics (Hanson, Paullin, & Borman, 1990), a rationally developed measure for predicting attrition from the Naval Reserve Officer Training Corps, and the Personal History Questionnaire (Mattson, Abrahams, & Hetter, 1985), an experimental instrument used for predicting success at the U.S. Naval Academy. Currently, these instruments are in an early stage of development. Other instruments that have been researched more extensively (the ASAP, EBIS, and the LEAP) will be discussed below.

Trent (in press) identified several biodata inventories that were used in the development of a Joint-Service biodata instrument, the Armed Services Applicant Profile (ASAP). These inventories included the History Opinion Inventory (HOI), a component of the Air Force Medical Evaluation Test, which has been used by the Air Force for referring recruits for additional psychological assessment; the Recruit Background Questionnaire (RBQ), which has been used primarily by the Navy for predicting recruit attrition; and the Military Applicant Profile (MAP), an Army instrument used to screen non-high school diploma graduates. The ASAP is based in part on items taken from the RBQ and the MAP and has been shown to be related to first-term military attrition. The ASAP, although it has never been used operationally, was combined with the ABLE and some experimental items (Laurence & Waters, in press) to form the Adaptability

Screening Profile (ASP), a background and personality measure that was to be used to predict adjustment to military life.

The ASAP has two forms that contain 50 items apiece. The items have been rationally categorized into the following topic areas (Trent, in press):

- Academic Involvement
- Nondelinquency
- Work Orientation
- Physical Condition
- Interests
- Conscientiousness
- Energy Level
- Independence
- Self-Esteem
- Traditional Values
- Sociability
- Demographics
- Military Career Intentions
- Dominance
- Cooperativeness
- Emotional Stability
- Miscellaneous

The average item-total score correlation for each form was found to be .21, indicating the heterogeneity of the items. ASAP items were empirically keyed using a Service completion criterion.

Although the ASAP was developed in an atheoretical manner, items from the two forms were factor analyzed to help understand the content of the instrument. A six-factor solution explained 27 percent of the total variability in each of the forms. These factors included School Achievement, Delinquency, Work Ethic, Independence, Social Adaptation, and Physical Involvement. As discussed above, construct validity has never been a strong characteristic of biodata measures, and the ASAP is no exception. The internal consistency reliability of these factors also signify their heterogeneity; the coefficients range from .17 (Independence, Form B) to .74 (School Achievement, Form B), with most of the coefficients in the .40s. Lower levels of internal consistency reliability are common for biodata predictors and, for this reason, it has been suggested that test-retest estimates are the most appropriate measure of reliability for biodata instruments (Mumford & Owens, 1987). Unfortunately, test-retest data are not available for the ASAP.

The Educational and Biographical Information Survey (EBIS; Means & Perelman, 1984) contains items concerning educational achievement, scholastic attitudes and behavior, family relations, work history, arrest record, and drug and alcohol use. The inclusion of several items dealing with moral issues makes the EBIS similar in some ways to a background investigation, where moral character issues are considered when selecting individuals for sensitive positions (McDaniel, 1989). The EBIS items have been empirically-keyed against early attrition from the military and factor analyzed using their keyed values (Steinhaus, 1988). Six factors were derived in this manner: Non-Conformity (i.e., getting into trouble), Alcohol and Drug Use, Criminal Offenses, Quitting Behavior, High School Achievement, and Employment Experience. A later study (McDaniel, 1989) factor analyzed raw item responses to the EBIS and found a similar factor structure but added one factor--Socioeconomic Status. A list of the factors that McDaniel derived, the number of items corresponding with each, and their respective reliabilities, is shown in Table 28.

Table 28

Factor Analysis and Reliability Results for the EBIS

Factor Label	No. of Items	Reliability	
		Alpha	Test-retest
1. School Suspension	7	.81	.80
2. Drug Use	9	.79	.46
3. Quitting School	8	.49	.86
4. Employment Experience	9	.60	.75
5. Grades and School Clubs	8	.67	.82
6. Legal System Contacts	7	.51	.55
7. Socioeconomic Status	5	.63	.82

Note: From "Biographical constructs for predicting employee suitability" by M. A. McDonald, 1989, Journal of Applied Psychology, 74.

The Leadership Effectiveness Assessment Profile (LEAP) is a biodata instrument that is being developed to measure specific traits that are predictive of Air Force Officer leadership behaviors (Appel, Quintana, Cole, Shermis, Grubb, Watson, & Headley-Goode, 1992). Development of the LEAP has proceeded using a conceptual model of officer effectiveness and retention, and the instrument was rationally developed such that items were written to correspond to specific constructs. These constructs were chosen to assess leadership potential, managerial potential, a propensity for commitment to the Air Force, and other related attributes. The constructs assessed in the latest version of the LEAP, the number of items corresponding to each, and their respective test-retest reliabilities are shown in Table 29. The constructs shown in Table 29 are the result of a rational grouping; a factor analysis of LEAP items has not yet been reported.

The LEAP, although it was rationally constructed around specific constructs, has been empirically keyed using an ordinal alternating least squares (ALS) approach (Appel et al., 1992).

The ordinal ALS scaling procedures allowed the researchers to maintain the rational basis for the instrument and at the same time optimize scale score weights using an empirical key. This is done by allowing the correct response for each item to be based on the rational key but weighting all other responses empirically. The test-retest correlations shown in Table 29 are based on the empirically-keyed items.

Table 29

Test-Retest Reliability for the LEAP

Scale	# of Items	Reliability
Transformational Leadership	23	.46
Charisma (a subcomponent of above)	(16)	.41
Transactional Leadership	8	.48
Managerial Decision Making	7	.67
Giving and Seeking Information	7	.66
Team Player Orientation	7	.54
Self-Sufficiency Orientation	7	.49
Physical Fitness Status	9	.63
Institutional Commitment	7	.66
Persistence to Excellence	7	.78
Toleration of Adversity	8	.64
Socialized Power	12	.58
Retention Propensity	7	.66
Faking Detection	12	.43

Note: N = 263.

Note: From "The Leadership Effectiveness Assessment Profile (LEAP): Officer instrument field testing and refinement" (AL-TR-1992) by V. H. Appel, C. M. Quintana, R. W. Cole, M. D. Shermis, P. D. Grubb, T. W. Watson, and A. Headley-Goode, 1992, Brooks Air Force Base, TX: Human Resources Directorate, Armstrong Laboratory.

Biodata Validity

Perhaps the most appealing aspect of biodata measures is the typically high relationships they show with targeted criteria (see Owens, 1976 and Barge & Hough, 1986 for extensive reviews of biodata validities). One of the first reviews of biodata validities (Ghiselli, 1955) found average correlations with trainability and job proficiency in the low .40s. More recently, Reilly and Chao (1982) found an average correlation of .35 between biodata measures and five different types of criteria when only cross-validated coefficients are considered. Barge and Hough (1986) examined median biodata validities across many studies, including both civilian and military research. Their review found a median validity of .25 for training criteria, .32 for job proficiency criteria, .30 with job involvement criteria (e.g., satisfaction, absenteeism, tenure), and .26 with adjustment criteria (e.g., delinquent behavior, unfavorable discharge). These reviews support the conclusion that biodata instruments can yield substantial validity; however, the stability of biodata validity is less assuring.

One of the primary disadvantages of biodata is the tendency for validities to decline rapidly over time. For example, Dunnette, Kirchner, Erickson, and Banas (1960) reported a decline from $r = .74$ to $r = .38$ over two years for a weighted application blank. In a military setting, use of the MAP was eventually discontinued due to validity failure (Walker, cited in Trent, in press). It is thought that this phenomenon may be due to compromise of the scoring key, shrinkage from the original empirical keying process, low sample sizes, and actual changes in applicants (Barge & Hough, 1986); although, few studies have explicitly studied the causes of the temporal degradation of biodata validities. Regardless of the cause, the rapid decline in validity shown for some biodata instruments bespeaks the need to maintain a vigilant effort at monitoring, revising, reweighing, and revalidating operational measures.

One recent study (Rothstein, Schmidt, Erwin, Owens, & Sparks, 1990) has examined factors leading to biodata validity generalization. The study included only biodata items that were based on a careful review of the job for which they were to be used (first-line supervisor), and that exhibited a relationship with performance criteria across organizations. Furthermore, items were included in the questionnaire only if their relationship to the job could be explained in psychological terms. A meta-analysis of validities obtained with the instrument indicated that moderation was produced only by factors such as organization, race, sex, collar (blue or white), years of experience, and education. This study also found that samples collected up to 11 years apart showed little decline in validity. This research suggests that, given some development conditions, the validity of biographical information may be more resilient than has been previously thought.

Validation research on the ASAP has focused on relationships with attrition during the first term of enlistment (cf. Trent, in press). A cross-validation coefficient of .29 was found in this research. Additionally, the ASAP was found to be a better predictor than other variables that are currently operational (high school diploma attainment and AFQT), and the ASAP added significant incremental validity (R change = .22) over these other measures. Validation work on the EBIS has shown that the instrument related .19 with unsuitability discharges from the military (McDaniel, 1989) when an optimally weighted composite was used.

LEAP researchers (Appel et al., 1992) studied the validity of the instrument by examining its relationship with supervisor ratings of ROTC cadet training performance. It was found that the rationally-keyed total score related with the supervisory ratings ($r = .11$); however, use of the empirically-keyed total score improved prediction of the ratings substantially ($r = .45$). Also, when individual scale scores were used to predict the criterion, the full model accounted for 27 percent of the variance in training performance ratings. Additionally, the LEAP showed substantial gains (R change = .51) in the prediction of these ratings over a cognitive predictor (the Air Force Officer Qualifying Test or AFOQT). Although, this may be due to the fact that the AFOQT only accounts for four percent of the criterion variance.

To date, little research has been conducted to examine changes in the validities of these measures over time. This is probably due to the fact that despite the research effort that has gone into the measures, none has been implemented. Because some of the factors that may lead to validity change (e.g., compromise of the items) are a function of operational use, an adequate test of the performance of these instruments over time will probably have to wait until implementation is possible.

Subgroup Differences

Barge and Hough (1986) concluded in their review that there is "little difference in the validities obtained between Whites and minority group members" on biodata inventories (p. 117). Recent research on the measures examined in this chapter tends to support this conclusion. This research also shows that, at least in the measures examined, the races do not differ much in intercept values either.

Race differences on the ASAP (cf. Trent & Quenette, 1992) showed that all minority groups except Native Americans score higher than Whites. Thus, at most of the practical passing score levels that could be set, a lower proportion of minorities would be excluded (rejected) than would White males. A comparison of the regression lines between racial groups indicated that very small but significant slope differences exist for Hispanic males and non-White females, leading to some underprediction for these groups. Trent (in press) concluded that the small practical difference in validities does not outweigh the goal of having a single ASAP scale across all subgroups.

Research with the EBIS (Steinhaus, 1988) evidenced similar findings: minority subgroups tended to score higher on the instrument than Whites, such that just about any practical passing score could be set and, as a consequence, a greater proportion of minority subgroup members would be selected compared to majority group members. Some small race and gender slope differences were found in individual item validities, however. It was suggested that subgroup differences in item validity were the product of differential item functioning across subgroups, but additional research is necessary to verify this hypothesis.

Appel et al. (1992) examined subgroup differences in intercept for the LEAP. Comparisons between males and females, Whites and non-Whites, and low socioeconomic status (SES) and high SES respondents indicated that mean subgroup differences in intercepts were virtually non-existent in the sample studied. Slope differences were not examined in the research.

Differences between males and females on biodata inventories have traditionally been more common than race differences. For example, Owens and Schoenfeldt (1979) demonstrated that male and female responses to a biographical questionnaire yield different factor structures. In fact, some reviewers of the biodata literature have suggested using different scoring keys for the sexes (Reilly & Chao, 1982). However, recent research (Rothstein et al., 1990) has shown that gender does not introduce substantial variability in biodata validities. In addition, work with the ASAP, EBIS, and LEAP indicates that separate keys may not be necessary, as gender differences on these inventories were not large.

Fakability

As was found to be the case with interest and personality measures, there is some evidence that people can respond more positively to biographical instruments when they are instructed to do so (e.g., Schrader & Osburn, 1977). However, there are fewer studies that have investigated faking in the biodata realm than there have been for other non-cognitive predictors (Barge & Hough, 1986), and there is contradictory evidence about the extent to which faking is likely to be a problem on these instruments. For example, Goldstein (1971) found high rates of distortion in a sample of 94 job applicants, but Cascio (1975) found a high degree of accuracy for biodata in a non-selection situation.

Several recent studies have looked at factors that may detect and reduce faking on biodata forms. Trent, Atwater, and Abrahams (1986) examined responses to a biographical questionnaire under conditions where groups of applicants who were motivated to fake received varying warnings cautioning against faking. They also examined the effects of verifiable items and an empirically-based scoring key on faking behavior. Trent et al. found that the use of warnings indicating that responses would be verified led to significant but small decreases in faking, that verifiable items tend to be faked to a lesser degree than non-verifiable items, and that the use of an empirically-based key minimized the impact of faking. Additionally, this study demonstrated that applicants who were told that their responses would affect their career showed less distortion than a control group who was instructed to "look good."

Another study (Kluger, Reilly, & Russell, 1991) compared faking on a biodata instrument when the form was item-keyed versus option-keyed. When items are option-keyed, each response option is empirically weighted; whereas item weights typically give the most weight to an option at one end of the response continuum and successively lower weights to each consecutive option. For example, if an item has five possible responses, option-keyed weights may be 0, -1, 0, 0, +1, while item-keyed weights may be 1, 2, 3, 4, 5. It was hypothesized, therefore, that option-keyed forms would be more difficult to fake. The hypothesis was supported by the study. Kluger et al. (1991) also examined the usefulness of response latencies for detecting response distortion. Latencies were found not to vary between subjects who were asked to answer truthfully and those who were asked to respond as if they were "actually applying for a job."

A more effective method of detecting faking was examined by Hanson, Hallam, and Hough (1989). That study used the ASP which includes both a biodata (ASAP) and a personality (ABLE) component. The Social Desirability scale from the ABLE was found to be moderately successful in the detection of faking and was best at detecting effective fakers compared to less effective fakers. This finding suggests that the combination of biodata items with personality items that can be used to detect faking may be another method of combating response distortion.

The recent studies on faking indicate that although it is reasonable to suspect that people can distort their responses on biodata instruments, steps can be taken to reduce the problem. That is, the use of verification warnings, verifiable items, and an empirical option-keyed scoring procedure may decrease the effect of distortion motivation. Further, the inclusion of a social desirability or faking scale may allow for the identification of people who do fake. However, the identification of fakers only raises the issue of what to do with suspect profiles. Additional research is necessary to determine how these responses should be treated. Finally, some research has indicated that biodata predictors are less susceptible to faking than are personality measures (Kleinke, 1992). There is also some evidence that even though people can fake their responses to biodata, the extent to which they actually fake under actual testing/selection conditions is not large (Becker & Colquitt, 1992).

Conclusions Regarding Biodata

The review of current military and civilian research on biographical information suggests several conclusions.

- Biodata are effective and valid predictors of a number of important criteria. Research has indicated that biodata validities can be made generalizable and stable (Rothstein et al., 1990), thus these measures are worthy of continued consideration as supplements to cognitive predictors of military performance.

- There is evidence that biodata may have incremental validity over cognitive measures, especially when predicting non-performance criteria such as attrition (e.g., Trent, in press).

- Biodata do not yield large differences between the races and evidence of differential validity is slight. However, more research is necessary to determine whether some items function differently for different races. Although it has been suggested that different keys be developed for each sex, research on military measures and a meta-analysis indicate that sex differences may not be as much of a problem as previously thought.

- Although biodata measures are possible to fake, research indicates that faking may not be prevalent. To the extent that faking does occur, steps can be taken to reduce and identify it.

- If biodata measures are made operational, it is critical to track their performance over time and maintain the instruments accordingly in order to avoid validity failure.

