Technical Report 1033

Development of a Roadmap for Special Forces Selection and Classification Research

Teresa L. Russell, Michelle R. Rohrback,
and Marguerite T. Nee
Human Resources Research Organization

Jennifer L. Crafts and Norman G. Peterson
American Institutes for Research

Fred A. Mael
U.S. Army Research Institute

October 1995

United States Army Research Institute
for the Behavioral and Social Sciences

U.S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Director

Research accomplished under contract
for the Department of the Army

Technical review by

Judith E. Brooks
Martha L. Teplitzky

NOTICES

DISTRIBUTION: Primary distribution of this report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, ATTN: PERI-POX, 5001 Eisenhower Ave., Alexandria, Virginia 22333-5600.

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE 1995, October	3. REPORT TYPE AND DATES COVERED FINAL 6/94 - 12/94

4. TITLE AND SUBTITLE
Development of a Roadmap for Special Forces Selection and Classification Research

5. FUNDING NUMBERS
MDA903-92-D-0206 DO 002
0603007A
A792
1223
C06

6. AUTHOR(S)
Teresa L. Russell, Michelle R. Rohrback, Marguerite T. Nee (HumRRO), Jennifer L. Crafts, Norman G. Peterson (AIR), and Fred A. Mael (ARI)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Human Resources Research Organization (HumRRO)
Alexandria, VA 22314
American Institutes for Research
Washington, DC 20007

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Institute for the Behavioral and Social Sciences
ATTN: PERI-RS
5001 Eisenhower Ave.
Alexandria, VA 22333-5600

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
ARI Technical Report 1033

11. SUPPLEMENTARY NOTES
COR: Fred A. Mael

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words): The purpose of this project was to develop an agenda for Special Forces (SF) selection and classification research. Job analysis data, interviews, field observation, and expert judgments about the quality of measures formed the foundation for the Roadmap. The resulting Roadmap is composed of eight projects. Projects 1 and 2, *Concurrent Criterion-Related Validation of Readily Available Predictor Measures Against on the Job Performance* and *Development and Implementation of Content Valid Job Sample Tests*, supplement SF selection and classification with measures of leadership, temperament, and communication and analytic skills that could be implemented quickly. Project 3, *Validation of Measures of Conventional Army Task Proficiency, Experience and Preference Against Training Performance*, addresses the fit between individuals and SF Jobs. Project 4, *Validation of Training Performance Against on the Job Performance*, would evaluate the usefulness of training data for predicting job performance. Project 5, *Predictive Validation of All Predictors Against on the Job Performance*, the ultimate test of any selection system, requires maintaining databases for validation purposes. Projects 6-8 involve the development of information to facilitate decision making at the U.S. Army John F. Kennedy Special Warfare Center and School. The are: *Development of a Selection and Training Decision Simulator* (Project 6), *Review of New Measures of Leader Problem Solving Performance* (Project 7), and *Training Performance Study* (Project 8).

14. SUBJECT TERMS				15. NUMBER OF PAGES 269
Job analysis	Validation	Research agenda	Predictors	16. PRICE CODE
Criteria	Special Forces			

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

Technical Report 1033

Development of a Roadmap for Special Forces Selection and Classification Research

Teresa L. Russell, Michelle R. Rohrback,
and Marguerite T. Nee
Human Resources Research Organization

Jennifer L. Crafts and Norman G. Peterson
American Institutes for Research

Fred A. Mael
U.S. Army Research Institute

Selection and Assignment Research Unit
Michael G. Rumsey, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600

Office, Deputy Chief of Staff for Personnel
Department of the Army

October 1995

Army Project Number
2O363007A792

Manpower and Personnel

FOREWORD

The U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) performs behavioral research to develop methods of selecting and training personnel for Army jobs. The increased variety and complexity of Special Forces missions throughout the world have created a need for systematic, comprehensive procedures for assessing Special Forces candidates. In response to this need, the U.S. Army John F. Kennedy Special Warfare Center and School (USAJFKSWCS) initiated the Special Forces Assessment and Selection (SFAS) program in June 1988. ARI has a commitment to support Special Forces through research on required skills and aptitudes.

The purpose of the current project was to develop an agenda--a Roadmap--for Special Forces selection and classification research. While SFAS has proven to be a useful tool for the selection of physically and mentally capable personnel, it does not measure a number of other skills that emerged in recent analyses of Special Forces jobs. This project expanded the job analysis work by identifying measures that could be used to assess important skills and concluded with recommendations for future research in eight areas.

ZITA M. SIMUTIS
Deputy Director
(Science and Technology)

EDGAR M. JOHNSON
Director

DEVELOPMENT OF A ROADMAP FOR SPECIAL FORCES SELECTION AND CLASSIFICATION RESEARCH

EXECUTIVE SUMMARY

Research Requirement:

The purpose of the current project was to develop an agenda—a Roadmap—for Special Forces (SF) selection and classification research. It had three specific objectives:

(1) Identify tests, exercises, and other measures (i.e., predictors and criteria) likely to be useful to the U.S. Army John F. Kennedy Special Warfare Center and School (USAJFKSWCS),

(2) Identify current and future SF selection directions based on SF missions and trends, and

(3) Organize information into projects that will lead to enhancement of SF selection and classification.

A recent analysis of SF jobs (Russell, Crafts, Tagliareni, McCloy, & Barkley, 1994) laid the foundation for this project. The job analysis identified 47 attributes relevant to successful performance in SF jobs and 26 critical incident-based job performance categories that describe SF jobs. The current project expanded the job analysis work by identifying measures that could be used to assess important attributes and concluded with recommendations for future research.

Procedure:

The first step was to identify potentially useful predictor and criterion measures through an expert judgment procedure. We began by interviewing U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) researchers and gathering documents describing tests, measures, exercises, and scales that could be made available to USAJFKSWCS. Using the interview and document information, we prepared descriptions of available and in-development measures and conducted an expert judgment exercise involving about 20 psychologists from ARI, the Human Resources Research Organization (HumRRO), and the American Institutes of Research (AIR). Experts rated the extent to which each exercise, test, or scale measured attributes or, in the case of criterion measures, job performance categories. The expert judgment exercises yielded reliable estimates of the extent to which tests, scales, and exercises measure SF attributes and SF performance categories. Intraclass correlation coefficients adjusted to the number of raters in the exercise ranged from .83 to .96 with a median of .90. Thus, expert judgment data that formed the basis for decisions about instruments were of high quality.

The second step was to gather information about trends in SF missions and future directions from key decision makers in SF USAJFKSWCS. We conducted one-on-one interviews with officers from the 1st Special Warfare Training Group (SWTG), Special Operations Proponency Office (SOPO), Directorate of Training Doctrine (DOTD), the 3rd Special Forces Group Airborne (SFG[A]), and the 7th SFG[A]. We learned that key decision makers in SF and USAJFKSWCS expect that their largest primary mission, foreign internal defense (FID), will continue to be the major focus of SF and that other types of missions involving cross-cultural interactions such as humanitarian aid and coalition warfare will grow. Missions without a cross-cultural emphasis such as direct action a re expected to diminish. Attributes relevant to building relationships with indigenous people are therefore expected to be highly important to success on future missions.

The third step was to organize information from the interviews and the expert judgments into a Roadmap for selection and classification research. Four principles guided Roadmap development:

(1) The measures selected for the Roadmap should be of high *quality* based on expert judgment.

(2) The measures selected for the Roadmap should be *feasible* with minimal development cost.

(3) As a whole, the measure selected should be *comprehensive*; that is, they should measure as many of the attributes needed for successful performance in SF as possible.

(4) Attributes related to the job performance category B. *Building effective relationships with indigenous people* are high in *priority* because this performance category is an emphasis for future SF missions.

Using those principles as decision rules, we examined the expert judgments and identified sets of test measures and scales likely to be useful for SF. We identified sets of predictors that could be codeveloped and covalidated. We developed projects based on the validation needs for each specific type of predictor set. Collectively, those projects formed the Roadmap.

Findings:

The Roadmap is composed of eight projects designed to enhance SF selection and classification. Five of the eight projects are predictor validation steps, and the remaining three projects involve the development of tools and information to facilitate decision making at USAJFKSWCS. The eight projects are:

Project 1 Concurrent criterion-Related Validation of Readily Available Predictor measures Against on the Job Performance.

Project 2	Development and Implementation of Content Valid Job Sample Test (Role Plays)
Project 3	Validation of Measures of Conventional Army Task Proficiency, Experience and Preference Against Training Performance
Project 4	Validation of Training Performance Against on the Job Performance
Project 5	Predictive Validation of All Predictors Against on the Job Performance
Project 6	Development of a Selection and Training Decision Simulator
Project 7	Review of New Measures of Leader Problem Solving Performance
Project 8	Training Performance Study

Projects 1 and 2, *Concurrent Criterion-Related Validation of Readily Available Predictor Measures Against* ***on the*** *Job Performance* and Development *and Implementation of Content Valid Job Sample Test*, are designed to supplement SF selection and classification with measures of leadership, temperament, and communication and analytic skills. Both projects would provide highly useful measures that address many of the SF attributes identified in the job analysis. Based on SF and USAJFKSWCS needs and priorities, Projects 1 and 2 should be conducted concurrently and as soon as possible. Project 1 will take about 8-12 months, and Project 2 will be shorter, perhaps 6-10 months (to the completion of the draft report). Those two projects together would provide strong measures in areas that are currently not well addressed in the selection system.