Finally, one additional strength of biodata is that some measures (e.g., the EBIS) account for variability in attrition that has traditionally been predicted by educational attainment criteria. Educational credentials have come under fire lately (cf. Laurence, in press) because they restrict entrance to the military for identifiable groups of individuals (e.g., GED recipients). Biodata instruments provide a compensatory measure such that no one particular characteristic will be likely to exclude an individual. Thus, biodata may face less implementation resistance than other predictors of military adjustment. It should also be noted, however, that this strength could become a weakness if biodata were perceived as being merely a proxy for questions that have been prohibited.

Conclusions

Our review of non-cognitive predictors suggests several general conclusions regarding these measures, as well as additional research that would be of valuable as the Services consider the implementation of an expanded predictor battery.

First, validation research indicates that the Services would be likely to improve prediction by adding non-cognitive measures. However, it is important to be clear about the criteria that are being predicted. Personality measures may add incremental validity over cognitive measures when "will-do" performance factors are examined. Interest measures increase the predictability of job satisfaction, suggesting their value as a classification tool. Biodata measures have been shown to be related to attrition. Thus, non-cognitive measures may become increasingly attractive as our conceptualization of the criterion domain unfolds.

Second, non-cognitive measure are less likely than traditional cognitive measures to show large differences between races. Although differences have been found on personality and interest measures, the direction of these difference (often in favor of minorities) indicates that adverse impact against minorities is unlikely to be a problem. The possibility that the use of these measures will lead to reverse discrimination is slim but may be worth investigating nonetheless.

Third, differences between the sexes on personality and interest measures are likely to be substantial. In diagnostic applications of these measures separate norms have often been used. The appropriateness of this procedure in a selection and classification context needs to be considered further.

Fourth, faking is possible on these measures, but it is possible to detect faking in may cases. Further research is necessary to determine how to best reduce socially desirable responding and purposeful faking, and how to deal with suspect response profiles. The conduct of a comprehensive review of the faking and social desirability literature would be an important step in organizing our knowledge in this important area. The literature reviewed here suggests that the possibility that faking may occur does not impoverish completely the utility of non-cognitive measures.

Finally, non-cognitive measures may be valuable because they offer alternatives to problematic variables that are currently used by the Services. Two examples were presented here. First, interest measures may make a valuable substitution for naive recruit preferences during the enlistment/classification process. Interest profiles could be used to pre-select an occupational category for a recruit, recruit preferences could then be used to reduce the choices within a category. Alternatively, interests could be used for guidance counseling purposes to influence recruit preferences, thereby making them less naive. Second, biographical data may accomplish similar goals to those of educational attainment variables. However, biodata may prove to be less controversial than measures that restrict military enlistment from identifiable subgroups.

A revised list of the constructs and measures reviewed in this Chapter appears in Table 30.

Table 30		
Individual Differences Attributes and Constructs and Selected Military Measures		
Broad Attributes	Related Constructs	Selected Measures Developed by the Services
Cognitive		
G_c - Knowledge or Crystallized Intelligence	Knowledge of general information Word knowledge	ASVAB [GS, WK, AS, MC, EI] OSB, AFOQT
G_f - Broad Reasoning or Fluid Intelligence	Inductive reasoning Conjunctive reasoning Deductive reasoning	AFOQT ECAT Mental Counters ECAT Sequential Memory ECAT Figural Reasoning
G_v - Broad Visual Intelligence	Spatial visualization Spatial orientation	BAT, AFOQT, OSB ECAT Assembling Objects ECAT Orientation Test ECAT Integrating Details
SAR - Short Term Acquisition and Retrieval	Recency memory Word span	BAT
TSR - Long Term Storage and Retrieval	Associational fluency Expressional fluency Ideational fluency	
G_s - Broad Speediness	Visual scanning Visual matching	ASVAB [CS, NO] BAT, AFOQT ECAT Target Identification
G_a - Auditory Intelligence	Discrimination among sound patterns Auditory cognition of relations	DLAB, ARC, Superdit
G_q - Quantitative Thinking	Computational fluency Numerical computation	ASVAB [AR, MK] OSB, AFOQT
Eng - English Adeptness	Word parsing Phonetic decoding	
Psychomotor		
Dexterity	Finger dexterity Manual dexterity	
Basic Movement Speed and Accuracy	Reaction time Control precision Speed of arm movement	
Perceptual-Motor Movement Control	Multi-limb coordination Rate control	BAT ECAT Tracking 1 ECAT Tracking 2

(Continued)

Table 30

Individual Differences Attributes and Constructs and Selected Military Measures (Continued)

Broad Attributes	Related Constructs	Selected Measures Developed by the Services
Physical		
Muscular Strength	Muscular tension Muscular power Muscular endurance	Air Force Strength Factor
Cardiovascular Endurance	Cardiovascular endurance	
Movement Quality	Flexibility Balance Coordination	
Personality		
Extraversion	Sociable, Gregarious Ambitious, Achievement-Oriented	OSB, ABLE, AAPP
Emotional Stability	Emotional, Anxious, Depressed	
Agreeableness	Good-natured, Cooperative	
Conscientiousness	Dependable, Responsible	
Intellectance	Curious, Broad-minded	
Interest		
Realistic	Practical, likes hand-on work	BAT, VOICE, AVOICE
Investigative	Curious, likes academic endeavors	
Artistic	Creative, likes self-expression	
Social	Friendly, likes people	
Enterprising	Ambitious, likes managing & directing	
Conventional	Concrete, likes exactness in work	
Biographical Information		
?	?	ASAP, EBIS, LEAP

Source: Cognitive (Horn, 1989); Psychomotor (Fleishman, 1967; Imhoff & Levine, 1981; McHenry, 1987), Physical (Hogan, 1991a); Personality (Barrick & Mount, 1991; Digman, 1990; Tett, Jackson, & Rothstein, 1991); Interests (Holland, 1983).

THIS

PAGE

IS

MISSING

IN

ORIGINAL

DOCUMENT

pg 100

V. DIRECTIONS FOR RESEARCH

Douglas H. Reynolds and Teresa L. Russell

Changing missions and limited resources are likely to result in changes in military job requirements--changes that may affect selection and classification individual differences measurement. For example, DoD involvement in the war on drugs or in the defense of our borders against illegal alien entry may require more small plane pilots and small intervention units that operate autonomously. Also, in response to funding limitations, the Services are redesigning jobs to make the workflow more efficient. This may mean that the Services are headed toward more general jobs and fewer specializations. It is also possible that structural changes in jobs will emphasize the importance of team performance or that new technology will result in cognitively demanding jobs.

This chapter has two parts. In the first part, we discuss change--the move to generalist jobs, team effectiveness, and technological advancement--within the context of individual differences measurement. And, we discuss changes underway in the civilian sector. In the second part, we revisit the research objectives outlined in Chapter I and present new, revised objectives based on research reviewed throughout this report.

Preparing for the Military Workplace of 2000 and Beyond

Specialization to Generalization

As we noted in our first report (Russell et al., 1992), the Services plan to move away from highly specialized jobs as they downsize. This transition has several implications for personnel operations in general (e.g., additional job analyses as jobs are combined) and selection and classification in particular. The current degree of specialization in jobs in the military has been driven by technology; complex hardware requires specialized knowledge and thus it is difficult to form general jobs. If the military is to be successful in moving toward a more generally capable workforce, individuals may need to be more versatile and capable of handling increasingly diverse and complex tasks. Training investment will probably need to increase in order to bring recruits up to a fully functioning level of performance in a number of areas. These changes also have implications for future aptitude requirements and selection standards.

If future increases are made in the training investment in each recruit, the cost of attrition will necessarily escalate. Additionally, as job tasks become more general, the marketability of the skills learned in the military may increase and perhaps affect the rate of attrition. Thus, predictors of job satisfaction, commitment to the military, and attrition will increase in their utility. As we have indicated in the previous chapter, interest measures have been shown to predict satisfaction in military occupational fields, and biographical information may be an effective predictor of early attrition.

The development of a more generally capable workforce depends not only upon the training of individuals, but also upon the selection of people who have the abilities to learn efficiently and perform complex jobs effectively. This may suggest that future recruits will need to be "smarter" in the traditional general cognitive ability sense of the term. Additionally, it has been suggested (e.g., Gorden, in press) that recruits will need to be more adaptable. Unfortunately, "adaptability" has not been well defined in the field of psychological assessment. Adaptability (or the ability to change oneself to meet the demands of a particular situation) may be simply a function of general intelligence, in which case the implications for personnel selection and classification are straightforward--select people of higher cognitive ability. It is also quite possible that adaptability is related to various personality constructs, such as Intellectance (Openness to Experience). Further work is necessary to operationally define what is meant by "adaptability" as a criterion, before it will be possible to indicate the best predictors of the construct.

It may be that adaptability will prove to be difficult to define outside the features of the situations to which people must adapt. Some general features of a diverse, rapidly changing, or unstable task environment may be hypothesized, however. It is likely that more diverse work environments (that is, those requiring people to be adaptable) will provide less structure than those that are less variable, thus people who are able to perform well under conditions that are less structured may be more adaptable. It is also possible that adaptability requires a certain level of self-motivation.

In all likelihood, performance in complex, rapidly changing work environments may be best predicted by considering the interactive effects of a range of individual difference variables. Some research that has looked at these interactive effects is discussed below.

Aptitude, Trait, and Environment Interactions

If we are to more fully understand performance in complex and changing environments, future research on the prediction of performance should focus not only on the constructs we measure and the quality of our measurement techniques, but also on the interrelationships between human characteristics and job environments. There is evidence that cognitive abilities and various non-cognitive traits and dispositions may interact with each other and with characteristics of the task environment, thereby affecting performance outcomes (e.g., Snow, 1989). For example, Snow (1989) reports the results of a series of studies that looked at the relations between prior knowledge of a subject area, the amount of structure provided during instruction, and several personality constructs. These studies indicate that differences in learning rate as a function of instructional technique are moderated both by cognitive aptitudes and non-cognitive characteristics such as Ascendancy (i.e., Extroversion) and Responsibility (i.e., Dependability). Another study (Peterson, cited in Snow, 1989), showed that high ability subjects who also have high levels of anxiety tend to learn better from structured instruction--just as do low-ability, low-anxiety subjects. However, able and non-anxious subjects, as well as less able and anxious subjects, learned better under less structured instruction. Non-cognitive and cognitive attributes therefore appear to jointly moderate learning outcomes; features of the environment, such as structure, moderate this effect.

The notion that prediction, and the understanding of the factors that affect prediction, will be improved by looking to a broad set of interacting variables has been emphasized by Sternberg (in press). Sternberg has proposed a "person-context" model for studying human potential. The model suggests that a person's capability in a given situation will be determined as a joint function of characteristics of the person (including abilities, knowledge, personality, motivation, and style), the role required (e.g., leader versus follower), the situation itself (e.g., physical comfort versus discomfort), values (such as valuing conformity over independence), and luck. Although the model specifies a range of variables that may be important for predicting success, it says little about the interrelationships among the variables. It is likely that the explication of the relationships among these variables will lead to significant gains in our understanding.

Motivation is also a key determinant of performance in just about any situation (e.g., Campbell, in press). Good performance in environments that require a broad set of skills may require greater levels of motivation than more specialized environments. For example, in environments where people have responsibilities for accomplishing a broader set of activities, it may be increasingly important for individuals to seek out information about how to perform a task rather than to rely on trained knowledge. The next section describes some research on individual difference variables that may impact motivation.

Variables Affecting Motivation

Motivation has been described as the direction of attentional effort, the proportion of total attentional effort directed toward a task, and the extent that effort is maintained over time (Kanfer & Ackerman, 1989). Motivation has been seen by some to be a choice behavior, while others have acknowledged that it is also a function of dispositional variables that impact goal-directed behavior (cf. Kanfer, 1990). These dispositional variables are of interest here because, if these variables can be measured, they may be of use in a selection and classification context. Other factors affecting motivation that are more likely to change over time (e.g., state variables such as expectancy and instrumentality) would be less useful for decision making.

Individual differences in dispositional characteristics that may affect motivation show potential as new predictor avenues. These variables may be especially promising for predicting later rather than more immediate performance. There are many different models of human motivation that account for motivational variation with a wide range of variables (cf. Kanfer, 1990). We focus here on two specific motivationally-related dimensions that are likely to predict performance in complex situations: Achievement Orientation and Action Control.

Achievement Orientation refers to the cognitive goal structures that serve to guide cognition and behavior. Recent findings from the Army's Project A demonstrate that an Achievement Orientation composite from the ABLE (consisting of Self-Esteem and Work Orientation variables) predicts "will-do" criteria such as Effort and Leadership, Personal Discipline, and Physical Fitness and Military Bearing (McHenry et al., 1990).

Several constructs have been proposed to be related to Achievement Orientation, including Work Orientation and Job/Task-Specific Self-Confidence. Work Orientation refers to a willingness to devote oneself to work, by working long hours and meeting organizational goals (e.g., Day & Silverman, 1989). Helmreich, Sawin, and Carsrud (1986) found that individual differences in Work Orientation did not predict performance during training in a study using airline ticket counter attendants, however it did predict performance after six months on the job. Job/Task-Specific Self-Confidence refers to the self-assessment of one's ability to successfully execute the behaviors required to produce the desired outcomes. Indicators of self-confidence--such as self-esteem--have been shown to relate to career choice, perceived ability to perform the job, and anticipated satisfaction (Brockner, 1988). Task-specific indicators, such as self-efficacy, have also been shown to affect performance and persistence in novel and difficult tasks (Bandura, 1986). Current research on self-efficacy will be reviewed in a later section of this chapter.

Action Control refers to the self-regulation of attention during learning (Bandura, 1986; Kanfer, 1990; Kuhl, 1981; 1984). The construct relates to the extent to which persons sustain and protect task-related cognitive processing in the face of distracting stimuli. Kanfer and Ackerman (1989) found that performance on a complex cognitive task was negatively associated with the degree to which individuals disengaged from task performance to monitor their level of performance and to the frequency of negative emotional reactions to performance. These effects are presumably due to the attentional drain that the intervening thought places on attention. Further, the effects are likely to reflect individual differences in self-regulatory skills that are required for both learning and sustaining performance. Sarason, Sarason, Keefe, Hayes, and Shearin (1986) also have provided evidence for stable individual differences in the frequency of intrusive thoughts during task performance. Interestingly, although self-regulation may impede performance during learning (due to interference), it is likely that it may also improve performance after the task has been learned (Kanfer & Ackerman, 1989).

These findings suggest that dispositional characteristics that affect motivation may have both interactive and direct influences on performance. Additional research is required to further explicate the role of these characteristics in motivation and the relationships between motivational characteristics, task complexity, and the passing of time.

Another related area of current research that may be relevant to the prediction of complex task performance under changing conditions is that of self-efficacy. A brief review of this research follows.

Self-Efficacy

Self-efficacy, the belief in the capacity to exercise control over one's own functioning and over environmental demands, has been hypothesized to affect thought, motivation, emotion, and performance (Bandura, 1986). Self-efficacy includes beliefs about one's basic skills and the capacity to orchestrate them into successful actions. Perceptions of self-efficacy concern an estimate of what one knows and extends to a prediction of how well one will be able perform in a given circumstance. Research on self-efficacious belief has shown relationships with these beliefs and dependent variables involving cognitive, motivational, affective, and selective processes (see Bandura, in press, for a summary).

In the cognitive arena, Bandura (in press) describes self-efficacy as being related to goal commitment and the visualization of successful performance scenarios that serve to guide behavior. Beliefs of self-efficacy may be critical for the performance of complex jobs. In a program of research on the influence of self-efficacy beliefs on cognitive processes, Bandura (e.g., Wood & Bandura, 1989) has manipulated subjects' beliefs about their ability to perform an experimental management task. He found that subjects with lowered levels of self-efficacy become more self-doubting, more erratic in their thinking, set lower goals, and become less productive. Bandura (in press) has thus postulated a causal role for self-efficacy in performance both directly and indirectly through the formulation of goals and performance strategies. However, the relationship is recursive--past performance influences self-efficacy beliefs which, in turn, affect subsequent performance.