After the completion of Projects 1 and 2, it would be reasonable to conduct projects 3 and 4. Project 3, *Validation of Measures of Conventional Army Task Proficiency, Experience and Preference Against Training Performance*, addresses the fit between individuals and SF jobs and could be conducted with a year's time. Project 4, *Validation of Training Performance Against* ***on the*** *Job Performance*, is of interest to USAJFKSWCS. It would evaluate the usefulness of training data for predicting job performance. Clearly, Projects 3 and 4 build on each other because Project 3 necessitates training criteria, and in Project 4 those criteria become predictors of on-the-job performance. It would be most efficient to begin Project 3 and then start Project 4 several months into Project 3.

Similarly, Projects 3 and 4 build up to Project 5, *Predictive Validation of All Predictors Against* ***on the*** *Job Performance*—a longitudinal project that involves careful database development and maintenance. But before starting predictive validation it would be wise to conduct Project 7, *Review of New Measures of Leader Problem Solving Performance*. The results of the expert judgment exercise showed that leader problem solving measures which are in development in ARI projects could be highly useful to SF, particularly for measuring officer attributes. It will be important to consider their potential usefulness again in 2 or 3 years—before beginning the predictive validation project.

Projects 6 and 8 could be conducted at any point in time. *The Development of a Selection and Training Decision Simulator* (Project 6) would result in a piece of software that would allow SWTG decision makers to analyze the potential impact of change in the sequence of selection and training activities. The eighth project, *Training Performance Study*, involves developing a procedure for measuring training gains of individuals trained by SF soldiers. Such a procedure would result in (1) feedback to teams on their training accomplishments and (2) information SF could use to illustrate its training accomplishments to its clients.

Utilization of Findings:

The Roadmap can be used to guide future research and the assignment of research priorities.

DEVELOPMENT OF A ROADMAP FOR SPECIAL FORCES SELECTION AND CLASSIFICATION RESEARCH

CONTENTS

Page

INTRODUCTION 1

Purpose 1
Overview of Special Forces 1
SF Selection and Classification Procedures 3
History of SF Selection and Classification Research 6
The Job Analysis 8
Objectives of the Roadmap 10

DEVELOPMENT OF EXPERT JUDGMENT LINKAGES 13

Collection of Expert Judgments About Predictor Measures 14
Collection of Expert Judgments About Criterion Measures 43
Collection of Expert Judgments About the Importance of SF Performance Categories for SF Missions 51
Conclusions 53

ORGANIZATION OF EXPERT JUDGMENTS AND OBSERVATIONS INTO A ROADMAP FOR SELECTION AND CLASSIFICATION RESEARCH 55

Development of Information About Future Trends 55
Development of Predictor Sets 56
Development of Criterion Sets 60
Development of the Roadmap 64
Recommendations 68

REFERENCES 71

APPENDIX A. Special Forces Assessment and Selection Data Analysis A-1

B. Cognitive Measures B-1

C. Biographical, Interest, and Temperament Measures C-1

D. Psychomotor/Physical Measures D-1

Page

APPENDIX E. Simulations, Administrative Indices, Ratings, and Other Performance Measures E-1

F. Predictor Description Bibliography F-1

G. Predictor Expert Judgment Exercise Instructions G-1

H. Predictor Measures Rated Most Highly for Each Attribute H-1

I. Descriptions of Criterion Measures I-1

J. Criterion Description Bibliography J-1

K. Criterion Expert Judgment Exercise Instructions K-1

L. Mission Performance Expert Judgment Exercise Instructions L-1

M. Recommendations for the Development of the Special Forces Biographical, Interest, and Temperament Survey (SFBITS) M-1

LIST OF TABLES

Table 1. Mean Extent of Measurement Ratings for Cognitive Instruments 21

2. Mean Extent of Measurement Ratings for Noncognitive Instruments 26

3. Mean Extent of Measurement Ratings for Physical and Psychomotor Instruments 31

4. Mean Extent of Measurement Ratings for Performance Measures 36

5. Intraclass Correlation Coefficients (ICCs) for Expert Judgments 41

6. Mean Judgments of the Relevance of Measures to SF Performance Categories 48

7. Mean Judgments of the Importance of Performance Categories for Missions 52

Page

LIST OF FIGURES

Figure 1. SF Roles and Performance Categories Based on Performance Examples 11

2. The Roles of Expert Judgments in Validation Paradigms 13

3. SF Selection and Classification Research Planning Databases 15

4. Measures included in the Expert Judgment Exercise 17

5. Example Predictor Description Form 18

6. Example Criterion Measure Description Form 45

7. Measures of Special Forces Training and On-the-Job Proficiency (Criteria) 46

8. Importance of Performance Categories A-K for the Five Primary Missions 54

9. Predictor Sets 58

10. Criterion Sets 62

11. Requirements, Strengths, and Deficiencies of the Five Roadmap Validation Projects 66

12. Roadmap Project Timeline 69

DEVELOPMENT OF A ROADMAP FOR SPECIAL FORCES SELECTION AND CLASSIFICATION RESEARCH
CHAPTER I
INTRODUCTION

Purpose

The Roadmap is a series of research projects that were gleaned from job analysis results, interviews with Special Forces (SF) and U.S. Army John F. Kennedy Special Warfare Center and School (USAJFKSWCS) decision-makers, field observations of SF selection and training, and judgments of experts in selection and classification. The starting point for the development of the Roadmap was a thorough job analysis (Russell, Crafts, Tagliareni, McCloy, & Barkley, 1994) which identified job performance dimensions and attributes that are important for successful performance. In turn, the goal of the Roadmap project was to extend the job analysis results by:

- identifying measures for important SF attributes and performance dimensions,
- ensuring that the selection system would meet predicted future needs as well as current requirements, and
- suggesting projects for the development and validation of measures.

This chapter provides an overview of SF in general, describes current SF selection and classification procedures, outlines the history of SF selection and classification research, and reviews the job analysis. It concludes with a discussion of the research rationale for the current Roadmap development project.

Overview of Special Forces

The basic unit within SF is the A detachment (or Operational Detachment - SFOD A). Ideally, an SF team is designed to have 12 members:

Officers
- 1 Detachment Commander (18A), usually a Captain
- 1 Assistant Detachment Commander (180A), a warrant officer, second in command

Advanced MOS
- 1 Operations Sergeant (18Z)
- 1 Assistant Operations and Intelligence Sergeant (18F)

Entry-Level (E-5 to E-7) Enlisted MOS
- 2 Weapons Sergeants (18B)
- 2 Engineer Sergeants (18C)
- 2 Medical Sergeants (18D)
- 2 Communications Sergeants (18E)

Operationally, the full contingent of 12 is not always realized. Shortages of officers, warrant officers, and medical sergeants result in smaller teams. It is common to find teams with a warrant officer and no Captain; in those instances the warrant officer is the team commander. Also, some teams only have one medic. Occasionally, teams are short on other MOS.

Each team is part of a larger structure defined by five active duty Special Forces Groups [Airborne] -- SFG[A]--each of which is responsible for a particular geographic area:

- 1st SFG[A] headquarters at Ft. Lewis, Southeast Asia orientation
- 3rd SFG[A] headquarters at Ft. Bragg, Africa orientation
- 5th SFG[A] headquarters at Ft. Campbell, Southwest Asia orientation
- 7th SFG[A] headquarters at Ft. Bragg, Latin America orientation
- 10th SFG[A] headquarters at Ft. Devens (in process of moving to Ft. Carson), Europe orientation

Geographic orientation influences language requirements for team members, types of missions, and training needs. For example, the 10th SFG[A] operates in cold weather environments; ski and cold weather survival training are important for 10th SFG[A] teams, and team members are likely to be trained in European languages such as Polish or Russian. On the other hand, the 1st SFG[A] works in the Southeast Asia environment, much of which is jungle; team members are likely to be trained in Vietnamese, Chinese, or other Asian languages. Obviously, cultures, social structures, and languages vary considerably across the various geographical orientations.

SF performs five primary missions (Department of Army, 1990):

- Unconventional Warfare (UW),
- Foreign Internal Defense (FID),
- Direct Action (DA),
- Special Reconnaissance (SR), and
- Counterterrorism (CT).

UW and FID missions both involve training indigenous forces, but UW includes guerrilla warfare (GW) and other direct offensive low-visibility, covert, or clandestine operations while FID missions are overt. FID involves training, organizing, and assisting forces for a Host Nation (HN). Both UW and FID missions can be of long duration. DA missions are short-duration, small-scale offensive actions. SR is reconnaissance and surveillance for data gathering purposes, and CT involves offensive measures to prevent, deter, and respond to terrorism.

In addition to the five primary missions, SF performs collateral activities (Department of the Army, 1990) including:

- Security Assistance,
- Humanitarian Assistance,

- Antiterrorism and other Security Activities,
- Counternarcotics,
- Search and Rescue, and
- Special Activities.

SF Selection and Classification Procedures

SF selection and classification is a multi-hurdle approach designed to ensure that SF personnel are well-qualified mentally and physically. There are three main phases: (1) initial screening of applicants, (2) a three-week assessment program (Special Forces Assessment and Selection [SFAS]), and (3) the SF Qualification Course (i.e., the Q-Course or SFQC). MOS assignment is made prior to the third hurdle (i.e., the Q-Course). Assignment to an SFG[A] is made during or after the Q-Course.

In order to apply for SF, specific requirements must be met. SF applicants must (Pleban, Thompson, Valentine, Dewey, Allentoff, & Wesolowski, 1988):

- be a male soldier (E4 to E7) or officer in a promotable status to the grade of captain;
- have a high school diploma or GED;
- have an Armed Services Vocational Aptitude Battery (ASVAB) General Technical (GT) score of 110 or higher;
- be airborne qualified or volunteer for airborne training;
- be able to swim 50 meters unassisted wearing boots;
- meet medical fitness standards as outlined in AR 40-501, DTD 15 May 1989
- pass the Advanced Physical Readiness Test (score 206 using 17-21 year group standards); and,
- be eligible for a Top Secret security clearance.

Applicants must not:

- be under suspension of favorable actions (AR 600-31);
- have been convicted by special or general court martial during current term of service;
- be barred from reenlistment;
- be a prior Special Forces or Airborne voluntary terminee; or
- have quit military school.

Selected applicants attend SFAS where they are tested and exposed to challenging field exercises. The SFAS battery comprises a number of mental, learning, and personality tests as well as a series of field-related assessment activities (Velky, 1990). Soldiers are required, for example, to swim 50 meters while wearing boots and fatigues, to test their agility on the obstacle course, and to go on long treks with a 45-55 pound rucksack -- otherwise known as the "pain bag." As land navigation is important in successful completion of the training, heavy emphasis is placed upon military orienteering events during SFAS (Pleban, Allentoff, & Thompson, 1989; Busciglio, Teplitzky, &

Welborn, 1991). After the first ten days, the candidates are evaluated by a board to determine whether each should continue. Soldiers may voluntarily withdraw from the program at any time. Those who are sent home are told the reasons why they cannot continue and how they may improve in order to reapply. Those who voluntarily withdraw can only be readmitted by exception. The remaining eleven days of activities are designed to evaluate how well individuals function as team members in a variety of physically demanding situations and how well they demonstrate leadership skills. On the twenty-first day, a final selection board determines whether or not each candidate is suitable to go on to the Q-Course. About 50 percent of the applicants who begin SFAS are selected for the Q-Course (Brooks, 1991; Fricke, 1990).

MOS assignment is made by a panel of senior SWC staff called the assignment board. Assignments are based upon the match between the candidate's background, aptitude level, and personal interests and the MOS requirements and SF needs.[1] In making assignments to SF MOS, the board considers the candidate's General Technical (GT), Skilled Technical (ST), and auditory perception test scores as well as the candidate's expressed interest and prior MOS. Some conventional Army MOS are viewed as highly relevant to particular SF MOS. For example, the conventional Army MOS 11B (Infantryman) is thought to have an SF counterpart, 18B (Weapons Sergeant). Other conventional Army to SF counterparts are: 12B (Combat Engineer) and SF 18C (Engineer Sergeant), 31C (Single Channel Radio Operator) and SF 18E (Communications Sergeant), and 91A (Medical Specialist) and SF 18D (Medical Sergeant).

Those who are selected for the Q-Course return to their original branches until they are called to participate (Fricke, 1990). The SFQC takes place primarily at Fort Bragg in North Carolina. The course lasts anywhere from 24 to 55 weeks, depending on the MOS that a candidate enters. Although the sequence of courses and activities has changed over the years and will change again in FY95, it includes several major activities: land navigation and small unit tactics, MOS specialty training, and a field assessment where soldiers are given an understanding of the Special Forces doctrine and organization while they are also trained in airborne and airmobile operations.

As mentioned earlier, there are four entry-level enlisted SF MOS. MOS 18B is SF Weapons Sergeant. The men are trained in such areas as tactics, anti-armor weapons utilization, the functions of all types of U.S. and foreign light weapons, indirect fire operations, manportable air defense weapons, weapons emplacement, and integrated fire control planning. Training lasts for 13 weeks. SF Engineer Sergeant (18C) training includes the topics of building and bridge construction, field fortification, and the use of explosives for both sabotage and demolitions. Again, training lasts for 13 weeks. MOS 18E, that of SF Communications Sergeant, requires an additional eight weeks of training that is actually completed before coming to the SFQC. During this prerequisite time, candidates participate in and pass the Advanced International Morse Code (AIMC)

---

[1]We attended an MOS assignment board. This description summarizes our observation of the board's process; it is not taken from formal documents.

course. Upon arriving at the SFQC, these individuals are then trained in the installation and operation of SF high-frequency and burst communications equipment; antenna theory; radio wave propagation; and communications operations, procedures, and techniques. Finally, MOS 18D is that of SF Medical Sergeant. Those entering this MOS must complete 31 weeks of the Special Operations Medical Course at Fort Sam Houston, Texas and 13 weeks at Fort Bragg. Training consists of advanced medical procedures that are to be administered both to the team and to indigenous populations. The topics covered include those of trauma management and surgical, dental, and veterinary procedures.

The final qualification period covers such topics as methods of instruction, unconventional warfare operations, and direct action operations. This phase culminates in a guerilla warfare exercise conducted in a national forest in the Fort Bragg area. Here, individuals are expected to be able to function as part of their 12 man team -- an "A-team" or "A-Detachment." Both specialty and common skills are evaluated in this environment as the team attempts to fulfill its mission. It should be noted that the basic objective of any "A-Detachment" is to raise, organize, train, equip, and lead in combat an indigenous light infantry battalion consisting of up to 1,500 members.

Attrition from the Q-Course varies substantially across MOS (Diana, Teplitzky, & Zazanis, 1994). The highest attrition rate is for the Medical Sergeant MOS (18D); only 18 percent of the students graduate on their first try through the course. Another 45 percent of the students eventually graduate 18D training, making the total graduation rate 63 percent. About 13 percent of Communications Sergeant (18E) trainees fail to graduate from training. Engineer Sergeant (18C) and Weapons Sergeant (18B) MOS have relatively low attrition rates, 16 and 15 percent respectively. In some cases, soldiers who fail training in one MOS are reassigned to a different MOS and proceed with SF training. SWC and ARI have been conducting additional research on attrition from the Q-Course and are studying ways to reduce attrition.

Those individuals who pass the SFQC receive language training. Individuals learn basic communication skills with an emphasis on military terminology and on speaking and listening skills. The languages learned range from those widely known, such as Spanish and French, to those many Americans deem obscure, such as Urdu (spoken in Pakistan) and Tagalog (spoken in the Philippines). Individuals are assigned to languages according to their SF Group assignment, language preference, and scores on the Defense Language Aptitude Battery (DLAB) (Petersen & Al-Haik, 1976). Foreign languages are divided into four difficulty levels, and different cut scores are applied to the DLAB for different languages. For example, Spanish, one of the easier languages for English speaking people to learn, is in the lowest difficulty category.

SF expects continuous training and honing of skills (Fricke, 1990). Once individuals are assigned to a team, they begin informal cross-training. SF soldiers are expected to acquire skills in at least one other specialty area. SF soldiers will also often attend the MOS portion of the SFQC to formally qualify in a second MOS. Cross-training does tend to blur differences between weapons sergeant (18B) and engineer (18C) over long periods of time. However, the skills required for communication

sergeant (18E) degrade without consistent practice, and medical sergeant (18D) skills are highly specialized. Thus, 18E and 18D tend to remain differentiated over the course of their SF careers.

History of SF Selection and Classification Research

Historically, SF selection and classification research dates back to the development of the Army Classification Battery (ACB), the forerunner to today's ASVAB (Berkhouse, 1963). In the early 1960's, Army researchers conducted validity studies to develop a special battery of tests, the SF Selection Battery (Berkhouse, 1963; Berkhouse & Cook, 1961; Berkhouse, Mendelson, & Cook, 1961). The experimental predictor battery contained a variety of noncognitive, self-description inventories as well as a situational judgment test and selected ACB aptitude area composites. Validity evidence led to the selection of four measures for the final battery: (1) the Infantry Aptitude Area composite from the ACB, (2) the Special Forces Suitability Inventory, a noncognitive measure of emotional stability or general psychological adjustment, (3) the Critical Decisions Test, a measure of risk-taking and practical judgment (where a few facts were presented with stringent time limits for deliberation), and (4) the Locations Test, a spatial orientation measure that required orienting oneself according to photographs of terrain. The four measures together yielded a multiple correlation of .63 with the hands-on performance criterion (N=216), .55 when corrected for shrinkage. The Special Forces Selection Battery became operational in 1961. Several noncognitive measures were later designed with the intent of supplementing the Special Forces Selection Battery (Marder & Medland, 1964) but there do not appear to be any citations to research using the newer noncognitive measures.