As previously mentioned, self-efficacy also plays a role in motivation and self-regulation. Self-efficacy beliefs may affect causal attributions for success and failure, and may moderate the motivating potential of rewards for successful performance. Through their effects on goal setting, efficacy beliefs may moderate the level of effort applied to a task after initial failure (Bandura & Cervone, 1983). Furthermore, a high sense of self-efficacy may lead to the development of higher goals after an initial goal has been reached (Bandura, in press).

In relation to affect, efficacy beliefs may affect how averse events are interpreted, the degree to which individuals have control over distressing thoughts, and the development of courses of action that reduce the adversity of hostile environments (Bandura, in press). Beliefs of personal efficacy are also related to the selection of the environments and activities people choose for personal and professional development (Lent & Hackett, 1987).

Despite the wide range of variables that have been shown to be related to self-efficacy, the concept may not function well as a predictor in the personnel selection arena. It may, in fact, be a better training goal, as self-efficacy beliefs are likely to change as a result of experience. Much of Bandura's work (e.g., 1982) involves systematic efforts to increase self-efficacy. To the extent that self-efficacy beliefs change over time, their long-term predictive ability is reduced. Additionally, self-efficacy is affected by the outcome of task performance. As such, while efficacious belief may improve performance, good performance also leads to efficacious belief--and performance is in part a function of ability. As a result, even though self-efficacy may be an important ingredient for success, the construct itself is likely to be of little use in a selection context. Future research needs to examine whether more stable dispositional characteristics (such as locus of control or self-esteem) could be used as similar but more stable predictors. It may be that self-efficacy is a more useful construct to consider when designing training for new and complex jobs.

Summary

As the Military Services move from very specialized jobs to jobs with more general areas of responsibility, the jobs will likely become more complex and require greater cognitive ability. Additionally, a greater degree of what some have called "adaptability" may also be important (e.g., Gorden, in press). This section has described current research in a few areas related to the problem of predicting performance for an increasingly complex set of jobs and to an ability to adapt. In particular, as military jobs become broader and more complex, selection and classification researchers may benefit by devoting attention to interactive combinations of predictor measures, dispositional characteristics that may affect motivation, and the identification of constructs that relate to the development of self-efficacious beliefs.

Team Emphasis

The second theme that emerged from our discussions with military personnel experts was that in the future, the Services may need the capability to form small, quick-reaction teams of specialized personnel for conflicts around the world. For example, the war on drugs and other mission changes may result in more special operations and low intensity warfare of short duration. The transition to reliance on teams suggests that selection and classification research should examine the characteristics that people need to perform successfully in a team environment, as well as how to best measure those characteristics. Additionally, it will be important to better understand the components of team performance, however, that topic is beyond the individual differences scope of this report.

It is probable that information about personality constructs will help to predict performance in social situations such as work teams. For example, information presented in Chapter IV indicated that personality constructs assessed with the ABLE added incremental validity for the prediction of several "will-do" performance criteria, such as Effort and Leadership, Personal Discipline, and Physical Fitness and Military Bearing. To the extent that these criterion constructs are important for performing effectively as a part of a team, personality constructs will be important variables to consider when selecting people for that environment. If the folk wisdom concerning what makes a valuable contribution to a team is correct (e.g., doing one's part to help achieve a mutual goal; taking responsibility for others), clearly the criteria predicted by the ABLE are relevant components of team performance.

In addition to the personality constructs already discussed, it may also be important to consider abilities that relate specifically to behavior in social settings when attempting to predict individual performance in a team environment. The issue of social abilities has received sporadic attention over the history of psychological research. The following section discusses some of the research in the domain of Social Intelligence.

Social Intelligence

Social intelligence was defined by Thorndike in 1920 as the ability to understand others and to act wisely in social situations. There has been periodic interest in the topic since that time, although many have used different terms for the topic and none has adequately validated the concept.

Comprehensive reviews of the literature on social intelligence have been conducted by Walker and Foley (1973) and Sternberg (1985). Based on these reviews, it is evident that social intelligence may have some potential as a predictor of performance in interpersonal situations. However, work remains to be done to define the concept and figure out how to best measure it. In fact, the two-part definition of the construct coined by Thorndike (1920) represents some of the difficulty researchers have had in determining the nature of the construct. That is, is social intelligence an ability or a way of behaving in social situations?

Despite the interpretive and definitional difficulties surrounding the concept, a number of tests have been developed to measure it; many of these tests were reviewed by Walker and Foley

(1973). Included in the study were the George Washington Social Intelligence Test developed by Moss, Hunt, Omwake, and Woodward (1955); the Social Insight Test developed by Chapin (1967); and a series of tests developed by Guilford (1967). The George Washington Social Intelligence Test contains several subtests, including one in which respondents must choose the most appropriate solution to a social problem, one in which respondents have to identify photographs that they had been shown earlier, and one in which they have to identify the emotion associated with a given statement. The Social Insight Test contains descriptions of different social situations. Examinees are asked to read a description and choose from among four different comments pertaining to the situation. Among the tests developed by Guilford was one in which the respondent had to match faces on the basis of mental state similarities and one in which respondents had to match a facial expression with a recording of a vocal expression. For each of these tests, it is easy to see that the test developers focused on measuring social understanding rather than sampling social behavior.

More recently, researchers at the Army Research Institute (Busciglio, Palmer, & King, 1991) have developed a model of social intelligence that makes hypotheses about the construct's antecedents, components, and consequences. The model posits that the interaction between general cognitive ability, dispositional characteristics, and social environment impacts the development of social awareness and social goal setting. These two components combine over time and experience to form tacit social knowledge--a sort of social common sense. It is the application of tacit social knowledge in a given situation and for the attainment of a specific, valued goal that leads to socially intelligent behavior. Busciglio et al. (1991) also cite research indicating that social intelligence may be related to leader, manager, and team effectiveness. Although the model specifies the relation between the two components of Thorndike's 1920 definition (i.e., social understanding and intelligent social action), research has yet to validate the structure of the construct as proposed by the model.

Although there is currently little support for its validity as a unitary construct, the concept of social intelligence may have some potential as a useful predictor after further research is conducted. One avenue of research was suggested by Walker and Foley (1973) when they observed that researchers in related areas of study (e.g., social and person perception) have discussed similar concepts; future research should identify common elements across these research areas. Another area of future research involves the specification of the performance domain that is likely to be related to social intelligence (e.g., Barnes & Sternberg, 1989). Once we have a better understanding of what socially-intelligent behavior is, we will have a better chance of determining its consequences and formulating its prediction. Such an approach has been suggested by Busciglio et al. (1991).

Social intelligence may be one of probably many characteristics that relate to interpersonal competence in social (team) situations. Although the attributes that best predict performance on teams have not been fully explicated, measurement methods have been developed for predicting performance in these environments. Specifically, simulations often have been used to replicate the types of activities that are important in a social work environment. The next sections describe these measurement methods and the constructs they typically purport to assess.

High-Fidelity Simulation (Assessment Centers)

The assessment center method has evolved from a number of attempts to measure complex performance with a combination of assessment techniques using several observers. Typical assessment centers involve the simulation of job activities in a manner that allows for multiple measures of an individual's performance and the integration of those measures into an overall evaluation of the individual (cf. Thornton & Byham, 1982). The use of job simulations that include role-plays and group discussions in assessment centers, allows for social skills that are relevant in team work environments to be assessed.

Assessment centers have an established history within psychology. In the early 1900s, German psychologists asserted that in order to assess leadership potential, an individual's total personality should be assessed--not separate abilities. Thus, these psychologists developed complex situations that allowed for multiple measures of behavior in a naturalistic setting (Ansbacher, 1941). Modern variants on the assessment center method often include management tasks such as an in-basket exercise, social situations such as a leaderless group discussion, analytical tasks such as a scheduling activity, and many also include paper-and-pencil tests of cognitive ability (Thornton & Byham, 1982).

A number of abilities and characteristics can be assessed in assessment centers, although, in the past, most applications have been oriented toward managerial selection and thus tend to measure ability dimensions that are important for higher level jobs. More recently, their use has expanded to cover a variety of jobs (e.g., police and fire personnel), and thus, they have been used to assess a broader variety of knowledge, skills, and ability. Assessment centers are commonly used to assess characteristics and abilities such as planning and organizing, leadership potential, decision making, interpersonal effectiveness, sensitivity, and flexibility. Because assessment centers allow for the observation of social behavior, they have also been suggested for use in research on social intelligence (Busciglio et al., 1991).

A recent meta-analysis of assessment center validities found a mean corrected validity of .37 (with a corresponding variance of .017), suggesting that assessment centers are valid predictors of performance across a range of situations (Gaugler, Rosenthal, Thornton, & Bentson, 1987). Although Gaugler et al. concluded that these validities generalize, a number of moderators were found. Specifically, assessment centers tended to show higher criterion-related validities when females were assessed, when several different exercises were used, when psychologists were used as assessors, and when peer ratings were included.

Assessment centers are likely to be appropriate for use in military settings where the careful measurement of leadership and managerial skills--and the prediction of subsequent performance in these areas--is critical. The measurement strengths of assessment centers notwithstanding, it should be noted that the procedure also has some practical disadvantages. Specifically, the method can be administratively complex and expensive to administer. The complexity often precludes testing at multiple sites and therefore assessment centers are often implemented at a central location--this often adds the cost of travel for the examinees to the expense of the procedure.

Low-Fidelity Simulation

In contrast to the established history of the assessment center, the low-fidelity simulation is a relatively new measurement procedure. A number of researchers have discussed similar techniques and applied different names. For example, Sternberg has referred to "tests of tacit knowledge" (e.g., Sternberg, 1990), Motowidlo, Dunnette, and Carter (1990) have discussed the "low-fidelity simulation," and Hanson and Borman (1990) describe a "situational judgment test."

The low-fidelity simulation has been developed from the same logic as the assessment center--that behavioral samples can be elicited from representations of the task environment and that these samples will be effective predictors of future performance. The low-fidelity simulation, however, attempts to replicate the task environment with less realism and thus the procedure is less costly. Some forms of the procedure are administered verbally, such as the situational interview (Latham & Saari, 1984), while others are paper-and-pencil measures.

Development of a low-fidelity simulation may involve the collection of critical incidents for the job(s) in question (e.g., Motowidlo et al., 1990). The critical incidents form the basis for the description of problem situations; the situations can be selected to tap specific performance areas. Next, solutions to each problem are collected and scaled in terms of their effectiveness. The final items provide a description of the situation and several response alternatives; examinees choose the best and worst alternative solutions. Scores resulting from the procedure are a function of the number of times the examinee selects the most highly scaled item as being the most effective solution and the lowest scaled item as being the worst option.

As with the assessment center, low-fidelity simulations have most often been used for the prediction of performance for supervisory or managerial jobs. The constructs tapped by the measures are in part domain-specific; that is, the tests are often developed to tap constructs that have been hypothesized to be important in specific situations. For example, Motowidlo et al. (1990) developed a test to tap interpersonal performance areas such as leadership, assertiveness, flexibility, and sensitivity, as well as the problem-solving areas of organization, thoroughness, drive, and resourcefulness. Hanson and Borman (1990) reported the use of a similar test as a criterion measure for supervisory performance in the Army; the test was shown to be related to supervisory job knowledge as measured by supervisory role-play simulations and ratings of supervisory performance.

The validity of the low-fidelity simulation has shown promise. Sternberg (1990, in press) reports that scores on a test of tacit knowledge show similar correlations to managerial performance as do intelligence tests for predicting school performance (about .40), although this finding is based on a small and restricted sample. Motowidlo et al. (1990) found their test to correlate with ratings of interpersonal effectiveness (.35), problem solving effectiveness (.28), communication effectiveness (.37), and overall effectiveness (.30) in a supervisory selection environment. Although the situational judgment test used by Hanson and Borman (1990) served as a criterion measure, relationships with other measures of supervisory performance help to understand what is being measured by the test. It was found that the test correlated with supervisory knowledge (.40), ratings of leading/supervising effectiveness (.24), and higher fidelity role-play simulations (.20).

The low fidelity simulation may prove to be a useful tool for predicting success in specific military jobs; however, its usefulness is limited by the relative newness of the procedure. More work is necessary to determine the constructs that are best assessed by these simulations. Because of the domain specificity of the items that are typically included in these measures, their usefulness across a broad range of jobs is limited--most applications have centered on supervisory/managerial prediction, suggesting that the procedure would be best suited for officer-level jobs. Another disadvantage of the technique is that the scaling of the responses can be difficult. Several strategies have been used (e.g., Sternberg, in press) and most are dependent upon the normative values of the individuals who perform the scaling tasks. More research is needed to determine the implications of alternative scaling procedures for these items.

Summary

It is likely that changes in the DoD mission will lead to an increased reliance on small teams. This change presents an opportunity for selection and classification researchers to identify new predictors of individual performance in a team environment and to refine current measures (e.g., the ABLE) for predicting team criteria. In this section, we reviewed one area, social intelligence, that may provide some direction about the types of variables that may be used to predict social performance components, as well as define social performance. Measurement methods that are likely to tap important social constructs were also discussed.

Preparing for Technological Advancement

The military is technologically dynamic. For example, advances in shipboard technology have resulted in phasing out the Navy's Boiler Technicians and phasing in a more complex Gas Turbine Technician rating (Russell et al., 1992). In addition, the Air Force has moved to using pneudraulic systems; hydraulics are out.

Expectations about technological advancement reinforce the continuing need for basic cognitive abilities measurement research. New systems may require different or more complex individual technical or cognitive attributes, depending upon how "smart" the systems are and how user-friendly maintenance and operational procedures are made. New technology may result in a general shift from concrete observable tasks to cognitively-demanding, non-observable activities (Glaser, Lesgold, & Gott, 1991). In sum, changes in technology may result in more jobs that require advanced technical attributes, or jobs that require attributes that are not well measured by current aptitude measures. Several key basic abilities measurement areas are summarized below.

Information Processing Constructs

The Air Force's LAMP is probably one of the best known basic abilities measurement research projects. Its goals are to denote the basic parameters of learning ability, to develop techniques to assess cognitive ability, and to investigate the feasibility of applying a cognitive model-based system to psychological assessment (Kyllonen, 1985). Over the course of the last six years, LAMP researchers have developed over 1000 computerized tests, four versions of their Cognitive Abilities Measurement (CAM) battery, a taxonomy of cognitive attributes, and theoretical notions about information processing (IP) and how IP relates to reasoning ability (Kyllonen, 1991; Kyllonen & Christal, 1990). Unlike traditional cognitive attribute-based taxonomies that are rooted in factor analysis, the CAM taxonomy is derived from an IP framework. It includes seven kinds of processing variables as shown in Figure 1. The seven process factors are fully crossed with three major types of stimuli: verbal, quantitative, and spatial. LAMP researchers have designed tests for each cell of the CAM taxonomy. For example, there are three working-memory capacity tests: verbal, quantitative, and spatial working-memory capacity.

	Types of Stimuli		
Processing Variables	Verbal	Quantitative	Spatial
Working-memory capacity			
Processing speed			
Declarative knowledge			
Procedural knowledge			
Declarative learning			
Procedural learning			
Temporal processing			

Figure 1. The Cognitive Abilities Measurement Taxonomy

The interrelationships among the CAM components have not been fully explored. Initial results, however, have made an important contribution to the understanding of IP and its linkage to cognition in general. Data suggest that there is a strong general factor underlying performance on the CAM tests, that working memory capacity measures load very highly on the first general factor, and that the working memory capacity factor is highly correlated with G_f (Kyllonen & Christal, 1990).

Selected experimental LAMP measures are currently being converted into applied tests in a separate Air Force Automated Personnel Testing (APT) project. The Air Force plans to examine the validity of APT measures against performance on self-paced intelligent tutor training and later against performance in training schools.

There is a need to continue linking the traditional factor-analytic based cognitive abilities with information processing constructs as Kyllonen and Christal have done (1990). What is the relationship between the speed of information processing and perceptual speed as measured by traditional tasks? Can all of these measures be subsumed by G_s or *g*? Kyllonen and Christal suggest that test/item content (e.g., quantitative, verbal, spatial) plays a minor role in information processing. What role does the sensory/perceptual mode of processing (e.g., auditory, visual) play? Finally, what implications do basic research findings have for applied research settings?