Another validation study examined the usefulness of the Special Forces Selection Battery and other measures for prediction of officers' academic grades, training performance, and peer ratings (Marder & Medland, 1965). The Special Forces Selection Battery, the Special Forces Qualifying Examination (verbal and math items extracted from other officer selection instruments), and a language aptitude test showed promise for predicting academic grades and to a lesser extent, peer ratings. None of the experimental measures predicted training performance evaluations.

A new experimental battery was developed and assessed in the early '70s (Olmstead, Caviness, Powers, Maxey, & Cleary, 1972). The battery contained the ACB, the Interest Opinion Questionnaire, Life History Inventory, Military Interest Blank, an inventory designed to assess attitudes toward SF activities, the Team-Task Motivation Questionnaire, the Cognitive Test Battery, physical endurance, and a personal information form, several of which had subtests or subscales. Criterion proficiency measures included job knowledge tests, hands-on tests, and self- and peer ratings. Based on stepwise regression results (N=100), researchers identified thirteen tests for the final battery. Several of the best predictors were cognitive; five were from the Cognitive Test Battery, and three were ACB subtests. "Fighter" scores from the life history and military interest instruments as well as a "despair" score, physical endurance, and the team task motivation score made the final battery.

Around the mid-70's the Army terminated use of special batteries for SF selection, relying primarily on the Army Physical Fitness Test, ASVAB GT score, and information available from administrative records such as training experiences for SF selection (Pleban, et al., 1988). These procedures continued for about a decade, until the Special Warfare Center (SWC) tasked ARI to assist in the development of SFAS--a program for screening applicants into SFQC (where attrition was about 50%).

Development of paper-and-pencil and other selected predictors for SFAS involved two major steps.[2] The first step was highly exploratory (Pleban, et al., 1988). The research team, along with the SWC psychologist, determined that predictors should tap three general domains (intelligence, personality, and physical fitness), selected measures for those domains, and compared profiles of SF and non-SF personnel on those measures. They administered the Wonderlic Personnel Test (WPT--a *g* measure), the Jackson Personality Inventory (JPI), the Myers-Briggs Type Indicator (MBTI), and a Biographical Questionnaire (BQ) to soldiers from the 197th Infantry Brigade (N=57), attending the Q-Course (N=339), and currently on A-Teams (N=19). The BQ contained 14 items tapping educational level, component (active-reserve), time in service, rank, specialized training received, MOS, marital status, race, and career plans. Based on practical concerns and comparisons between the samples and between Q-Course students who were successful and unsuccessful in Phase I of the Q-Course, they eliminated the MBTI from further consideration.

The second step was a criterion-related validation study (Pleban, Allentoff, & Thompson, 1989). The WPT, JPI, and BQ were administered to SFQC Phase I candidates. At that time, Phase I was a four-week course focusing on general subjects, teaching, leadership, patrolling, land navigation, and physical conditioning. Phase I status, the criterion, was based on six variables: (1) a map reading written exam, (2) a land navigation field exercise (FTX), (3) a confidence course, (4) a patrolling written exam, (5) a patrolling FTX, and (6) rated performance as a patrol leader. The six scores were noncompensatory; failure to reach the specified cut score on any one variable resulted in termination from SFQC. The best single predictor of Phase I status was WPT (r = .29). Four of the 16 JPI scales correlated significantly with Phase I status. Consequently, the authors recommended use of and further research on the WPT and the four JPI scales.

The BQ items pertaining to specialized prior training were examined. Pleban et al. found that prior Ranger training was related to Phase I status; eighty-four percent of the candidates who had graduated from Ranger school successfully completed Phase I. Reconnaissance and Jungle Warfare training also appeared to be associated with Phase I success. Analyses of the other BQ items (e.g., marital status) were not reported.

---

[2]SFAS includes a number of predictors other than those mentioned here. Some of them are classified, such as the Ruckmarch. Literature reviewed here is limited to reported and unclassified documents.

There have been two recent, relevant investigations of physical fitness requirements for Ranger training and SFAS. Burke and Dyer (1984) collected self-report information about recent Advanced Physical Fitness Test (APFT) scores and administered a physical fitness test consisting of the Harvard Step Test, push-ups, and pull-ups to 906 students in the Ranger Course on the day before training. They found that many of the physical test and APFT scores were related to both graduation from Ranger training and self-reports on the occurrence of nonserious injuries.

Teplitzky (1991) showed that the SFAS Phase I selection boards give considerable weight to the ruckmarch scores in making decisions about candidates. She correlated physical ability components of SFAS, the Ruckmarch and the APFT, with graduation (yes or no) from SFAS. The data were operational (not experimental), and the selection boards had reviewed scores on these events when deciding whether to allow poor performing students to continue. She computed average correlations across three years of SFAS (N=approximately 2,000 per year). The correlations of .25 (APFT) and .43 (Ruckmarch) with SFAS graduation suggest that physical abilities, particularly the Ruckmarch are a major component of the graduation decision.

Recent SF selection and classification research has investigated the usefulness of predictors from the Army's Project A (Peterson, Hough, Dunnette, Rosse, Houston, & Toquam, 1990). Busciglio et al. (1991) found that spatial tests developed in Project A yielded moderate validities for predicting two land navigation criteria collected during SFAS. DeMatteo, White, Teplitzky, & Sachs (1991) administered three scales from the Assessment of Background and Life Experiences (ABLE) to 1023 SF candidates on the third day of SFAS. Approximately 49% of the candidates graduated successfully from SFAS. Scores on the three ABLE scales (Energy Level, Emotional Stability, and Internal Control) were highly skewed, concentrated on the positive end of each scale. Internal Control, which was most severely skewed, failed to demonstrate a significant correlation with SFAS graduation. Energy Level and Emotional Stability yielded low, but significant positive correlations with graduation. Additional analyses suggested that ABLE scores were differentially related to the reasons for attrition. Nearly half of the unselected candidates had withdrawn voluntarily while others were involuntarily cut. The 74 candidates with very low ABLE scores had a disproportionately high rate of voluntary attrition compared to candidates with higher ABLE scores.

The Job Analysis

An analysis of SF jobs was recently conducted (Russell et al., 1994). The primary goal of the job analysis was to provide a solid foundation for the development of selection/classification and criterion measures for MOS in the 18 CMF. The specific objectives for meeting this goal were to describe: (a) the job performance domain and (b) the domain of individual attributes likely to be associated with job performance.

Our approach for achieving these goals (a) coupled task and performance (behavioral) information to form a complete description of the performance domain, (b) relied on individual differences research literature and subject matter expert (SME) input to specify individual attributes, and (c) used professional and subject matter expert

(SME) judgment to link these two domains. Together the attribute and performance information provide the building blocks for the identification of predictors, development of criteria, and conduct of a criterion-related validation study.

An important aspect of this research was the focus on job performance behaviors afforded by the critical incident approach (Flanagan, 1954; Pulakos & Borman, 1985; Smith & Kendall, 1963). Critical incidents define in concrete, behavioral terms the critical performance requirements of the jobs. These behavioral analyses tend to illuminate critical performance components that are a function of motivation, interpersonal skills, communication skills, etc., which are often less likely to emerge in task analyses. The behavioral analyses provided the basic data for constructing job performance rating scales for SF jobs--a major product of the job analysis.

Another important component of the entire project was the inclusion of a subject matter expert panel (SMEP) composed of officers and NCOs from USAJFKSWCS. We briefed the SMEP at key stages of the project--prior to each data collection. SMEP members provided advice on data collection plans, made specific suggestions on forms and materials, and helped us obtain information. Although most of our contact with the SMEP was in formal briefings, several members provided informal feedback on draft materials and sent us articles or other documents.

The job analysis involved five major steps:

(1) Development of workshop materials and logistics,
(2) Administration of workshops to collect critical incidents and task and attribute ratings,
(3) Analysis of critical incident, task, and attribute data,
(4) Development of performance categories and behavior based rating scales, and
(5) Analysis of linkages between attributes and performance categories.

Step 1, Development of workshop materials and logistics, involved: (1) collecting and reviewing documents to form initial lists of job tasks and personal attributes relevant to SF jobs, (2) conducting interviews with SF officers and NCOs to obtain critical incidents and feedback on the initial lists of tasks and attributes, and (3) preparing and pilot testing job analysis data collection procedures.

Step 2 involved a total of 175 NCOs, officers, and warrant officers representing the five major SFG[A]. On average, the participants had 13 years of Army experience and 8 years of SF experience. Seventy-seven percent of participants were currently assigned to A Detachments (B Detachment = 17%, C Detachment = 6%). The participants in Step 2 provided three major types of information:

(1) judgments about individual attributes (such as judgment and decision making ability, non-verbal communication ability, endurance, motivation)
(2) judgments about task areas relevant to SF MOS, and

(3) descriptions of critical incidents (scenarios that describe a situation, an SF individual's behavior in that situation, and the outcome of the individual's actions).

In total, the participants provided 1,767 critical incidents.

In Step 3, the research staff: (1) edited and categorized critical incidents to form performance categories, (2) computed means, standard deviations, and reliability coefficients for the task ratings, and (3) computed means, standard deviations, and reliability ratings for the attribute ratings.

Step 4 involved collecting and analyzing additional information on the performance categories and critical incidents. It had two goals: (1) to get input from SF NCOs, officers, and warrant officers on the performance categories and (2) to obtain judgments about the effectiveness of different behaviors that are represented in the critical incidents from SF NCOs, officers, and warrant officers. One hundred and thirteen soldiers representing the five SFG[A] made the judgments.

We used the results of the analyses of the effectiveness ratings to develop behavior-based performance evaluation scales relevant to each of the performance categories. The names of the performance categories and the major roles of SF jobs that they reflect are listed in Figure 1.

Step 5, Analysis of linkages between attributes and performance categories, involved collecting judgments from NCOs, officers, and researchers familiar with SF jobs about the importance of each attribute for effective performance in each of the job performance categories.

Objectives of the Roadmap

The primary goal of the Roadmap project was to extend the job analysis results by:

- identifying useful, readily available measures for important SF attributes and performance dimensions,
- ensuring that the selection system would meet predicted future needs as well as current requirements, and
- suggesting projects for the development and validation of measures.

The identification of measures for important SF attributes and performance dimensions was accomplished through a series of expert judgment exercises linking measures to attribute and performance constructs. Chapter II describes this process in detail.

Role		Performance Category(ies)
Teacher	A.	Teaching Others
Diplomat	B.	Building and Maintaining Effective Relationships with Indigenous Populations
	C.	Handling Interpersonal Situations
	D.	Using and Enhancing Own Language Skills
Professional	E.	Contributing to the Team Effort and Morale
	F.	Showing Initiative and Extra Effort
	G.	Displaying Honesty and Integrity
Mission Planner	H.	Planning and Preparing for Missions
	I.	Decision Making
Soldier	J.	Confronting Physical and Environmental Challenges
	K.	Navigating in the Field
	L.	Troubleshooting and Solving Problems
	M.	Being Safety Conscious
	N.	Administering First Aid and Treating Casualties
	O.	Managing Administrative Duties
Weapons Expert	P.	Operating and Maintaining Direct-Fire Weapons
	Q.	Employing Indirect-Fire Weapons and Techniques
Engineer	R.	Employing Demolitions Techniques
	S.	Constructing for Mission-Related Requirements
Communi-cations	T.	Following Communication Policies and Procedures
	U.	Assembling and Operating Commo Equipment
Medic	V.	Evaluating and Treating Medical Conditions and Injuries
	W.	Determining and Administering Medications and Dosages
	X.	Ensuring Standards of Health-Related Facilities, Conditions, and Procedures
Leader	Y.	Showing Consideration for Subordinates
	Z.	Providing Direction

Figure 1.
SF Roles and Performance Categories Based on Performance Examples

To ensure that the selection system would meet future and current needs, we discussed future mission changes with SF decision makers and observed SFAS and part of the Q-Course. We then integrated the job analysis results, interviews with SF and SWC decision-makers, field observations of SF selection and training, and judgments of experts in selection and classification to form the Roadmap. The final Roadmap and its development are described in Chapter III.

CHAPTER II
DEVELOPMENT OF EXPERT JUDGMENT LINKAGES

The overall goal of this project was to develop an agenda--a Roadmap--for Special Forces (SF) selection and classification research. Key questions for the Roadmap project were: *What predictors should be included in future validation projects to enhance SF selection and classification?* and *What criterion measures are available, how well do they cover the performance domain, and what measures could be developed?* Moreover, one of the specific objectives of the project was to identify tests, exercises, and other measures (i.e., predictors and criteria) likely to be useful to USAJFKSWCS.

Validation involves assembling evidence about relationships between predictor measures, attribute definitions, job descriptors, and criterion measures, (Society for Industrial Psychology, 1987). Often expert judges are one source of evidence in this network (see Figure 2 for one depiction of such judgments). The appropriate set of judges to use depends on the purpose of the expert judgment exercise and the way in which descriptors and mesures are defined. For example, using expert judgments in validation may involve assessing the relationship between job descriptors and attribute descriptors and between job descriptors and criterion measures, judgments that requires knowledge of job descriptors, attribute descriptors, and criterion measures. Either job experts or psychologists may be appropriately used for these kinds of judgments, depending on how much prior job knowledge psychologists have and how the judgment task is defined. Assessing the relationship between attribute descriptors and predictor measures, a judgment that requires knowledge of the psychometric and individual difference literature bases, is probably best completed by psychologists.

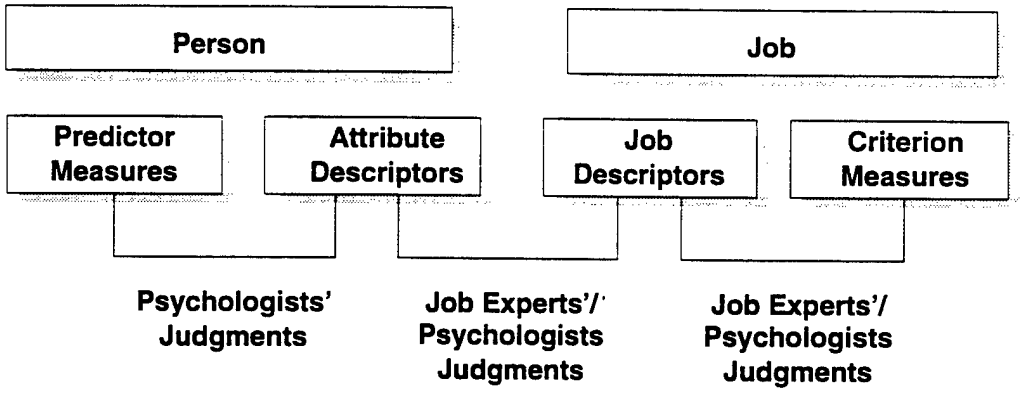

Figure 2
The Roles of Expert Judgments in Validation Paradigms

Research suggests that such experts can make reliable, accurate judgments of these kinds. Studies typically report reliabilities in the .80-.90 range for experts' judgments of relationships among constructs (Peterson & Bownas, 1982; Wing, Peterson, & Hoffman, 1984). Indeed, experts can make reasonably accurate estimates of empirical validities (Schmidt, Hunter, Croll, & McKenzie, 1983).

The best validation efforts assemble information from multiple sources, and that is the foundation of our approach to identifying predictor and criterion measures for SF selection and classification research. The first step in that direction, taken in the SF Job Analysis project, was to obtain expert judgments to establish relationships between attribute and job descriptors. These judgments are represented by Box B in Figure 3 which shows the linkages between SF attributes and performance categories made by USAJFKSWCS officers and NCOs and the research team. Boxes A, C, and D represent three types of expert judgments made in the current project. Box C represents psychologists' judgments about the extent to which tests and scales measure attributes. Similarly, Box D shows a linkage between criterion measures and performance categories. Lastly, the matrix at the top (Box A) shows the mapping of SF primary missions against SF performance categories. This mapping has implications for selection and training. Once projections are made about what missions will be emphasized in the future, the linkage of missions to performance categories allows the targeting of selection practices toward specific attributes critical for those missions or the targeting of training to specific job performance categories.

In this chapter, we describe the preparatory activities, the procedures, and the results of the expert judgment exercises we conducted to continue the process of assembling validation information. Specifically, we describe data collection and results for these three types of expert judgments:

(1) the predictor expert judgment exercise (Box C),
(2) the criterion expert judgment exercise (Box D), and
(3) the mission performance expert judgment exercise (Box A).