Componential Analysis

Information processing research in the 1960s and 1970s, such as Shepard and Metzler's (1971) studies showing that mental rotation time scores increase linearly as the angle of rotation required increases, spawned research on componential analysis and decomposition of cognitive problems into mental processes. Such work has enhanced our knowledge of time scores, how to use them, and how not to use them (e.g., individual slope scores have consistently been proven unreliable) (Lohman, in press).

It is now becoming apparent that componential analysis will not revolutionize abilities measurement (Lohman, in press; Sternberg, in press). For example, Sternberg (in press) noted that tests of elementary cognitive processes neither correlate well with each other nor show relationships with external criteria; additionally a great deal of time is required to produce reliable component scores. Furthermore, individual differences on basic information processing components do not explain differences in overall performance on tasks (Lohman, in press).

Definition of New Cognitive Constructs

Cognitive processing research, together with technological advancements in personal computers, has made it possible to identify and measure new constructs. Yet, scientific progress may not have much impact in applied settings simply because cognitive measures are highly correlated with each other. There is also strong evidence for a general spatial factor underlying performance on all cognitive measures.

Even so, some research has pointed to constructs that are conceptually and theoretically quite novel and may be of interest in future individual differences research. For example, the Navy has sponsored experimental work with dynamic displays (e.g., Hunt, Pelligreno, Abate, Alderton, Farr, Frick, & McDonald, 1987). Hunt et al. developed several measures that involved extrapolation of time. In an Arrival Time test, subjects watched an object proceed toward a fixed point. One-quarter to one-half of the way across the screen, the object disappeared from view. The subjects had to press a key to indicate when they thought the object reached the fixed point. In other tests, subjects made extrapolations of incomplete paths as well as for time.

As suggested by Knapp et al. (1992), as well as Glaser et al. (1991), job analysis can and should be a useful tool for delineating cognitive attributes needed for jobs and, in turn, for basic research on abilities. Recognizing this, Kyllonen (1985) suggested linking the specific content of cognitive tests to the cognitive requirements of jobs.

Newly evolving cognitive job analysis procedures, designed to delineate experts' mental models of a problem, also may prove useful for linking the CAM constructs to work behaviors. These procedures generally involve interviewing individuals who are experts in a particular area to map out decision points in a pre-selected job task and to identify segments of a task that are difficult for novice performers, but not for the experts (e.g., Eggemeier, Fisk, Robbins, Lawless, & Spaeth, 1988; Glaser et al., 1991). The primary result is a model of the cognitive processes involved in the accomplishment of the task and a description of differences between processing skills of experts and novices.

Although cognitive analyses were originally designed to delineate cognitive processes and to aid in the development of expert systems, the type of information these methods yield is potentially useful for a variety of other purposes. It can be used to reorient training programs to address specific segments of a task that are problematic for novices. With regard to the development of predictor and/or criterion measures, cognitive processing information can be used to build realistic task simulations and to develop protocols for scoring task performance. Even so, most of these methods are new, unvalidated, and labor-intensive. Further research is needed before they will be broadly applicable job analysis tools.

Dynamaticity and Automaticity of Performance on Cognitive Tasks

The predictive validity of cognitive ability measures changes over time (cf. Murphy, 1989). Several examples of this phenomenon have appeared in the literature. For example, Fleishman and his associates (e.g., Fleishman & Hempel, 1955) demonstrated that cognitive and psychomotor abilities show a changing pattern of correlations with task performance as a function of practice on that task. Ackerman (1986; 1987) suggested that the changing relationships are dependent upon the nature of the task. Specifically, tasks that entail a consistent series of information-processing steps may be highly dependent on cognitive ability only as the task is first learned; as practice continues, information processing for the task becomes automatic. Tasks that require a varying series of steps, however, continue to require controlled processing, and thus cognitive ability, even with practice. Furthermore, some research has shown that while cognitive ability may predict early performance on a job, personality characteristics may become more predictive as time passes (Helmreich et al., 1986).

Understanding the dynamaticity of performance on cognitive tasks affects decisions about the measurement of attributes and the amount of practice needed for accurate measurement of the intended construct. For example, Embretson (1987) showed that post-training spatial test scores were more internally consistent and more predictive of the criterion than pre-training scores. Continued research in this arena and interpretation of the research into applied practices may facilitate more accurate selection and classification measurement and decision-making.

Civilian Sector Preparations for Change

In response to societal concerns about education and preparation of youth to enter a competitive marketplace, the Department of Labor (DOL) established the Secretary's Commission on Achieving Necessary Skills (SCANS; Department of Labor, 1991, 1992). The SCANS Commission was composed of executives of major corporations (e.g., Motorola, Inc., General Electric Company). Its charter was to identify the skills, or competencies, high school graduates need in order to be prepared for a job upon completion of high school. DOL, with assistance from contractors, conducted a literature review and job analysis interviews with incumbents in a variety of entry level occupations to define job competencies and their relevance to specific jobs.

SCANS identified two major types of "Workplace Know-How:" Foundation skills and functional skills. Foundation skills are the basic academic and behavioral characteristics needed for the development of functional skills. Foundation skills include: (1) basic skills such as reading writing arithmetic, speaking, and listening, (2) thinking skills defined as the ability to learn, to reason, to think creatively, to make decisions and to solve problems, and (3) personal qualities such as individual responsibility, self-esteem, self-management, sociability, and integrity. Functional skills, or workplace competencies, are skills that build on the foundation skills. There are five functional skills: (1) skill in using resources (i.e., allocating time, money, materials, space, and staff), (2) interpersonal skills (e.g., working on teams, teaching others, serving customers), (3) information skills (e.g., acquiring and evaluating data, organizing and maintaining files), (4) systems understanding (e.g., understanding how social, organizational, and technological systems work), and (5) ability to use technology (e.g, to select equipment and tools for tasks, to maintain and troubleshoot equipment). Each skill is defined more specifically in terms of the tasks relevant to occupations that were included in the job analysis.

The SCANS Commission advocates widespread focus on the competencies. Specifically SCANS (DOL, 1992) recommended:

1. The nation's school systems should make the SCANS foundation skills and workplace competencies explicit objectives of instruction at all levels.

2. Assessment systems should provide students and workers with a resume documenting attainment of the SCANS know-how.

3. All employers, public and private should incorporate the SCANS know-how into all their human resource development efforts.

4. The Federal Government should continue to bridge the gap between school and the high-performance workplace, by advancing the SCANS agenda (p. xv).

In short, DOL and SCANS advocate the standardized measurement (for high school age youth) of certain basic workplace competencies. DOL awarded a contract in the summer of 1992 to develop measures of the these constructs, and in a separate effort education specialists are developing hands-on measures of SCANS competencies. The plan is to eventually give employers (including DoD) access to information about the student's current standing on the competencies. It is, therefore, likely that the SCANS competencies will receive increasing attention in the future and that the Services should watch, if not take an active role, in further development of SCANS assessments.

Review of Individual Differences Variables

We have mentioned a variety of individual differences constructs and measures throughout this report. They are listed in Table 31. Reliable measures exist for many of the constructs, particularly the cognitive variables. With regard to cognitive variables, future research should focus on (1) understanding the effects of practice and developing a mechanism for dealing with practice effects, (2) identifying novel cognitive constructs that, through improved technology, can now be measured, and (3) finding ways to minimize adverse impact and predictive bias. The bulk of previous research has been in the cognitive arena. Wherever possible, research should build on that knowledge base and avoid duplication.

Some measures of non-cognitive variables have shown promise for use in selection and classification; however, complications preclude their immediate use. For example, personality measures are good candidates for obtaining incremental validity over the ASVAB, but concerns about fakability and coachability block their implementation. Psychomotor tests may also enhance accuracy of prediction of job performance, but the Services will need to determine how to deal with the practice effects on these measures before they become operational. Finding ways to overcome the obstacles to implementing predictors that are very likely to be useful should be a primary research objective for the Services.

Measurement of other constructs such as adaptability is much more tenuous. Some measures that are already available (e.g., the ABLE) may provide a useful starting point for the measurement of new constructs. Yet, considerable work needs to be done to define definitions of these constructs before the development of good measures can proceed.

Individual Differences Measurement Research Objectives

The primary outcome of the Roadmap project will be a selection and classification research agenda. Task 1 yielded a set of research objectives and information about military selection and classification experts' perceptions of the importance and urgency of such objectives. Information gathered during Task 2 has suggested expansion of the Task 1 objectives with regard to individual differences measurement. The Task 1 objectives and their modifications are discussed below.

Determine which existing (but not implemented) predictors are most useful for classification purposes (Objective 7).

The ASVAB is a highly useful general purpose predictor. ASVAB subtests, composites, and the ASVAB general factor are valid predictors of job and training performance. The ASVAB predicts training success in a host of schools, for a variety of jobs, and in all the Services. Job performance validity information is limited but what is available indicates that the ASVAB predicts performance of the technical aspects of jobs (e.g., hands-on tasks).

Table 31

Individual Differences Attributes and Constructs and Selected Military Measures

Broad Attributes		Related Constructs	Selected Measures Developed by the Services
Cognitive			
G_c -	Knowledge or Crystallized Intelligence	Knowledge of general information Word knowledge	ASVAB [GS, WK, AS, MC, EI] OSB, AFOQT
G_f -	Broad Reasoning or Fluid Intelligence	Inductive reasoning Conjunctive reasoning Deductive reasoning	AFOQT ECAT Mental Counters ECAT Sequential Memory ECAT Figural Reasoning
G_v -	Broad Visual Intelligence	Spatial visualization Spatial orientation	BAT, AFOQT, OSB ECAT Assembling Objects ECAT Orientation Test ECAT Integrating Details
SAR -	Short Term Acquisition and Retrieval	Recency memory Word span	BAT
TSR -	Long Term Storage and Retrieval	Associational fluency Expressional fluency Ideational fluency	
G_s -	Broad Speediness	Visual scanning Visual matching	ASVAB [CS, NO] BAT, AFOQT ECAT Target Identification
G_a -	Auditory Intelligence	Discrimination among sound patterns Auditory cognition of relations	DLAB, ARC, Superdit
G_q -	Quantitative Thinking	Computational fluency Numerical computation	ASVAB [AR, MK] OSB, AFOQT
Eng -	English Adeptness	Word parsing Phonetic decoding	
Psychomotor			
Dexterity		Finger dexterity Manual dexterity	
Basic Movement Speed and Accuracy		Reaction time Control precision Speed of arm movement	
Perceptual-Motor Movement Control		Multi-limb coordination Rate control	BAT ECAT Tracking 1 ECAT Tracking 2

(Continued)

Table 31

Individual Differences Attributes and Constructs and Selected Military Measures (Continued)

Broad Attributes	Related Constructs	Selected Measures Developed by the Services
Physical		
Muscular Strength	Muscular tension Muscular power Muscular endurance	Air Force Strength Factor
Cardiovascular Endurance	Cardiovascular endurance	
Movement Quality	Flexibility Balance Coordination	
Personality		
Extraversion	Sociable, Gregarious Ambitious, Achievement-Oriented	OSB, ABLE, AAPP
Emotional Stability	Emotional, Anxious, Depressed	
Agreeableness	Good-natured, Cooperative	
Conscientiousness	Dependable, Responsible	
Intellectance	Curious, Broad-minded	
Interest		
Realistic	Practical, likes hand-on work	BAT, VOICE, AVOICE
Investigative	Curious, likes academic endeavors	
Artistic	Creative, likes self-expression	
Social	Friendly, likes people	
Enterprising	Ambitious, likes managing & directing	
Conventional	Concrete, likes exactness in work	
Biographical Information		
?	?	ASAP, EBIS, LEAP

Table 31

Individual Differences Attributes and Constructs and Selected Military Measures (Continued)

Broad Attributes	Related Constructs	Selected Measures Developed by the Services
Other Constructs		
Adaptability	?	Personality (?)
Motivational Predisposition	Achievement orientation Action control	Personality measures (?) ABLE (?)
Self-Efficacy Beliefs	Self-esteem Locus of control	Personality measures (?)
Social Intelligence	?	Assessment centers Low-fidelity simulations
Information Processing Constructs		CAM

Source: Cognitive (Horn, 1989); Psychomotor (Fleishman, 1967; Imhoff & Levine, 1981; McHenry, 1987), Physical (Hogan, 1991a); Personality (Barrick & Mount, 1991; Digman, 1990; Tett, Jackson, & Rothstein, 1991); Interests (Holland, 1983).

Efforts to improve the ASVAB need to focus on two major areas: (1) broadening its coverage of cognitive constructs and (2) reducing its adverse impact. Assuming that broadening the coverage of cognitive constructs measured by the ASVAB is a worthwhile goal, future supplements should focus on G_f, G_v, SAR, TSR, and perhaps G_a constructs. Within the context of Horn's 1989 framework, G_c, G_s, and G_q are covered by the ASVAB. G_v, *Broad Visualization*, G_f, *Fluid Intelligence*, SAR, *Short Term Acquisition and Retrieval*, TSR, *Long Term Storage and Retrieval*, and G_a, *Auditory Intelligence* are not. Also, comparisons of the factor structure of the

ASVAB with other published tests have provided empirical evidence that the ASVAB lacks a G_v measure (McBride, 1991; Wise & McDaniel, 1991). Sex and race differences in ASVAB scores are not trivial, particularly on the technical subtests. AS, MC, and EI yield the largest sex and race differences. Sex differences range from .80 SD for MC to 1.18 SD for AS. Black-White differences are greater than 1.25 SD for each test. When the three tests are unit weighted to form a "technical" score, the sex difference is 1.06 SD and the Black-White difference is 1.45 SD (Peterson, Russell et al., 1990).

The Services recognized these deficiencies in the ASVAB when preparing the ECAT. Chapter III provides a summary of research to date on the ECAT measures; additional data are currently being collected in a Joint Service research project. So far, several ECAT measures look like good candidates for inclusion in the ASVAB. With regard to G_v, the available data suggest that ECAT Assembling Objects is a test the Services will want to examine closely when supplementing the ASVAB. It has yielded small sex differences (relative to other spatial measures) in three large samples. It has been a useful predictor in studies conducted by the Marine Corps as well as the Army, although its incremental validity over the ASVAB is small and there is debate among researchers about the worth of small increments in validity.

In sum, with the Joint Service ECAT project, the Services are well on the way to identifying changes in new versions of the ASVAB. Short-term research projects do, however, continue to be necessary to identify the impact of removing specific subtests from the ASVAB and inserting new ones. These efforts are also underway in each of the Services.

Develop and evaluate measures of new predictors likely to be useful for classification purposes (Objective 8).

Cognitive Predictors

Several new predictors from the APT project, Project A, and Navy projects still hold promise for future ASVABs, although they are not in the current ECAT battery. Research using already developed cognitive measures should be encouraged wherever possible. Doing so would not only reduce costs associated with test development but also enable us to build a richer base of knowledge about tests. Also basic research on cognitive abilities is needed to identify abilities, enhance measurement of abilities, learn more about how abilities change over time or with practice, and to link information processing and traditional abilities domains.

We add two objectives:

8a. Include selected, already developed cognitive predictors in validation studies, across Services--to identify candidates for inclusion in future ASVABs.

8b. Continue to sponsor basic cognitive abilities measurement research.

Psychomotor Predictors

Addition of ECAT Tracking tests to the ASVAB would represent measurement of a new domain, and there is reason to expect these psychomotor tests would supplement the validity of the ASVAB. However, both tests are probably not necessary. ECAT Tracking 1 and 2 have virtually identical items and are highly correlated with each other (Peterson, Russell et al., 1990). Also, before implementing the psychomotor tests, the Services will need to decide how to deal with the large practice effects associated with them. Perhaps testing practice stations could be set up in the MEPS or in recruiting stations where applicants would be encouraged to try out practice items on tests. Another alternative may be to include a number of practice items on the tests.

Sex differences on the ECAT tracking measures are large. As long as these tests are used for selection and classification for combat jobs and combat jobs remain off-limits for women, this is a moot point. If, however, combat exclusion policies and laws are removed in the future, a number of issues arise. First, perhaps it will be more important to use psychomotor measures to make classification decisions because a wider range of individuals may be considered for combat jobs. Second, because the sex differences are so large, it will be necessary to show that psychomotor tests, if used, are based on job requirements identified through job analyses. Otherwise, it could be alleged that the Services adopted such tests as a surrogate for combat exclusion policies/laws, since psychomotor measures could exclude women from these jobs.