Collection of Expert Judgments about Predictor Measures

The predictor expert judgment task was designed to identify tests, exercises, and scales that would be most useful in measuring the individual attributes necessary to perform SF jobs. We developed and conducted the expert judgment task in the following four steps:

- Gather written documents and interview researchers
- Prepare materials for the exercise
- Collect expert judgments
- Analyze the data and interpret the results

Gather Information and Conduct Interviews. We interviewed ARI personnel to identify ARI tests, measures, and scales that were likely to be useful for SF selection. For each interview, we first explained the purpose of the Roadmap project and explained the role of the researchers in it. We used an unstructured format to elicit information from researchers about the research they had either participated in, were familiar with, or had heard about, that might prove useful for measuring SF job attributes. In some cases, we were already familiar with the research of the interviewee, and therefore probed for certain descriptive and psychometric information about specific predictors. We also asked researchers to provide copies of technical or other reports that we could

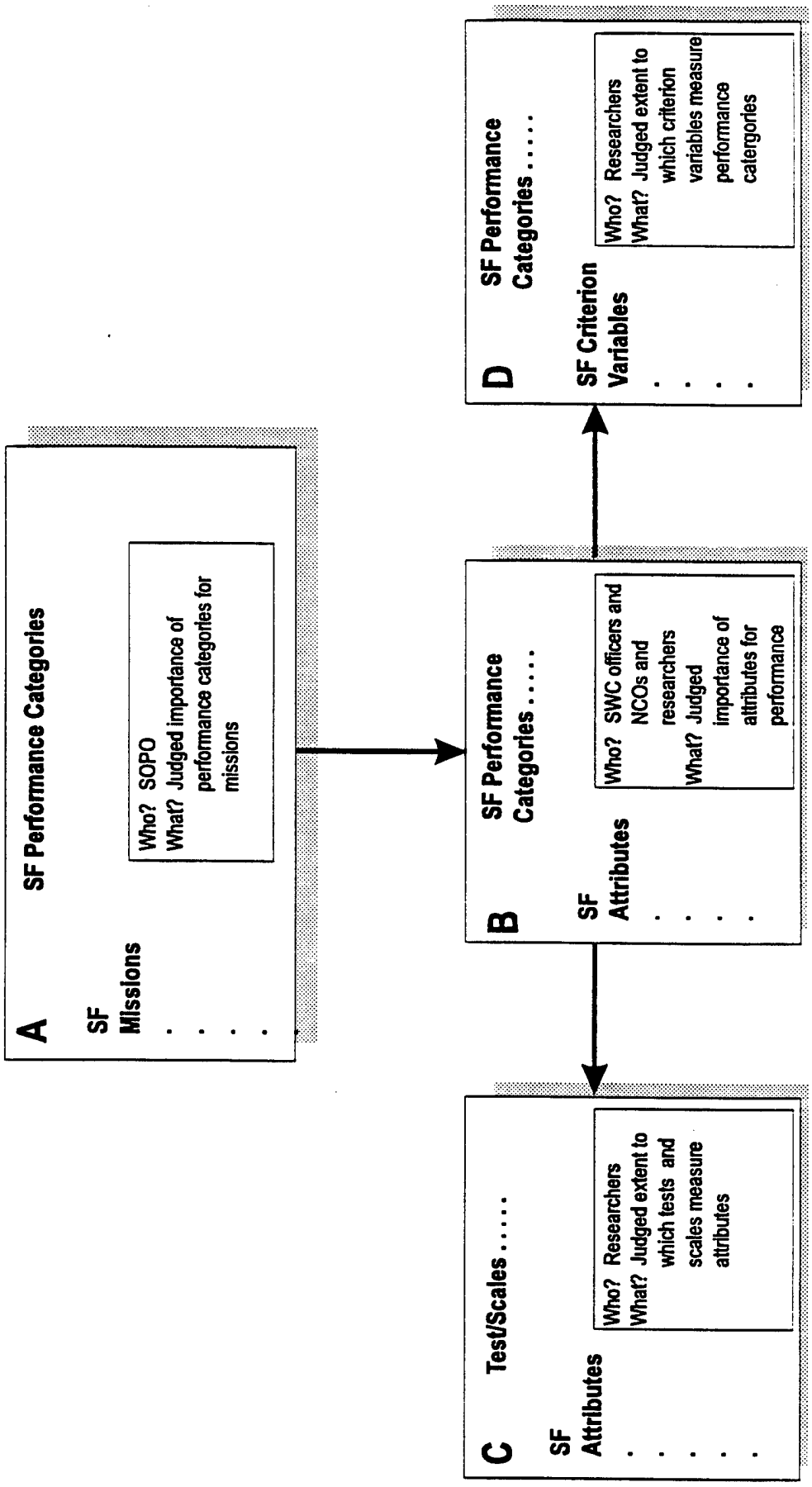

Figure 3. SF Selection and Classification Research Planning Data Bases

use to locate specific descriptive details for the variety of predictors. A critical question asked of each interviewee was whether they could think of any additional predictors or any additional projects that we should investigate.

We also reviewed codebooks for the SFAS database to identify variables that would be useful as predictors in a validation study. We found too many (very detailed) variables to include as individual measures in a research plan, so we looked for meaningful ways to aggregate the data. We conducted a factor analysis and formed composites based on the results. The composites we formed were for the variety of physical tests taken during SFAS (e.g., physical endurance and physical fitness) and for performance on military orienteering and situation reaction exercises. A summary of these analyses is included in Appendix A.

At the end of the information gathering stage, we had a set of potential predictor measures that could be administered prior to and during SFAS. Many of the instruments have been used in the Army previously and have accumulated validity evidence (e.g., Armed Services Vocational Aptitude Battery), yet other measures are still in the development stages and require further validity research (e.g., measures of wisdom). Figure 4 lists all the measures used in the predictor expert judgment exercises: cognitive, non-cognitive (i.e., biographical, personality, and interest measures), physical and psychomotor, and performance-based measures.

Prepare Materials. We consolidated all of the information from the interviews, technical reports, test manuals, and other materials, by preparing a predictor description form (an example is provided in Figure 5). We listed the following information for each predictor (except where the information was not available from any source):

- Short description of test
- Psychometric properties:
  - Scoring
  - Correlations among constructs
  - Correlations with other measures
  - Reliability
  - Subgroup differences
  - Fakability
  - Validity Evidence

We assigned labels to each predictor to indicate its stage of development:

- Proposed -- instruments under consideration for development for Special Forces.
- Experimental -- instruments that have been developed and field tested but are not currently in use.
- Operational -- instruments that are currently in use.
- Published -- instruments developed and controlled by a test publisher.

Measures Included in the Expert Judgment Exercises

Cognitive Measures

C1. ASVAB General Science	C2. ASVAB Arithmetic Reasoning	C3. ASVAB Word Knowledge
C4. ASVAB Paragraph Comp.	C5. ASVAB Number Operations	C6. ASVAB Coding Speed
C7. ASVAB Auto/Shop	C8. ASVAB Math Knowledge	C9. ASVAB Mechanical Comp.
C10. ASVAB Electronics Inf.	C11. Wonderlic	C12. Project A Map Test
C13. Assembling Objects	C14. TABE Language	C15. TABE Reading
C16. TABE Mathematics	C17. ABLE Vocabulary	C18. ABLE Reading
C19. ABLE Language	C20. ABLE Mathematics	C21. Perc. Speed/Accuracy
C22. Target Identification	C23. Number Memory	C24. Short Term Memory
C25. Defense Language Apt. B.	C26. Problem Construction	C27. Information Encoding
C28. Category Search and Spec.	C29. Category Combination	C30. Wisdom 1
C31. Problem Solving Skills	C32. Solution Characteristics	C33. Problem Evaluation
C34. Plan and Implement	C35. Leadership Knowledge	C36. Wisdom 2
C37. Leadership Problems	C38. Army Radio Code Test	C39. Superdit-Sound Mem.
C40. Superdit-Sound Mem Plus	C41. Superdit-Motor Prog.	

Biographical, Personality, and Interest Measures

NC1. ABI Academic Performance	NC2. ABI Formal Leadership	NC3. ABI Ruggedness
NC4. ABI Mechanical Activities	NC5. ABI Work Experience	NC6. ABI Home Economics
NC7. ABI Nondelinquency	NC8. ABI Team Sports/Group Orient.	NC9. ABI Work Skills
NC10. ABI Family/Community	NC11. ABI Cross Cultural	NC12. RBI Cognition Under Stress
NC13. RBI Mature Team Commit.	NC14. RBI Self Esteem	NC15. RBI Combat Motivation
NC16. RBI Need for Achievement	NC17. RBI Outdoor Orientation	NC18. RBI Physical Endurance
NC19. RBI Physical Strength	NC20. RBI Object Belief	NC21. FCABLE-Work Orientation
NC22. FCABLE-Dominance	NC23. FCABLE-Dependability	NC24. FCABLE-Agreeableness
NC25. FCABLE-Emotional Stability	NC26. AVOICE-Rugged/Outdoors	NC27. AVOICE-Audiovisual Arts
NC28. AVOICE-Interpersonal	NC29. AVOICE-Skilled Technical	NC30. AVOICE-Administrative
NC31. AVOICE-Food Service	NC32. AVOICE-Protective Services	NC33. AVOICE-Structural/Machines
NC34. JOB-High Expectations	NC35. JOB-Routine	NC36. JOB-Autonomy
NC37. JCQ-SF Scale	NC38. JCQ-Weapons	NC39. JCQ-Engineer
NC40. JCQ-Commo	NC41. JCQ-Medic	NC42. Organizational Identity
NC43. Occupational Stress A.I.	NC44. Social Intelligence Biodata	

Physical and Psychomotor Measures

PS1. Project A Target Tracking 1	PS2. Project A Target Tracking 2	PS3. Project A Target Shoot
PS4. Cannon Shoot Test	PS5. Army Physical Fitness Test	PS6. SFAS Physical Endurance
PS7. SFAS Physical Fitness	PS8. SFAS Swim Test	

Measures of Conventional Army Proficiency and Performance-Based Measures

P1. Self Development Test	P2. SFAS Military Orienteering	P3. SFAS Peer Rankings
P4. SFAS Situation Reaction Exer.	P5. # of Awards and Certific.	P6. # of Article 15s and Flags
P7. Promotion Rate	P8. Work and Training Portfolio	P9. Language Training Record
P10. Army Wide Perf. Ratings	P11. Teaching Role Play	P12. Cultural Adaptability Role Play
P13. Structural Interview	P14. Situational Judgment Test	P15. Hi Fi Simulation
P16. NCO Role Plays		

ASVAB=Armed Services Vocational Aptitude Battery; TABE=Test of Adult Basic Education; ABLE=Adult Basic Learning Examination; ABI=Army Biodata Inventory; RBI=Ranger Biodata Inventory; FCABLE=Forced Choice Assessment of Background and Life Experiences; AVOICE= Army Vocational Interest Career Examination; JOB=Job Orientation Blank; JCQ=Job Compatibility Questionnaire; JRTC=Joint-Readiness Training Center

Figure 4
Measures Included in the Expert Judgment Exercise

12. Experimental: Project A MAP Test (MP)

Construct Measured:

Spatial Orientation--This test measures one's ability to "appreciate one's location relative to land marks in the environment" (Peterson et. al., 1987).

Short Description of Test:

Subjects are given a map with various landmarks such as a campsite, a forest, a lake, and so on. Several items refer to each map, within each item, subjects are provided compass directions by indicating the direction from one landmark to another (e.g., "the forest is North of the campsite") and they are informed of their present location on the map. Given this information, the subject must determine which direction to take to reach another landmark.

Number of Items: 20 Time Limit: 12 minutes
Apparatus: Paper and pencil

Psychometrics:

Scoring: The score is the total number of correct answers.

Correlations with other constructs: The Map test correlates with Assembling Objects $r = .50, .52$; Object Rotation $r = .39, .42$; MAZE $r = .44, .42$; Orientation $r = .53, .54$; Reasoning $r = .52, .51$ all N's = 9332, 6941 respectively. Factor analytic research including the Map Test suggests that it represents a first order Orientation factor and loads highly on a second order general spatial factor. Busciglio & Teplitzky (1994) found the MAP correlated with MAZE $r = .48$; Orientation $r = .52$; and the Wonderlic $r = .66$ with N = 232.

Subgroup Differences: Whites tend to score 1 SD higher than blacks (large sample effect sizes range from .98 to 1.08). Whites score .4 to .6 SD higher than Hispanics. Males tend to score higher than females with effect sizes (standardized mean differences) between .28 to .30.

Reliability: Cronbach's alpha: .88 (N = 6754); .89 (N = 9332); .90 (N = 290). Test-Retest Reliability: .78 (N = 499); .84 (N = 97).

Practice Effects: Test performance on spatial ability tests is to some degree malleable; test scores improve with practice (Lohman, 1988). However, the gains are not substantially larger than those observed for tests of other abilities (Russell et al., 1994). There is also some evidence that gains from practice are larger for speeded tests than for power tests (Dunnette, Corpe, & Toquam, 1987). Gains from practice on the Map test have been low in two studies. With a one week interval between testing sessions (N = 100), subjects' scores went up .08 sd from testing 1 to testing 2 (Peterson, 1987). With one month between testing sessions (N = 473) subjects' scores again went up .08 sd from testing 1 to testing 2 (Toquam, Peterson, Rosse, Ashworth, Hanson, & Hallam, 1986)

Validity Evidence: In Project A, McHenry et al. (1990) combined six Project A spatial tests to form one composite score. The spatial score yielded modest incremental validity (beyond that afforded by the ASVAB) for predicting technical proficiency in Army enlisted MOS and hands-on performance. Similar results were obtained for a longitudinal validation sample.

Busciglio & Teplitzky (1994) found that the MAP test is predictive of performance in the SFQC land navigation exercises, adding unique variance over other variables in predicting success in this exercise. They found the MAP test to be the best single predictor of first time land navigation success ($F = 7.97$, $p < .01$).

Busciglio & Teplitzky (1990) also found the MAP test to predict success in SFAS military orienteering exercises. They found that high MAP scores are related to higher ratings and less time needed to complete the military orienteering exercises. Ratings on the Early (Task I Day and Night and Task II Day) and Later (Task II Night, Task III and Task IV) Orienteering scores are correlated with the Map Test .31, .23; $p < .0001$ respectively. The time for the Early Orienteering scores is related to the Map test -.24 $p < .0001$.

Figure 5
Example Predictor Description Form

Many of the predictors have multiple subtests (e.g., the Armed Services Vocational Aptitude Battery (ASVAB)) which we described and rated separately. Due to the number of predictors involved, we grouped the predictor description forms into the following four categories:

(1) Cognitive measures (e.g., Wonderlic)

(2) Noncognitive measures (e.g., biographical, interest, temperament, and preference measures)

(3) Psychomotor/physical measures (e.g., Project A Target tracking tests; SFAS physical fitness tests)

(4) Performance measures (e.g., awards and certificates; SFAS performance-peer rankings and situational reaction exercises).

The complete packets of these predictors are included in Appendices B - E. Appendix F contains the bibliography for these predictor descriptions.

We designated expert judges by first identifying individuals with test and measurement experience at ARI, HumRRO, and AIR. We sent out a very brief questionnaire asking these people to indicate their familiarity with the four predictor categories listed above. Based on their responses, we designated specific predictor subsets for each of 20 expert judges to rate. Therefore, each judge completed one to four subsets of the four predictor categories.

We developed an expert judgment task to gather ratings of how well each Proposed, Experimental, Operational, and Published predictor measured each SF attribute. Each packet included four documents (see Appendix G):

- Background--a statement, describing the project and the expert judgment task in very general terms,

- Supporting Information--the Executive Summary of the Job Analysis project and a listing of several important acronyms.

- Background Information Form--a form which experts completed to describe their level of knowledge about each of the four areas of measures,

- Instructions--instructions for making extent-of-measurement judgments

We chose a rating scale that has been used successfully in similar projects to collect expert judgments (Peterson, Owens-Kurtz, Hoffman, Arabian, & Whetzel, 1990).

We asked respondents to consider *To what extent does each predictor measure each attribute?* They used the following scale to quantify their judgments:

0----------1----------2----------3----------4----------5----------6----------7----------8

This attribute is not at all measurable by the predictor (it is almost useless)	This attribute is measured partly by the predictor (it is of some use)	This attribute is entirely measured by the predictor (it is highly useful)

We developed instructions specifying that experts should scan a set of predictors, then read all of the attribute definitions. They were to rate each predictor according to the extent it measures each attribute. We asked that raters consider the following factors in completing the ratings: *What does the predictor measure? How well does the predictor measure the construct? What is the predictor's track record?*

We developed an additional exercise for expert judges to complete after they had finished all of their ratings (which ranged from one to four sets). We instructed each judge to identify a "best bet" predictor -- the predictor they felt would be the best measure for each attribute on the list. In making "best bet" judgments, we asked that judges integrate information about the extent of measurement along with other factors (e.g., subgroup differences) that they thought were important in the use of the test or scale.

Collect Data. Individuals were chosen to participate in the exercise based on their experience with the different types of predictors and their knowledge of SF jobs. Individuals had studied these types of predictors in undergraduate and graduate courses, and often had performed work using these measures, at times supervising the work of others in these areas. Many of the experts had also taught others in this task, and some have published articles or books in the areas of their expertise.

A total of 20 expert judges completed ratings to link the SF attributes with the predictor measures. Five individuals who were very familiar with the project and the predictors were asked to complete all of the rating exercises; each of the other fifteen completed ratings for one to three of the predictor subsets.

All but one of the rating exercises required several hours to complete. The cognitive measures required a commitment of six hours, the noncognitive measures exercise required six hours; the psychomotor/physical measures exercise required only one hour and the performance measures exercise took approximately three to four hours to complete. The participants received information at the end of September (1994) and were asked to complete the exercises and return them within a two-week period.

Analyze and Interpret Data. We prepared and checked the data for analysis. We calculated means and standard deviations for the ratings of each predictor for each attribute. Means for cognitive, noncognitive, physical/psychomotor, and performance measures appear in Tables 1 through 4 respectively.