We add two objectives:

8c. If psychomotor tests are to be used, develop a mechanism for dealing with practice effects.

8d. If psychomotor tests are to be used, establish a job analytic mechanism for demonstrating the job relatedness of psychomotor abilities.

Physical Abilities Predictors

It is reasonable to expect that physical abilities measures would supplement the ASVAB for the prediction of performance in physically demanding jobs. Also, taxonomies of physical abilities are now available and can facilitate generalizability of validation results from civilian jobs to the domain of military jobs, making research less costly and more efficient. Therefore, physical abilities predictors are good candidates for inclusion in future testing efforts.

The issues involved in implementing physical abilities and psychomotor tests are similar. Specialized job analysis information would be needed to determine the physical and psychomotor requirements of the jobs. Both types of tests will yield some, if not a great deal of, adverse impact. In the same vein, the issues of if, how, and where to appropriately set cut-off scores for the tests utilized would need to be addressed.[1] Another consideration would be the cost of acquiring special equipment to conduct physical abilities and psychomotor testing. For physical abilities testing, test administrators would also have to be hired and/or trained to validly and reliably measure individuals. In addition, there may well be a space problem to deal with should such testing be implemented at MEPS. Rooms for testing and space for equipment storage would be needed. Perhaps physical abilities testing should occur during or at the end of Basic Military Training instead of at the MEPS.

Despite these concerns, assessing the capacity of military applicants to handle physical tasks would appear to be fundamental to selecting individuals to perform in certain fields. Hogan (in press) does point out that as physical abilities tests have been found to be valid predictors of job performance and are statistically independent, they provide incremental validity to the prediction of the criterion space. The capability, then, exists to further calculate and thereby improve upon the performance of those entering and working in positions that require physical effort. We add these objectives:

8e. If physical abilities tests are to be used, establish a job analytic mechanism for demonstrating the job relatedness of physical abilities.

8f. Examine and estimate the logistical requirements associated with physical abilities and psychomotor test administration.

8g. Identify physical abilities measures that are likely to be good predictors with minimal adverse impact.

Personality Predictors

Personality predictors are promising candidates as supplements to the cognitive measures traditionally used by the Services for several reasons. First, recent advances in the area of personality structure have led to new agreement on basic factors around which traits may be organized. These factors have helped researchers to be specific about the nature of the criterion relationships that may be expected for personality variables. Second, meta-analyses have shown personality variables to have consistent useful relationships with a variety of criteria. Research indicates that personality measures are good candidates as supplemental measures to existing and experimental cognitive tests, especially for the prediction of "will-do" criteria such as Effort and Leadership, Personal Discipline, and Physical Fitness and Military Bearing, as well as training attrition. Third, personality measures appear to show fewer differences among races than do cognitive measures, and the differences that have been shown tend to favor minority respondents. Fourth, the Services have already developed some personality measures that appear to work well.

[1]When we interviewed selection and classification experts in earlier phases of this project, experts voiced some concern that cut scores on physical tests (as well as other physical restrictions on height, for example) lack job analytic support.

The primary issue regarding actual implementation of personality measures is the potential for fakability and coachability. Faking is possible on these measures, but it is possible to detect faking in many cases. Further research is necessary to determine how to best reduce socially desirable responding and purposeful faking and how to deal with suspect response profiles. The conduct of a comprehensive review of the faking and social desirability literature would be an important step in organizing our knowledge in this important area. The literature reviewed here suggests that the possibility that faking may occur does not impoverish completely the utility of non-cognitive measures. It is also possible that there are ways to prevent faking that have not been explored (e.g., giving periodic, tactful feedback on a computer-administered form). An objective we give very high priority is:

8h. Investigate fakability/coachability of personality measures, particularly how to prevent fakability/coachability and how to determine the impact of faking when it does occur.

Interest Measures

The Air Force and the Navy currently use individual information about job preferences in their classification process (Russell et al., 1992). It is possible that interest inventories would more accurately identify interests than the current methods where recruits rate occupational categories. Validation findings indicate that interest measures predict later occupational membership and job satisfaction, however interests do not appear to add much in the prediction of job performance over that accounted for by cognitive and personality predictors. These findings suggest that interest measures may by more useful for classifying people into jobs rather than as selection measures.

There are two major obstacles to the implementation of interest measures: (1) adverse impact and (2) coachability. Implications for research are:

8i. Analyze adverse impact issues regarding interest measures.

8j. Identify ways to prevent faking/coaching on interest inventories.

Biodata Predictors

Biodata are effective and valid predictors of a number of important criteria. Research has indicated that biodata validities can be made generalizable and stable (Rothstein et al., 1990), thus these measures are worthy of continued consideration as supplements to cognitive predictors of military performance. There is also evidence that biodata may have incremental validity over cognitive measures, especially when predicting non-performance criteria such as attrition (e.g., Trent, In Press). Biodata do not yield large differences among the races and evidence of differential validity is slight. Although biodata measures are possible to fake, research indicates that faking may not be prevalent. Finally, one additional strength of biodata is that some measures (e.g., the EBIS) have predicted attrition which has traditionally been predicted by educational attainment criteria. Educational credentials have come under fire lately (cf. Laurence, in press) because they restrict entrance to the military for identifiable groups of individuals (e.g., GED recipients). Biodata instruments provide a compensatory measure such that no one particular characteristic will be likely to exclude an individual. Thus, biodata may face less implementation resistance than other predictors of military adjustment.

We think that biodata measures are probably one of the best candidates for improving enlisted selection and classification. If biodata measures were made operational, it would be critical to track their performance over time and maintain the instruments accordingly. This leads to another objective:

8k. Continue research to determine the utility of biodata predictors.

Multi-Domain Research

In the first task of this project (Russell et al., 1992), military personnel specialists were asked about the future needs of the Services. Three major themes regarding future changes emerged from our data. First, it was noted that the Services will move from highly specialized jobs to jobs with more generalized responsibilities. Second, the mission of the armed forces is changing from large scale operations to smaller scale intervention, and with this change comes an increased emphasis on smaller teams that may be deployed quickly. Third, technological advancement will continue to change the nature of military work.

Such trends will likely lead to increased job complexity, greater social interdependence, and cognitive ability requirements that are beyond our current measurement capability. This suggests that these jobs may require higher cognitive ability, but also that selection and classification researchers will need to investigate the predictive utility of the interactions between cognitive and dispositional characteristics, basic differences in motivational predisposition and social intelligence, and the measurement of basic cognitive processes. Another objective is:

8l. Conduct basic multi-domain predictor research.

Identify and/or develop classification measures that minimize adverse impact and/or predictive bias (Objective 17).

Although the ASVAB does not typically result in predictive bias, there is adverse impact in test scores. There are three ways to reduce the impact. One method is to develop/identify cognitive tests that yield minimal adverse impact, with little or no reduction in validity. There is evidence that some cognitive tests yield differences that are smaller than those from other tests of the same broad construct (Linn & Petersen, 1984). Meta-analyses of sex and race differences in abilities may help shed light on aspects of tests that are related to magnified differences. Another way to reduce overall adverse impact is to use non-cognitive, particularly personality, measures that traditionally yield either no difference or differences favoring minority groups. Since such measures also add incremental validity over the ASVAB in the prediction of job performance criteria, they are very attractive candidates for future selection and classification testing. Finally, adverse impact results not only from the nature of the ASVAB itself but also from policy. The Air Force requires applicants to meet minimum standards on MAGE, which yields a sex difference, while the other Services use the AFQT which results in a smaller sex differences. The third way to reduce adverse impact, against women anyway, is to recommend policy changes.

We suggest the following refinements to Objective 17:

17a. Identify/develop cognitive measures that minimize adverse impact without loss of validity.

17b. Examine the effect of coupling cognitive and non-cognitive measures on overall adverse impact (and predictive validity).

17c. Identify policies that reinforce adverse impact and recommend changes.

REFERENCES

Abrahams, N. M., Neumann, I., & Githens, W. H. (1971). Faking vocational interests: Simulated versus real-life motivation. Personnel Psychology, 24, 5-12.

Ackerman, P. L. (1986). Skill acquisition, individual differences and human abilities. Proceedings of the 30th annual meeting of the Human Factors Society, 270-274.

Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. Psychological Bulletin, 102, 3-27.

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. Journal of Experimental Psychology: General, 117, 288-318.

Alderton, D. L. (1989a). Development and evaluation of Integrating Details: A complex spatial problem solving test (NPRDC TR 89-6). San Diego, CA: Navy Personnel Research and Development Center.

Alderton, D. L. (1989b). Integrating Details: Proposal for joint service validation (A proposal to the Technical Advisory Selection Panel). San Diego, CA: Navy Personnel Research and Development Center.

Alderton, D. L. (1990, April). Revisiting a cognitive framework for test design: Application for a computerized Perceptual Speed Test. Paper presented at the meeting of the American Educational Research Association, Boston, MA.

Alderton, D. L. (1991, August). Improved measurement of perceptual speed ability through computer administration. Paper presented at the meeting of the American Psychological Association, San Francisco, CA.

Alley, W. E., & Matthews, M. D. (1982). The Vocational Interest Career Examination: A description of the instrument and possible applications. The Journal of Psychology, 112, 169-193.

Alley, W. E., Treat, B. R., & Black, D. E. (1988). Classification of Air Force jobs into aptitude clusters (AFHRL-TR-88-14). Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory.

Alley, W. E., Wilbourn, J. M., & Berberich, G. L. (1976). Relationships between performance on the Vocational Interest Career Examination and reported job satisfaction (AFHRL-TR-76-89). Lackland Air Force Base, TX: U.S. Air Force Human Resources Laboratory.

Allport, G. W., & Odbert, H. S. (1936). Trait names: A psycholexical study. Psychological Monographs, 47 (1, Whole No. 211).

American Institutes of Research (1992). A guide to test validation. Washington, DC: ERIC Clearinghouse on Tests, Measurement, and Evaluation.

Anastasi, A. (1958). Differential psychology: Individual and group differences in behavior (3rd ed.). New York: MacMillan.

Anastasi, A. (1983). Evolving trait concepts. American Psychologist, 38, 175-184.

Ansbacher, H. L. (1941). German military psychology. Psychological Bulletin, 38, 370-379.

Appel, V. H., Quintana, C. M., Cole, R. W., Shermis, M. D., Grubb, P. D., Watson, T. W., & Headley-Goode, A. (1992). The Leadership Effectiveness Assessment Profile (LEAP): Officer instrument field testing and refinement (AL-TR-1992-xx). Brooks Air Force Base, TX: Human Resources Directorate, Armstrong Laboratory.

Arth, T. O. (1986). Validation of the AFOQT for non-rated officers (AFHRL-TP-85-50). Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory.

Arth, T. O., & Skinner, J. (1986, November). Aptitude selections for Air Force officer non-aircrew jobs. Paper presented at the meeting of the Military Testing Association, New London, Connecticut.

Baker, H. G. (1985). Antecareer crisis: Military recruiting and the youthful job applicant. Armed Forces & Society, 11, 565-580.

Bandura, A. (1982). Self-efficacy mechanism in human agency. American Psychologist, 37, 122-147.

Bandura, A. (1986). Social foundations of thought and action: A social-cognition theory. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Bandura, A. (in press). Regulative function of perceived self-efficacy. Proceedings of the ARI Conference on Selection and Classification. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Bandura, A., & Cervone, D. (1983). Self-evaluative and self-efficacy mechanisms governing the motivational effects of goal systems. Journal of Personality and Social Psychology, 45, 1017-1028.

Barge, B. N., & Hough, L. M. (1986). Utility of interest assessment for predicting job performance. In L. M. Hough, J. D. Kamp, & B. N. Barge (Eds.), Utility of temperament, biodata, and interest assessment for predicting job performance: A review and integration of the literature. Minneapolis, MN: Personnel Decisions Research Institute.

Barnes, M. L., & Sternberg, R. J. (1989). Social intelligence and decoding of nonverbal cues. Intelligence, 13, 263-287.

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. Personnel Psychology, 44, 1-26.

Becker, T. E., & Colquitt, A. L. (1992). Potential versus actual faking of a biodata form: An analysis along several dimensions of item type. Personnel Psychology, 45, 389-406.

Bem, D. J., & Allen, A. (1974). On predicting some of the people some of the time: The search for cross-situational consistencies in behavior. Psychological Review, 81, 506-520.

Bennett, G. K., & Cruikshank, R. M. (1974). Sex differences in the understanding of mechanical problems. Journal of Applied Psychology, 26, 121-127.

Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1974). Manual for the differential aptitude tests: Forms S and T (5th ed.). New York: The Psychological Corporation.

Berger, F. R., & Berger, R. M. (1977). Vocational Interest Career Examination: Norming and standardization on a nation-wide high school sample (AFHRL-TR-77-69). Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory.

Bloxom, B. (1992, March). Armed Services Vocational Aptitude Battery and its use in selection and classification. Briefing presented to the Armed Services Vocational Aptitude Battery Revision Workshop, Washington, DC.

Bordelon, V. P., & Kantor, J. E. (1986). Utilization of psychomotor screening for USAF pilot candidates: Independent and integrated selection methodologies (AFHRL-TR-86-4). Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory.

Bosshardt, M. J. (1987). Utility of psychomotor tests for prediction of Navy enlisted performance (Institute Report No. 141). Minneapolis, MN: Personnel Decisions Research Institute.

Boyd, V. S. (1978). Neutralizing sexist titles in Holland's Self-Directed Search: What difference does it make? In C.K. Tittle & D.G. Zytowski (Eds.), Sex-fair interest measurement: Research and implications. Washington, DC: National Institute of Education.

Brinkmann, E. H. (1966). Programmed instruction as a technique for improving spatial visualization. Journal of Applied Psychology, 50, 179-184.

Brockner, J. (1988). Self-esteem at work: Research, theory, and practice. Lexington, MA: Lexington Books.

Brown, D. C. (1989). Officer aptitude selection measures. In M. F. Wiskoff & G. M. Rampton (Eds.), Military personnel measurement (pp. 107-167). New York: Praeger.

Brown, W., & Stephenson, W. A. (1933). A test of the theory of two factors. British Journal of Psychology, 23, 352-370.

Burnett, S. A. (1986). Sex differences in spatial ability: Are they trivial? American Psychologist, 41, 1012-1014.

Busciglio, H. H., & Palmer, D. R. (1992, August). An empirical assessment of coaching and practice effects on three Army tests of spatial aptitude. Paper presented at the annual meeting of the American Psychological Association, Washington, DC.

Busciglio, H. H., Palmer, D. R., & King, I. H. (1991). What is "social intelligence" and how do we measure it? Proceedings of the 33rd Annual Conference of the Military Testing Association, 640-645.

Campbell, C. H., Ford, P., Rumsey, M. G., Pulakos, E. D., Borman, W. C., Felker, D. B., DeVera, M. V., & Riegelhaupt, B. J. (1990). Developmental multiple job performance measures in a representative sample of jobs. Personnel Psychology, 43, 277-300.

Campbell, D. P. (1971). Handbook for the Strong Vocational Interest Blank. Stanford, CA: Stanford University Press.

Campbell, D. P., & Hansen, J. C. (1981). Manual for the SVIB-SCII. Stanford, CA: Stanford University Press.

Campbell, J. P. (Ed.). (1987). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report, 1985 Fiscal Year (ARI Technical Report 746). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Campbell, J. P. (1986). Validation analysis for new predictors (RS-WP-86-09). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Campbell, J. P. (Ed.). (1989). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report, 1988 Fiscal Year. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Campbell, J. P. (Ed.). (1990). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report, 1985 fiscal year (ARI Technical Report 746). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Campbell, J. P. (Ed.). (1992). Building a joint-service classification research roadmap: Methodological issues in selection and classification. Alexandria, VA: Human Resources Research Organization.

Campbell, J. P. (in press). Alternative models of job performance and their implications for selection and classification. Proceedings of the ARI Conference on Selection and Classification. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Campbell, J. P., & Johnson, J. W. (in press). Results of the second-tour validation (CVII). In J. P. Campbell & L. M. Zook (Eds.), Building and retaining the career force: New procedures for accessing and assigning Army enlisted personnel: Annual report, 1991 fiscal year (ARI Technical Report 952). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Cantor, J., Engle, R. W., & Hamilton, G. (1991). Short-term memory, working memory, and verbal abilities: How do they relate? Intelligence, 15, 229-246.

Carey, N. B. (1992, August). New predictors of mechanics' job performance: Marine Corps findings. Paper presented at the annual meeting of the American Psychological Association, Washington, DC.

Carretta, T. R. (1987a). Basic Attributes Test (BAT) System: Development of an automated test battery for pilot selection (AFHRL-TR-87-9). Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory.

Carretta, T. R. (1987b). Basic Attributes Test (BAT) System: A preliminary evaluation (AFHRL-TP-87-20). Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory.

Carretta, T. R. (1987c). Field dependence - independence and its relationship to flight training performance (AFHRL-TP-87-36). Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory.

Carretta, T. R. (1990). Cross-validation of experimental USAF pilot training performance models (AFHRL-TR-89-68). Brooks Air Force Base, TX: Manpower and Personnel Division, U.S. Air Force Human Resources Laboratory.

Carretta, T. R. (1991). Short-term test-retest reliability of an experimental version of the Basic Attributes Test battery (AL-TP-1991-0001). Brooks Air Force Base, TX: Human Resources Directorate, Armstrong Laboratory.

Carretta, T. R. (1992). Predicting pilot training performance: Does the criterion make a difference? (AL-TP-1991-0055). Brooks Air Force Base, TX: Human Resources Directorate, Armstrong Laboratory.

Carroll, J. B. (1989). Factor analysis since Spearman: Where do we stand? What do we know? In R. Kanfer, P. L. Ackerman, & R. Cudek (Eds.), Learning and individual differences: Abilities, motivation, and methodology. Hillsdale, NJ: Erlbaum.

Carroll, J. B. (1991a). No demonstration that *g* is not unitary, but there's more to the story: Comment on Kranzler and Jensen. Intelligence, 15, 423-436.

Carroll, J. B. (1991b). Still no demonstration that *g* is not unitary: Further comment on Kranzler and Jensen. Intelligence, 15, 449-453.

Carter, G. W. (1991). A study of relationships between measures of individual differences and job satisfaction among U.S. Army enlisted personnel. Unpublished doctoral dissertation, University of Minnesota, Minneapolis.

Cascio, W. F. (1975). Accuracy of verifiable biographical information blank responses. Journal of Applied Psychology, 60, 767-769.

Cattell, R. B. (1971). Abilities: Their structure, growth and action. Boston: Houghton-Mifflin.

Cattell, R. B., Eber, H. W., & Tatsuoka, M. M. (1970). Handbook for the Sixteen Personality Factor Questionnaire (16PF). Champaign, IL: Institute for Personality and Ability Testing.

Chapin, F. S. (1967). The Social Insight Test. Palo Alto, CA: Consulting Psychologists Press.

Christal, R. E. (1991). Comparative validities of ASVAB and LAMP tests for Logic Gates Learning (AL-TP-1991-0031). Brooks Air Force Base, TX: Human Resources Directorate, Armstrong Laboratory.

Cleary, T. A. (1968). Test bias: Prediction of grades of Negro and White students in integrated colleges. Journal of Educational Measurement, 5, 115-124.

Comrey, A. L. (1970). EITS manual for the Comrey Personality Scales. San Diego: Educational and Industrial Testing Service.

Cortina, J. M., Doherty, M. L., Schmitt, N., Kaufman, G., & Smith, R. G. (1992). The "big five" personality factors in the IPI and MMPI: Predictors of police performance. Personnel Psychology, 45, 119-140.

Cowan, D. K., Barrett, L. E., & Wegner, T. G., Capt. (1989). Air Force Reserve Officer Training Corps selection system validation (AFHRL-TR-88-54). Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory.

Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. (1972). An MMPI handbook, Volume I: Clinical interpretation. Minneapolis: University of Minnesota Press.

Dawis, R. V. (1991). Vocational interests, values, and preferences. In M. D. Dunnette and L. M. Hough (Eds.), Handbook of Industrial and Organizational Psychology (2nd Ed.) (Vol. 2, pp. 833-872). Palo Alto, CA: Consulting Psychologists Press.

Day, D. V., & Silverman, S. B. (1989). Personality and job performance: Evidence of incremental validity. Personnel Psychology, 42, 25-36.

Department of the Army. (1987). Alternate flight aptitude selection test (AFAST). Fort Rucker, AL: U.S. Army Research Institute for the Behavioral and Social Sciences.

Department of Defense (1982). Profile of American Youth: 1980 Nationwide Administration of the Armed Services Vocational Aptitude Battery. Washington, DC: Office of the Assistant Secretary of Defense.

Department of Defense (1984a). Armed Services Vocational Aptitude Battery (ASVAB) Test Manual, Forms 8 - 14 (DoD 1304.12AA). North Chicago, Illinois: United States Military Entrance Processing Command.

Department of Defense (1984b). Armed Services Vocational Aptitude Battery (ASVAB) Information Pamphlet, Forms 8 - 14 (DoD 1304.12Z). Washington, DC: U.S. Government Printing Office.

Department of Labor (1970). General Aptitude Test Battery Manual, Section III. Washington, DC: Employment and Training Administration, U.S. Employment Service.

Department of Labor (1991). What work requires of schools: A SCANS report for America 2000. Washington, D.C.: Author.

Department of Labor (1992). Learning a living: A blueprint for high performance: A SCANS report of America 2000. Washington, D.C.: Author.

Diamond, E. E. (Ed.). (1975). Issues of sex bias and sex fairness in career interest measurement. Washington, DC: Department of Health, Education, and Welfare: National Institute of Education.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. Annual Review of Psychology, 41, 417-440.

Dunnette, M. D., Corpe, V. A., & Toquam, J. L. (1987). Cognitive paper-and-pencil measures: Field test. In N. G. Peterson (Ed.), Development and field test of the trial battery for Project A (ARI TR-739). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Dunnette, M. D., Kirchner, W. K., Erickson, J., & Banas, P. (1960). Predicting turnover among female office workers. Personnel Administration, 23, 45-50.

Dunnette, M. D., McCartney, J., Carlson, H. C., & Kirchner, W. K. (1962). A study of faking behavior on a forced-choice self-description check-list. Personnel Psychology, 15, 13-24.

Eastman, R.F., & McMullen, R.L. (1978). Item analysis and revision of the flight aptitude selection tests (ARI Field Unit Research Memorandum 78-4). Fort Rucker, AL: U.S. Army Research Institute for the Behavioral and Social Sciences.

Eberhardt, B. J., & Muchinsky, P. M. (1982). Biodata determinants of vocational typology: An integration of two paradigms. Journal of Applied Psychology, 67, 714-727.

Edwards, A. L. (1959). Edwards Personal Preference Schedule Manual (Rev.). New York: Psychological Corporation.

Eggemeier, F. T., Fisk, A. D., Robbins, D., Lawless, M. T., & Spaeth, R. L. (1988). High-performance skills task analysis methodology: An automatic human information processing theory approach (AFHRL-TP-88-32). Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory.

Eitelberg, M. J., Laurence, J. H., Waters, B. K., & Perelman, L. S. (1984). Screening for service: Aptitude and education criteria for military entry. Washington, DC: Office of the Assistant Secretary of Defense.

Ekstrom, R. B., French, J. W., & Harman, H. H. (1979). Cognitive factors: Their identification and replication. Multivariate Behavioral Research Monographs, 79, 1-84.

Embretson, S. E. (1987). Improving the measurement of spatial aptitude by dynamic testing. Intelligence, 11, 333-358.

England, G. W. (1971). Development and use of weighted application blanks (Bulletin 55). Minneapolis, MN: Industrial Relations Center, University of Minnesota.

Eysenck, H. J., & Eysenck, S. B. G. (1969). Personality structure and measurement. San Diego, CA: Knapp.

Eysenck, H. J., & Eysenck, S. B. G. (1975). Manual for the Eysenck Personality Questionnaire. San Diego, CA: Educational and Industrial Testing Service.

Finegold, L., & Rogers, D. (1985). Relationship between Air Force Officer Qualifying Test scores and success in air weapons controller training (AFHRL-TR-85-13). Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory.

Fischl, M. A., Edwards, D. S., Claudy, J. G., & Rumsey, M. G. (1986). Development of Officer Selection Battery Forms 3 and 4 (ARI Technical Report 603). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Fleishman, E. A. (1954). Dimensional analysis of psychomotor abilities. Journal of Experimental Psychology, 8, 437-454.

Fleishman, E. A. (1955). Predicting code proficiency of radio telegraphers by means of aural tests. Journal of Applied Psychology, 39, 3, 150-155.

Fleishman, E. A. (1964). Structure and measurement of physical fitness. In J.A. Hogan (1991), Physical abilities. In M.D. Dunnette and L.M. Hough (Eds.), Handbook of industrial and organizational psychology (Vol. 2, pp. 751-831). Palo Alto, CA: Consulting Psychologists Press.

Fleishman, E. A. (1967). Performance assessment based on an empirically derived task taxonomy. Human Factors, 9, 349-366.

Fleishman, E. A. (1972). On the relation between abilities, learning, and human performance. American Psychologist, 27, 1017-1032.

Fleishman, E. A., & Ellison, G. D. (1962). A factor analysis of fine manipulative tests. Journal of Applied Psychology, 46, 96-105.

Fleishman, E. A., & Hempel, W. E. (1954a). Changes in factor structure of a complex psychomotor test as a function of practice. Psychometrika, 19, 239-252.

Fleishman, E. A., & Hempel, W. E. (1954b). A factor analysis of dexterity tests. Personnel Psychology, 7, 15-32.

Fleishman, E. A., & Hempel, W. E. (1955). The relation between abilities and improvement with practice in a visual discrimination reaction task. Journal of Experimental Psychology, 49, 301-312.

Fleishman, E. A., & Hempel, W. E. (1956). Factorial analysis of complex psychomotor performance and related skills. Journal of Applied Psychology, 40, 96-104.

Fleishman, E. A., & Mumford, M. D. (1988). Ability requirements scales. In S. Gael (Ed.), Job analysis handbook for business, industry, and government (Vol. 2, pp. 917-935). New York: Wiley.

Gaugler, B. B., Rosenthal, D. B., Thornton, G. C., & Bentson, C. (1987). Meta-analysis of assessment center validity. Journal of Applied Psychology, 72, 493-511.

General Accounting Office (1991, August). Women in the military: Air Force revises job availability but entry screening needs review (GAO/NSIAD-91-199). Washington, DC: Author.

Ghiselli, E. E. (1955). The measurement of occupational aptitude. Berkeley, CA: University of California Press.

Ghiselli, E. E. (1973). The validity of aptitude tests in personnel selection. Personnel Psychology, 26, 461-477.

Gibb, G. D. (1987). Development of a computer-based Naval selection battery [Summary]. Proceedings of the 29th Annual Conference of the Military Testing Association, 619-624.

Gibb, G. D. (1990). Initial validation of a computer-based secondary selection system for student Naval aviators. Military Psychology, 2, 205-219.

Glaser, R., Lesgold, A., & Gott, S. (1991). Implications of cognitive psychology for measuring job performance. In A.K. Wigdor & B.F. Green (Eds.), Performance assessment for the workplace: Volume II technical issues. Washington, D.C.: National Academy Press.

Goldberg, L. W. (1992, May). Basic research on personality structure: Implications of the emerging consensus for applications to selection and classification. Proceedings of the ARI Conference on Selection and Classification. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Goldstein, I. L. (1971). The application blank: How honest are the responses? Journal of Applied Psychology, 55, 491-492.

Gorden, F. A. (in press). Introductory comments. Proceedings of the ARI Conference on Selection and Classification. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Gordon, L. V. (1978). Gordon Personal Profile - Inventory manual. New York: Psychological Corporation.

Gough, H. G. (1975). Manual for the California Psychological Inventory. Palo Alto, CA: Consulting Psychologists Press.

Griffin, G.R., & McBride, D.K. (1986). Multitask performance: Predicting success in Naval aviation primary flight training. NAMRL-1316, Pensacola, FL: Naval Aerospace Medical Research Laboratory.

Guilford, J. P. (1967). The nature of human intelligence. New York: McGraw-Hill.

Guilford, J. P., & Lacey, J. I. (1947). Printed Classification Tests, A.A.F., aviation psychological progress research report, 5. Washington, DC: U.S. Government Printing Office.

Guilford, J. S., Zimmerman, W. S., & Guilford, J. P. (1976). The Guilford-Zimmerman Temperament Survey handbook. San Diego, CA: Eudcational and Industrial Testing Service.

Guion, R. M., & Gottier, R. F. (1965). Validity of personality measures in personnel selection. Personnel Psychology, 18, 135-164.

Hanson, M. A., & Borman, W. C. (1990). A situational judgment test of supervisory knowledge in the U.S. Army. Proceedings of the 32nd Annual Conference of the Military Testing Association, 268-273.

Hanson, M. A., Hallam, G. L., & Hough, L. M. (1989). Detection of response distortion in the Adaptability Screening Profile (ASP). Proceedings of the 31st Annual Conference of the Military Testing Association, 422-427.

Hanson, M. A., Paullin, C., & Borman, W. C. (1990). Development of an experimental Biodata/temperament inventory for NROTC selection. Proceedings of the 32nd Annual Conference of the Military Testing Association, 498-503.

Hartke, D. D., & Short, L. O. (1988). Validity of the academic aptitude composite of the Air Force Officer Qualifying Test (AFOQT) (AFHRL-TP-87-61). Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory.

Hedges, L. V. (1982). Fitting categorical models to effect sizes from a series of experiments. Journal of Educational Statistics, 7, 119-137.

Heist, P., & Yonge, G. (1968). Manual for the Omnibus Personality Inventory, Form F. New York: Psychological Corporation.

Helmreich, R.L. Foushee, C.H., Bensen, R., & Russini, W. (1985). Proceeding of the Third Aviation Psychology Symposium, Columbus, Ohio, The Ohio State University.

Helmreich, R. L., Sawin, L. L., & Carsrud, A. L. (1986) The honeymoon effect in job performance: Temporal increases in the predictive power of achievement motivation. Journal of Applied Psychology, 71, 185-188.

Hoffman, K. I., Guilford, J. P., Hoepfner, R., & Doherty, W. J. (1968). A factor analysis of the figural-cognitive and figural-evaluation abilities (Report No. 40). Los Angeles, CA: University of Southern California Psychological Laboratory.

Hogan, J. A. (1984, April). A model of physical performance for occupational tasks. Paper presented at the annual meeting of the American Psychological Association, Toronto, Canada.

Hogan, J. A. (1991a). Physical abilities. In M. D. Dunnette and L. M. Hough (Eds.), Handbook of industrial and organizational psychology (Vol 2., pp. 751-831). Palo Alto, CA: Consulting Psychologists Press.

Hogan, J. A. (1991b). Structure of physical performance in occupational tasks. Journal of Applied Psychology, 76, 495-507.

Hogan, J. A. (in press). Theoretical and applied developments in models of individual differences: Physical abilities. Proceedings of the ARI Conference on Selection and Classification. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Hogan, R. (1982). A socioanalytic theory of personality. In M. Page (Ed.), Nebraska symposium on motivation (Vol. 30, pp. 56-89). Lincoln: University of Nebraska Press.

Hogan, R. (1986). Manual for the Hogan Personality Inventory. Minneapolis, MN: National Computer Systems.

Holland, J. L. (1966). The psychology of vocational choice. Waltham, MA: Blaisdell.

Holland, J. L. (1973). Making vocational choices: A theory of careers. Englewood Cliffs, NJ: Prentice-Hall.

Holland, J. L. (1976). Vocational preferences. In M. D. Dunnette (Ed.), Handbook of industrial and organizational psychology (pp. 521-570). Chicago, IL: Rand McNally.

Holland, J. L. (1983). Vocational preferences. In M.D. Dunnette (Ed.), Handbook of industrial and organizational psychology (2nd ed.; pp. 521-570). New York: John Wiley and Sons.

Horn, J. L. (1989). Cognitive diversity: A framework of learning. In P. L. Ackerman, R. J. Sternberg, & R. Glaser (Eds.), Learning and individual differences (pp. 61-116). New York: Freeman.

Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities [Monograph]. Journal of Applied Psychology, 75, 581-595.

Hough, L. M., Kamp, J. D., & Barge, B. N. (Eds.). (1986). Utility of temperament, biodata, and interest assessment for predicting job performance: A review and integration of the literature. Minneapolis, MN: Personnel Decisions Research Institute.

Humphreys, L. G. (1979). The construct of general intelligence. Intelligence, 3, 105-120.

Humphreys, L. G. (1986). Commentary. Journal of Vocational Behavior, 29, 421-437.

Humphreys, L. G. (1992). Ability testing commentary: What both critics and users of ability tests need to know. Psychological Science, 3, 271-274.

Hunt, E., Pelligrino, J. W., Abate, R., Alderton, D. L., Farr, S. A., Frick, R. W., & McDonald, T. P. (1987). Computer-controlled testing of visual-spatial ability (NPRDC TR 87-31). San Diego, CA: Navy Personnel Research and Development Center.

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. Journal of Vocational Behavior, 29, 340-362.

Hunter, J. E. & Schmidt, F. L. (1990). Methods of meta-analysis. Newbury Park: Sage Publications.

Hunter, D., & Thompson, N. (1978). Pilot selection system development (AFHRL-TR-78-33). Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory.

Imhoff, D. L., & Levine, J. M. (1981). Perceptual-motor and cognitive performance task battery for pilot selection (AFHRL-TR-80-27). Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory.

Intano, G.P., Gabriel, P., Howse, W.R., & Lofaro, R.J. (1987). The selection of an experimental test battery for aviator cognitive, psychomotor abilities and personal traits. (ARI Interim Report). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Intano, G.P., & Howse, W.R. (1991). Predicting performance in Army aviation primary flight training. (ARI Report ARI-RN-92-06). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Intano, G.P., Howse, W.R., & Lofaro, R.J. (1991a). The selection of an experimental test battery for aviation cognitive, psychomotor abilities and personal traits. (ARI Report ARI-RN-

91-21). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Intano, G.P., Howse, W.R., & Lofaro, R.J. (1991b). Initial validation of the Army aviator classification process. (ARI Report ARI-RN-91-38). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Jackson, D. N. (1967). Personality Research Form manual. Goshen, NY: Research Psychologists Press.

Jackson, D. N. (1976). Jackson Personality Inventory manual. Goshen, NY: Research Psychologists Press.

Jensen, A. R. (1977). Cumulative deficit in IQ of Blacks in the rural south. Developmental Psychology, 13, 184-191.

Jensen, A. R. (1980). Bias in mental testing. New York: The Free Press.

Jensen, A. R. (1986). *g*: Artifact or reality? Journal of Vocational Behavior, 29, 301-331.

Jensen, A. R. (1992). Ability testing commentary: Vehicles of *g*. Psychological Science, 3, 275-278.

Kamp, J. D., & Hough, L. M. (1986). Utility of temperament for predicting job performance. In L. M. Hough, J. D. Kamp, & B. N. Barge (Eds.), Utility of temperament, biodata, and interest assessment for predicting job performance: A review and integration of the literature. Minneapolis, MN: Personnel Decisions Research Institute.

Kanfer, R. (1990). Motivation theory in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), Handbook of industrial and organizational psychology (Vol. 1, pp. 75-170). Palo Alto, CA: Consulting Psychologists Press.

Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition [Monograph]. Journal of Applied Psychology, 74, 657-690.

Kaplan, H. (1965). Prediction of success in Army aviation training (Technical Research Report 1142). Washington, DC: U.S. Army Personnel Research Office.

Kass, R. A., Mitchell, K. J., Grafton, F. C., & Wing, H. (1983). Factorial validity of the Armed Services Vocational Aptitude Battery (ASVAB), Forms 8, 9, and 10: 1981 Army applicant sample. Educational and Psychological Measurement, 43, 1077-1087.

Kettner, N. (1977). Armed Services Vocational Aptitude Battery (ASVAB Form 5): Comparison with GATB and DAT tests (Technical Research Report 77-1). Fort Sheridan, IL: Directorate of Testing, U.S. Military Enlistment Processing Command.

Kleinke, D. J. (1992, August). Edison Electric Institute's Employee Selection Testing Consortia: Lessons learned. Presented at the monthly meeting of the Personnel Testing Council of Metropolitan Washington, Arlington, VA.

Kluger, A. N., Reilly, R. R., & Russell, C. J. (1991). Faking biodata tests: Are option-keyed instruments more resistant? Journal of Applied Psychology, 76, 889-896.

Knapp, D. K., Russell, T. L., & Campbell, J. P. (1992). Building a joint-service classification research roadmap: Job analysis methodologies. Alexandria, VA: Human Resources Research Organization.

Korman, A. K. (1967). Relevance of personal need satisfaction for overall satisfaction as a function of self-esteem. Journal of Applied Psychology, 51, 533-538.

Kranzler, J. H., & Jensen, A. R. (1991a). The nature of psychometric *g*: Unitary process or a number of independent processes? Intelligence, 15, 397-442.

Kranzler, J. H., & Jensen, A. R. (1991b). Unitary *g*: Unquestioned postulate or empirical fact. Intelligence, 15, 437-448.

Kroemer, K. H. E. (1970). Human strength: Terminology, measurement, and interpretation of data (AMRL-TR-69-9, AD710593). Wright-Patterson Air Force Base, OH: Aerospace Medical Research Laboratory.

Kuhl, J. (1981). Motivational and functional helplessness: The moderating effect of state versus action orientation. Journal of Personality and Social Psychology, 40, 155-170.

Kuhl, J. (1984). Volitional aspects of achievement motivation and learned helplessness: Toward a comprehensive theory of action control. Progress in Experimental Personality Research, 13, 99-171.

Kyllonen, P. C. (1985a). Dimensions of information processing speed (AFHRL-TR-85-87). Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory.

Kyllonen, P. C. (1985b). Theory-based cognitive assessment (AFHRL-TP-85-30). Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory.

Kyllonen, P. C. (1991). Editorial: Principles for creating a computerized test battery, Intelligence, 15, 1-15.

Kyllonen, P. C., & Christal, R. E. (1988). Cognitive modeling of learning abilities: A status report of LAMP (AFHRL-TP-87-66) Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning is (little more than) working memory capacity. Intelligence, 14, 389-433.

Kyllonen, P. C., Lohman, D. F., & Snow, R. E. (1984). Effects of aptitudes, strategy training, and task facets on spatial test performance. Journal of Educational Psychology, 76, 130-145.

Larson, G. E. (1989). Mental counters and sequential memory: Proposals for joint services validation (A proposal to the Technical Advisory Selection Panel). San Diego, CA: Navy Personnel Research and Development Center.

Larson, G. E., & Alderton, D. L. (1992, August). Test/retest results for the enhanced computer-administered test (ECAT) battery. Paper presented at the annual meeting of the American Psychological Association, Washington, DC.

Latham, G. P., & Saari, L. M. (1984). Do people do what they say? Further studies on the situational interview. Journal of Applied Psychology, 69, 569-573.

Lau, A. W., & Abrahams, N. M. (1971). Reliability and predictive validity of the Navy Vocational Interest Inventory (Research Report SRR 71-16). San Diego, CA: Navy Personnel and Training Research Laboratory.

Laurence, J. H. (in press). Education standards and military selection: From the beginning. In J. H. Laurence & T. Trent (Eds.), Adaptability screening for the military. Washington, DC: Department of Defense.

Lent, R. W., & Hackett, G. (1987). Career self-efficacy: Empirical status and future directions. Journal of Vocational Behavior, 30, 347-382.

Levine, J. M., Brahlek, R. E., Eisner, E. J., & Fleishman, E. A. (1979). Trainability of abilities: Training and transfer of abilities related to electronic fault-finding (ARRO Technical Report, 3010 R79-2). Washington, DC: Advanced Research Resources Organization.

Levine, J. M., Schulman, D., Brahlek, R. E., & Fleishman, E. A. (1980). Trainability of abilities: Training and transfer of spatial visualization (ARRO Technical Report, 3010 TR3). Washington, DC: Advanced Research Resources Organization.

Linn, R. L. (1986). Comments on the *g* factor in employment testing. Journal of Vocational Behavior, 29, 340-362.

Linn, M. C., & Peterson, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. Child Development, 56, 1479-1498.

Lockwood, R.E., & Shipley, B.D., Jr. (1984). Evaluation of the revised flight aptitude selection test (Technical Report ASI479-020-84). Fort Rucker, AL: Anacapa Sciences, Inc.

Lohman, D. F. (1979). Spatial ability: A review and reanalysis of the correlational literature (Aptitude Research Project Technical Report No. 8). Stanford, CA: Stanford University.

Lohman, D. F. (1988). Spatial abilities as traits, processes, and knowledge. In R. J. Sternberg (Ed.), Advances in the psychology of human intelligence (Vol. 4, pp. 181-248). Hillsdale, NJ: Lawrence Erlbaum Associates.

Lohman, D. F. (1992, May). Implication of cognitive psychology for selection and classification: Three critical assumptions. Proceedings of the ARI Conference on Selection and Classification. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Maccoby, E. E., & Jacklin, C. N. (1974). The psychology of sex differences. Stanford, CA: Stanford University Press.

Mael, F. A. (1991). A conceptual rationale for the domain and attributes of biodata items. Personnel Psychology, 44, 763-792.

Maier, M. H. & Fuchs, E. F. (1969). Development of improved aptitude area composites for enlisted classification (ARI TR 1159, AD 701134). Alexandria, VA: U.S. Army Behavioral Science Research Laboratory.

Maier, M. H. & Grafton, F. C. (1981). Aptitude composites for ASVAB 8, 9, and 10 (ARI TR 1308). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Maier, M. H., & Mayberry, P. W. (1989). Evaluating minimum aptitude standards (CRM 89-9). Alexandria, VA: Center for Naval Analyses.

Maier, M. H. & Truss, A. R. (1985). Validity of the Armed Services Vocational Aptitude Battery forms 8, 9, and 10 with application to forms 11, 12 ,13, and 14 (CNR 102). Alexandria, VA: Center for Naval Analyses.

Martin, C. J. (1992, March). Computerized testing. Briefing presented to the Armed Services Vocational Aptitude Battery Revision Workshop. Arlington, VA: Defense Manpower Data Center.

Mattson, J. D., Abrahams, N. M., & Hetter, R. D. (1985). Use of personal history information to predict Naval Academy disenrollment. Proceedings of the 27th Annual Conference of the Military Testing Association, 2, 503-508.

Mayberry, P. W., & Hiatt, C. M. (1990). Incremental validity of new tests in prediction of infantry performance (CRM 90-110). Alexandria, VA: Center for Naval Analyses.

McBride, J. R. (1991, August). Content and structural comparisons of ASVAB and DAT. Paper presented at the meeting of the American Psychological Association, San Francisco, CA.

McCrae, R. R., & Costa, P. T., Jr. (1985). Updating Norman's adequate taxonomy: Intelligence and personality dimensions in natural language and questionnaires. Journal of Personality and Social Psychology, 49, 710-721.

McDaniel, M. A. (1989). Biographical constructs for predicting employee suitability. Journal of Applied Psychology, 74, 964-970.

McGee, M. G. (1979). Human spatial abilities: Psychometric studies and environmental, genetic, hormonal, and neurological influences. Psychological Bulletin, 86, 889-918.

McHenry, J. J. (1987). Development and evaluation of three computerized psychomotor tests. Unpublished doctoral dissertation, University of Minneapolis, Minnesota.

McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. A., & Ashworth, S. (1990). Project A validity results: The relationship between predictor and criterion domains. Personnel Psychology, 43, 335-354.

McHenry, J. J., Toquam, J. L., Rosse, R. L., Peterson, N. G., & McGue, M. K. (1987). Perceptual/psychomotor computer-administered measures: Field test. In N. G. Peterson (Ed.), Development and field test of the trial battery for Project A (ARI TR-739). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

McHenry, J. J., & Rose, S. R. (1988). Literature review: Validity and potential usefulness of psychomotor ability tests for personnel selection and classification (ARI Research Note 88-13). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

McLaughlin, D. J., Rossmeissl, P. G., Wise, L. L., Brandt, D. A., & Wang, M. (1984). Development and validation of Army selection and classification measures, Project A: Validation of current and alternative ASVAB Area Composites, based on training and SQT information on FY 1981 and FY 1982 enlisted accessions (ARI TR 651). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Means, B., & Perelman, L. S. (1984). The development of the Educational and Background Information Survey (Report No. FR-PRD-84-3). Alexandria, VA: Human Resources Research Organization.

Michael, W. B., Guilford, J. P., Fruchter, B., & Zimmerman, W. S. (1957). The description of spatial-visualization abilities. Educational and Psychological Measurement, 17, 185-199.

Mittelholtz, D. J., & Lohman, D. F. (1986, April). Effects of extensive practice without feedback on a spatial synthesis task. Paper presented at the meeting of the American Educational Research Association, San Francisco, CA.

Moreno, K. E., Wetzel, C. D., McBride, J. R., & Weiss, D. J. (1984). Relationship between corresponding Armed Services Vocational Aptitude Battery (ASVAB) and computerized adaptive testing (CAT) subtests. Applied Psychological Measurement, 8, 155-163.

Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. Journal of Applied Psychology, 75, 640-647.

Moss, F. A., Hunt, T., Omwake, K. T., & Woodward, L. G. (1955). Manual for the George Washington Series Social Intelligence Test. Washington, DC: The Center for Psychological Service.

Mumford, M. D., & Owens, W. A. (1987). Methodology review: Principles, procedures, and findings in the application of background data measures. Applied Psychological Measurement, 11, 1-31.

Murphy, K. R. (1989). Is the relationship between cognitive ability and job performance stable over time? Human Performance, 2, 183-200.

Myers, D. C., Gebhardt, D. L., Crump, C. E., & Fleishman, E. A. (1984). Validation of the military entrance physical strength capacity test (ARI-TR-610). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Neiner, A. G., & Owens, W. A. (1982). Relationships between two sets of biodata with seven years separation. Journal of Applied Psychology, 67, 146-150.

Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. Journal of Abnormal and Social Psychology, 66, 574-583.

North, R. A., & Griffin, G. R. (1977). Aviator selection 1919-1977 (NAMRL Special Report 77-2; AD-A048105). Pensacola, FL: Naval Aerospace Medical Research Laboratory.

O'Dell, J. W. (1971). Method for detecting random answers on personality questionnaires. Journal of Applied Psychology, 55, 380-383.

Oppler, S. H., Peterson, N. G., & Russell, T. L. (in press). Basic LVI validation results. In J. P. Campbell and L. Zook (Eds.), Building and retaining the career forces: FY 1991 annual report. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Oppler, S. H., Peterson, N. G., Whetzel, D. L., Steele, D., Childs, R. A., Park, R. K., Rosse, R. L., Rehling, J. F., Brantorer, T. M., & Kieckhaefer, W. F. (1992). Selection and classification tests for critical military occupational specialties (ARI Research Note, in processing). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Owens, W. A. (1976). Background data. In M. D. Dunnette (Ed.), Handbook of industrial and organizational psychology (pp. 609-644). Chicago, IL: Rand McNally.

Owens, W. A., & Schoenfeldt, L. F. (1979). Towards a classification of persons [Monograph]. Journal of Applied Psychology, 64, 569-607.

Palmer, P., Haywood, C. S., & Curran, L. T. (1989). Predictive validity of the computerized adaptive ASVAB [Summary]. Proceedings of the 31st Annual Conference of the Military Testing Association, 7-12.

Petersen, C. R., & Al-Haik, A. R. (1976). The development of the Defense Language Aptitude Battery (DLAB). Educational and Psychological Measurement, 36, 369-380.

Peterson, N. G. (Ed.) (1987). Development and field test of the trial battery for Project A (ARI Technical Report 739). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Peterson, N. G., Gialluca, K. A., Borman, W. C., Carter, G. W., & Rosse, R. L. (1990). An investigation of methods for simplifying Navy classification (Institute Report 189). Minneapolis, MN: Personnel Decisions Research Institute.

Peterson, N. G., Hough, L. M., Dunnette, M. D., Rosse, R. L., Houston, J. S., & Toquam, J. L. (1990). Project A: Specification of the predictor domain and development of new selection/classification tests. Personnel Psychology, 43, 247-276.

Peterson, N. G., Russell, T. L., Hallam, G., Hough, L. M., Owens-Kurtz, C., Gialluca, K., & Kerwin, K. (1990). Analysis of the experimental predictor battery: LV Sample. In J. P. Campbell and L. M. Zook (Eds.), Building and retaining the career force: New procedures for accessing and assigning Army enlisted personnel, Annual Report, 1990 Fiscal Year (ARI FR-PRD-90-6). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Popham, S. M., & Holden, R. R. (1990). Assessing MMPI construct through the measurement of response latencies. Journal of Personality Assessment, 54, 469-478.

Rayman, J. R. (1976). Sex and the single interest inventory: The empirical validation of sex-balanced interest inventory items. Journal of Counseling Psychology, 23, 239-246.

Ree, M. J., & Earles, J. A. (1991a, April). Estimating psychometric *g*: An application of the Wilk's theorem. Paper presented at the annual meeting of the American Psychological Association, San Francisco, California.

Ree, M. J., & Earles, J. A. (1991b). Predicting training success: Not much more than *g*. Personnel Psychology, 44, 321-332.

Ree, M. J., & Earles, J. A. (1992a). Subtest and composite validity of ASVAB forms 11, 12, and 13 for technical training courses (AL-TR-1991-0107). Brooks Air Force Base, TX: Human Resources Directorate, Armstrong Laboratory.

Ree, M. J., & Earles, J. A. (1992b). Intelligence is the best predictor of job performance. Current Directions in Psychological Science, 1, 86-89.

Ree, M. J., Earles, J. A., & Teachout, M. S. (1992). General cognitive ability predicts job performance. (AL-TP-1991-0057). Brooks AFB, TX: Human REsources Directorate, Armstrong Laboratory.

Ree, M. J., Mullins, C. J., Matthews, J. J., & Massey, R. H. (1982). Armed Services Vocational Aptitude Battery: Item and factor analysis of Forms 8, 9, and 10 (AFHRL-TR-81-55). Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory.

Reilly, R. R., & Chao, G. T. (1982). Validity and fairness of some alternate employee selection procedures. Personnel Psychology, 35, 1-62.

Reilly, R. R., Zedeck, S., & Tenopyr, M. L. (1979). Validity and fairness of physical ability tests for predicting craft jobs. Journal of Applied Psychology, 64, 262-274.

Robertson, D. W. (1982). Development of an occupational strength test battery (STB) (NPRDC TR 82-42). San Diego, CA: Navy Personnel Research and Development Center.

Roe, A. (1956). The psychology of occupations. New York, NY: Wiley.

Rogers, D. L., Roach, B. W., & Short, L. O. (1986). Mental ability testing in the selection of Air Force officers: A brief historical overview (AFHRL-TP-86-23). Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory.

Rothstein, H. R., Schmidt, F. L., Erwin, F. W., Owens, W. A., & Sparks, C. P. (1990). Biographical data in employment selection: Can validities be made generalizable? Journal of Applied Psychology, 75, 175-184.

Ruch, F. L., & Ruch, W. W. (1967). The K factor as a (validity) suppressor variable in predicting success in selling. Journal of Applied Psychology, 51, 201-204.

Russell, T. L., Knapp, D. J., & Campbell, J. P. (1992). Building a joint-Service classification research roadmap: Defining research objectives (HumRRO IR-PRD-92-10). Alexandria VA: Human Resources Research Organization.

Sands, W. A. (1987). The joint-Service CAT-ASVAB program: Progress and plans [Summary]. Proceedings of the 29th Annual Conference of the Military Testing Association, 335-338.

Sands, W. A. (1988). Computerized adaptive testing: The CAT-ASVAB program [Summary]. Proceedings of the 30th Annual Conference of the Military Testing Association, 458-462.

Sands, W. A. (1990). Joint-service computerized aptitude testing [Summary]. Proceedings of the 32nd Annual Conference of the Military Testing Association, 245-250.

Sarason, I. G., Sarason, B. R., Keefe, D. E., Hayes, B. E., & Shearin E. N. (1986). Cognitive interference: Situational determinants and traitlike charactistics. Journal of Personality and Social Psychology, 51, 215-226.

Schmitt, N., Gooding, R. Z., Noe, R. A., & Kirsch, M. P. (1984). Meta-analyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. Personnel Psychology, 37, 407-422.

Schrader, A. D., & Osburn, H. G. (1977). Biodata faking: Effects of induced subtlety and position specificity. Personnel Psychology, 30, 395-404.

Schratz, M. K., & Ree, M. J. (1989). Enlisted selection and classification: Advances in testing. Military personnel measurement: Testing, assignment, evaluation (pp. 1 - 40). New York: Praeger.

Schwab, D. P. (1971). Issues in response distortion studies of personality inventories: A critique and replicated study. Personnel Psychology, 24, 637-647.

Segall, D. O. (1991). CAT-ASVAB equating update. Briefing presented to the Defense Advisory Committee on Military Personnel Testing, Washington, DC.

Sevy, B. A. (1983). Sex-related differences in spatial ability: The effects of practice. Unpublished doctoral dissertation, University of Minnesota, Minneapolis.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. Science, 171, 701-703.

Siegel, A. I., Federman, P. J., & Welsand, E. H. (1980). Perceptual/psychomotor requirements basic to performance in 35 Air Force specialties (AFHRL-TR-80-26). Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory.

Siem, F. M. (1990). Predictive validity of an automated personality inventory for Air Force pilot selection C (AFHRL-TP-90-55). Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory.

Siem, F. M. (1991, October). Predictive validity of response latencies from computer-administered personality tests. Paper presented at the 33rd Annual Conference of the Military Testing Association, San Antonio, TX.

Siem, F. M. (1992, August). Incremental validity of personality measures for predicting pilot training performance. Paper presented at the meeting of the American Psychological Association, Washington, DC.

Silva, J. M., White, L. A., & Rumsey, M. G. (1991, month). Relationship of cognitive aptitudes to success in foreign language training. Paper presented at the meeting of the American Psychological Association, San Francisco, CA.

Skinner, J., & Ree, M. J. (1987). Air Force Officer Qualifying Test (AFOQT): Item and factor analysis of form O (AFHRL-TR-86-68). Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory.

Smith, I. M. (1948). Measuring spatial abilities in school pupils. Occupational Psychology, 22, 150-159.

Snow, R. E. (1989). Cognitive-conative aptitude interactions in learning. In R. Kanfer, P. L. Ackerman, & R. Cudek (Eds.), Learning and individual differences: Abilities, motivation, and methodology (pp. 435-474). Hillsdale, NJ: Erlbaum.

Society of Industrial and Organizational Psychology, Inc. (1987). Principles for the validation and use of personnel selection procedures (Third Edition). College Park, MD: Author.

Spearman, C. (1927). The abilities of man. New York: MacMillan Co.

Sperl, T. C., & Ree, M. J. (1990). Air Force Officer Qualifying Test (AFOQT): Development of quick score composites for forms P1 and P2 (AFHRL-TR-90-3). Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory.

Sperl, T. C., Ree, M. J., & Steuk, K. W. (1990). Air Force Officer Qualifying Test (AFOQT) and Armed Service Vocational Aptitude Battery (ASVAB): Analysis of crewman measurement and attributes (AFHRL-TR-90-37). Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory.

Sperl, T. C., Ree, M. J., & Steuk, K. W. (1992). Air Force Officer Qualifying Test (AFOQT) and Armed Service Vocational Aptitude Battery (ASVAB): Analysis of crewman measurement and attributes. Military Psychology, 4, 175-188.

Steinhaus, S. D. (1988). Predicting military attrition from educational and biographical information (Report No. FR-PRD-88-06). Alexandria, VA: Human Resources Research Organization.

Steinhaus, S. D., & Waters, B. K. (1991). Biodata and the application of a psychometric perspective. Military Psychology, 3, 1-23.

Sternberg, R. J. (1985). Beyond IQ: A triarchic theory of human intelligence. Cambridge, England: Cambridge University Press.

Sternberg. R. J. (1990). T & T is an explosive combination: Technology and testing. Educational Psychologist, 25, 201-222.

Sternberg, R. J. (in press). The PRSVL model of person-context interaction in the study of Human potential. Proceedings of the ARI Conference on Selection and Classification. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Steuck, K. W., Watson, T. W., & Skinner, J. (1988). Air Force Officer Qualifying Test (AFOQT): Forms P pre-implementation analyses and equating (AFHRL-TP-88-6). Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory.

Stoker, P., Hunter, D. R., Kantor, J. E., Quebe, J. C., & Siem, F. M. (1987). Flight screening program effects on attrition in undergraduate pilot training (AFHRL-TP-86-59). Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory.

Stringer, P. (1975). Drawing training and spatial ability. Ergonomics, 18, 101-108.

Sympson, J. B., Weiss, D. J., & Ree, M. J. (1982). Predictive validity of conventional and adaptive tests in an Air Force training environment (AFHRL-TR-81-40). In P. Palmer, C. S. Haywood, & L. T. Curran (1989), Predictive validity of the computerized adaptive ASVAB [Summary]. Proceedings of the 31st Annual Conference of the Military Testing Association, 7-12.

Synk, D. J. (1983, August). The effect of gender on general Aptitude Test Battery validity and test scores. Paper presented at the meeting of the American Psychological Association, Anaheim, CA.

Synk, D. J., & Swarthout, D. (1987). Comparison of black and nonminority validities for the General Aptitude Test Battery (USES Test Research Report No. 51). Washington, DC: U.S. Department of Labor.

Tellegen, A. (1982). Brief manual for the Differential Personality Questionnaire. Unpublished manuscript, University of Minnesota, Minneapolis.

Tellegen, A., Kamp, J., & Watson, D. (1982). Recognizing individual differences in predictive structure. Psychological Review, 89, 95-105.

Tett, B. P., Jackson, D. N., & Rothstein, M. R. (1991). Personality measures as predictors of job performance: A meta-analytic review. Personnel Psychology, 44, 703-742.

Thorndike, E. L. (1920). Intelligence and its use. Harper's Magazine, 140, 227-235.

Thorndike, R. L. (1986). The role of general ability in prediction. Journal of Vocational Behavior, 29, 332-339.

Thornton, G. C., & Byham, W. C. (1982). Assessment centers and managerial performance. New York: Academic Press.

Thurstone, L. L. (1938). Primary mental abilities. Chicago: University of Chicago Press.

Thurstone, L. L., & Thurstone, T. G. (1941). The Primary Mental Abilities Tests. Chicago: Science Research Associates.

Tittle, C. K., & Zytowski, D. G. (Eds.). (1978). Sex-fair interest measurement: Research and implications. Washington, DC: National Institute of Education.

Toquam, J. L., Corpe, V. A., & Dunnette, M. D. (1989). Literature review: Cognitive abilities - theory, history, and validity (ARI Research Note 91-28). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Toquam, J., Peterson, N. G., Rosse, R. L., Ashworth, S., Hanson, M. A., & Hallam, G. (1986). Concurrent validity data analyses: Cognitive paper-and-pencil and computer-administered predictors. Presentation to the Project A Scientific Advisory Group, Minneapolis, MN.

Trent, T. (in press). The Armed Services Applicant Profile (ASAP). In J. H. Laurence & T. Trent (Eds.), Adaptability screening for the military. Washington, DC: Department of Defense.

Trent, T., Atwater, D. C., & Abrahams, N. M. (1986). Biographical screening of military applicants: Experimental assessment of item response distortion. Paper presented to the Tenth Psychology in the DoD symposium.

Trent, T., & Quenette, M. A. (1992). Armed Services Applicant Profile (ASAP): Development and validation of operational forms (NPRDC-TR-92-9). San Diego, CA: Navy Personnel Research and Development Center.

Tupes, E. C., & Christal, R. E. (1961). Recurrent personality factors based on trait ratings (USAF ASD Technical Report Number 61-97). Lackland Air Force Base, TX: U.S. Air Force.

Tyler, L. E. (1965). The psychology of human differences (3rd edition). New York: Appleton-Century-Crofts.

Tyler, L. E. (1986). Back to Spearman? Journal of Vocational Behavior, 29, 445-450.

Uniform Guidelines in Employee Selection Procedures (1978). Federal Register, 43, 38290-38315.

Valentine, L. (1977). Navigator-observer selection research: Development of new Air Force Officer Qualifying Test navigator-technical composite (AFHRL-TR-77-36). Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory.

Vernon, P. E. (1950). The structure of human abilities. London: Methven.

Vernon, P. A. (1990). The use of biological measures to estimate behavioral intelligence. Educational Psychologist, 25, 293-304.

Walker, C. B. (1989). ARI's nominations of new tests for the Technical Advisory Selection Panel (TASP) (ARI-WP-RS-89-6). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Walker, R. E. & Foley, J. M. (1973). Social intelligence: Its history and measurement. Psychological Reports, 33, 838-864.

Waters, B. K. (in press). The politics of adaptability screening research in the Department of Defense. In J. Laurence & T. Trent (Eds.), Adaptability screening for the military. Washington, DC: Department of Defense.

Waters, B. K., Barnes, J. D., Foley, P., Steinhaus, S. D., & Brown, D. C. (1988). Estimating the reading skills of military applicants: The development of an ASVAB to RGL conversion table (Final Report No. 88-22). Alexandria, VA: Human Resources Research Organization.

Watson, T. W., Alley, W. E., & Southern, M. E. (1979). Initial development of operational composites for the Vocational Interest Career Examination. Paper presented at the annual meeting of the Military Testing Association, San Diego, CA.

Weeks, J. L., Mullins, C. J., & Vitola, B. M. (1975). Airman classification batteries from 1948 to 1975: A review and evaluation (AFHRL-TR-75-78, AD-A026470). Lackland Air Force Base, TX: Air Force Human Resources Laboratory.

Wegner, T. G., & Ree, M. J. (1986). Alternative Armed Forces Qualification Test composites (AFHRL-TP-86-27). Brooks Air FOrce Base, TX: U.S. Air Force Human Resouces Laboratory.

Welsh, J. R., Jr., Kucinkas, S. K., & Curran, L. T. (1990). Armed Services Vocational Aptitude Battery (ASVAB): Integrative review of validity studies (AFHRL-TR-90-22). Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory.

Welsh, J. R., Jr., Watson, T. W., & Ree, M. J. (1990). Armed Services Vocational Aptitude Battery (ASVAB): Predicting military criteria from general and specific abilities (AFHRL-TR-90-63). Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory.

Whetstone, R. D., & Hayles, V. R. (1975). The SVIB and black college men. Measurement and Evaluation in Guidance, 8, 105-109.

White, L. A., Hanser, L. M., & Park, R. K. (1988, October). A preliminary investigation of the relationship between the ASVAB and DLAB. Paper presented at the meeting of the Military Testing Association, Arlington, VA.

Willerman, L. (1979). The psychology of individual and group differences. San Francisco, CA: W. H. Freeman and Company.

Wise, L. L., & McDaniel, M. A. (1991, August). Cognitive factors in the Armed Services Vocational Aptitude Battery and the General Aptitude Test Battery. Paper presented at the meeting of the American Psychological Association, San Francisco, CA.

Wolfe, J. H., Alderton, D. L., & Larson, O. E. (1992, August). Incremental validity of new computerized aptitude tests for predicting training performance in nine Navy technical schools. Paper presented at the meeting of the American Psychological Association, Washington, DC.

Wood, R. E., & Bandura, A. (1989). Impact of conceptions of ability on self-regulatory mechanisms and complex decision making. Journal of Personality and Social Psychology, 56, 407-415.