Table 1
Mean Extent of Measurement Ratings for Cognitive Instruments

	Attributes									
	Judgment	Planning	Adaptability	Creativity	Auditory Ability	Mechanical Ability	Spatial Ability	Perceptual Ability	Basic Math	Advanced Math
Test or Scale	1	2	3	4	5	6	7	8	9	10
1. ASVAB GS	3.92	2.25	2.00	2.25	0.17	3.17	2.67	2.08	2.58	2.75
2. ASVAB AR	4.83	2.83	1.67	2.25	0.50	3.33	2.92	1.75	7.42	6.17
3. ASVAB WK	4.00	2.25	1.75	2.83	0.42	1.67	1.92	1.33	1.92	2.00
4. ASVAB PC	5.25	2.92	2.58	2.92	0.33	1.92	2.00	1.75	1.92	2.50
5. ASVAB NO	2.67	1.42	1.25	1.50	0.75	1.50	1.67	2.75	5.92	4.75
6. ASVAB CS	2.33	1.33	1.08	1.08	0.75	1.33	2.00	3.50	2.25	1.42
7. ASVAB AS	2.92	1.92	1.58	2.08	0.33	6.42	3.75	3.33	2.92	2.33
8. ASVAB MK	3.92	2.58	1.75	2.67	0.50	2.67	3.42	2.25	6.58	7.50
9. ASVAB MC	4.50	2.58	2.25	3.25	0.50	7.33	5.08	3.58	3.83	3.75
10. ASVAB EI	3.92	2.25	2.08	2.25	0.50	5.25	3.50	2.83	3.75	3.50
11. Wonderlic	5.92	3.83	3.58	3.75	0.58	2.75	3.25	2.33	4.83	4.58
12. Project A Map Test	4.00	3.08	2.50	2.08	0.17	2.83	6.92	4.58	2.75	2.42
13. Project A Assembling Objects	4.50	3.00	2.67	3.33	0.17	4.00	7.17	4.50	2.83	2.67
14. TABE Language Composite	3.50	2.00	1.42	1.75	0.67	1.67	1.67	1.42	1.83	1.67
15. TABE Reading Composite	4.00	2.17	1.67	1.83	0.42	1.50	1.67	1.42	1.50	1.50
16. TABE Mathematics Composite	4.08	2.33	1.25	1.75	0.50	2.42	2.58	1.92	7.17	6.25
17. ABLE Vocabulary Composite	2.75	2.00	1.50	1.58	0.58	1.42	1.67	1.33	1.58	1.50
18. ABLE Reading Composite	3.42	2.17	1.58	2.08	0.50	1.75	2.00	1.58	1.50	1.50
19. ABLE Language Composite	3.00	1.92	1.50	1.42	0.58	1.33	1.92	1.75	1.17	1.42
20. ABLE Mathematics Composite	4.17	2.33	1.75	2.08	0.33	2.67	2.67	1.83	6.83	6.17
21. Project A PSA	2.42	1.17	1.83	1.08	0.67	2.17	3.08	6.25	1.92	1.50
22. Project A TID	2.67	1.08	1.67	1.17	0.50	2.00	4.25	6.00	1.42	1.25
23. Project A NM	2.75	1.67	1.58	0.92	0.67	1.58	2.00	3.42	5.67	2.67
24. Project A STM	2.00	1.33	1.17	0.75	1.00	1.50	2.00	3.42	1.33	1.08
25. DLAB	3.17	1.75	2.50	2.25	3.50	1.33	1.75	1.92	1.83	2.17
26. Problem Construction	4.67	4.08	4.17	4.67	0.00	0.75	0.83	0.67	0.67	1.25
27. Information Encoding	4.83	3.83	3.25	4.17	0.00	0.75	0.75	1.42	0.67	1.00
28. Category Search and Specification	5.08	3.17	3.33	3.50	0.00	0.50	0.58	1.25	0.50	0.67
29. Category Combination	4.83	2.92	3.25	4.75	0.17	0.67	1.25	1.17	0.42	0.50
30. Wisdom 1	4.25	2.17	2.58	3.75	0.17	0.33	0.75	1.08	0.33	0.50
31. Problem Solving Skills	5.33	4.42	3.42	3.75	0.17	0.50	0.67	1.08	0.50	0.58
32. Solution Characteristics	5.17	4.83	3.75	4.33	0.17	0.33	0.67	1.00	0.42	0.58
33. Problem Evaluation	4.83	3.83	3.08	3.33	0.08	0.42	0.75	1.25	0.42	0.50
34. Planning and Implementation	4.92	5.50	3.33	3.83	0.00	0.33	0.83	1.17	0.42	0.50
35. Leadership Knowledge	3.92	3.75	2.67	3.08	0.08	0.50	0.75	1.00	0.42	0.50
36. Wisdom 2	5.00	3.33	3.17	3.50	0.08	0.42	0.75	1.00	0.50	0.50
37. Leadership Problems Inventory	4.83	4.92	3.00	2.83	0.08	0.58	0.83	0.83	0.42	0.42
38. Army Radio Code Test	1.00	0.42	1.00	0.33	5.50	1.08	0.58	2.42	0.92	0.42
39. Superdit-Sound Memory	1.17	0.25	0.67	0.25	6.92	0.50	0.50	3.00	1.00	0.50
40. Superdit-Sound Memory with Interference	1.50	0.42	1.08	0.25	6.50	0.50	0.58	3.17	1.17	0.58
41. Superdit-Motor Programming	1.50	0.67	1.00	0.25	5.17	0.83	0.58	3.17	1.00	0.50

Note 0-2 = attribute not measured at all by predictor 3-5 = attribute measured partly by predictor 6-8 = attribute entirely measured by predictor

Table 1
Mean Extent of Measurement Ratings for Cognitive Instruments

	Attributes									
	Reading Ability	Writing Ability	Language Ability	Communication	Non-Verbal Comm	Persuasiveness	Cultural/Interperson	Maturity	Autonomy	Team Playership
Test or Scale	11	12	13	14	15	16	17	18	19	20
1. ASVAB GS	4.08	2.17	1.75	1.42	0.83	0.42	0.83	1.08	0.75	0.25
2. ASVAB AR	4.00	1.83	2.00	1.17	0.67	0.42	0.42	0.67	0.75	0.25
3. ASVAB WK	6.25	4.83	4.25	3.75	1.42	2.17	1.17	1.08	0.92	0.50
4. ASVAB PC	7.33	4.92	4.17	4.00	1.67	1.92	1.50	1.33	0.92	0.42
5. ASVAB NO	2.00	1.00	1.08	0.67	0.33	0.08	0.25	0.75	0.67	0.17
6. ASVAB CS	2.50	1.33	1.33	0.67	0.33	0.17	0.25	0.67	0.58	0.17
7. ASVAB AS	2.42	1.00	0.75	0.50	0.33	0.17	0.17	0.75	1.08	0.33
8. ASVAB MK	3.42	1.67	1.50	0.92	0.58	0.42	0.42	1.00	1.00	0.33
9. ASVAB MC	2.83	1.33	1.25	1.00	0.67	0.17	0.33	0.92	1.08	0.33
10. ASVAB EI	3.08	1.17	1.08	0.75	0.42	0.08	0.17	0.75	0.67	0.25
11. Wonderlic	5.33	3.67	3.08	3.25	2.08	2.42	2.25	1.67	1.67	1.08
12. Project A Map Test	2.58	1.08	1.08	0.92	0.83	0.42	0.67	0.75	1.00	0.50
13. Project A Assembling Objects	2.17	1.17	1.25	1.00	0.67	0.42	0.58	0.58	0.75	0.50
14. TABE Language Composite	5.00	6.33	3.42	3.25	1.08	1.58	0.67	0.58	0.42	0.42
15. TABE Reading Composite	7.25	4.58	3.25	2.92	1.25	1.67	1.00	0.75	0.50	0.42
16. TABE Mathematics Composite	2.67	1.33	1.50	1.00	0.58	0.67	0.50	0.50	0.42	0.33
17. ABLE Vocabulary Composite	5.08	4.08	3.00	3.33	1.00	1.42	0.75	0.58	0.42	0.33
18. ABLE Reading Composite	7.25	4.33	3.17	2.83	1.25	1.50	0.92	0.67	0.50	0.42
19. ABLE Language Composite	4.75	6.00	3.58	3.00	1.08	1.42	0.83	0.67	0.42	0.42
20. ABLE Mathematics Composite	3.00	2.00	1.58	1.17	0.50	0.58	0.50	0.50	0.42	0.33
21. Project A PSA	1.67	0.75	1.00	0.58	0.58	0.25	0.42	0.17	0.33	0.25
22. Project A TID	1.25	0.58	0.92	0.42	0.33	0.25	0.25	0.17	0.33	0.25
23. Project A NM	1.25	0.58	0.83	0.42	0.33	0.17	0.08	0.08	0.33	0.33
24. Project A STM	1.25	0.75	1.08	0.75	0.33	0.42	0.25	0.08	0.25	0.25
25. DLAB	3.67	2.75	6.67	3.08	1.33	1.25	1.42	0.50	0.42	0.42
26. Problem Construction	2.08	1.58	1.17	1.17	0.83	1.50	1.50	1.25	1.67	1.08
27. Information Encoding	2.92	2.67	1.33	1.58	0.67	1.50	1.50	1.25	1.08	1.00
28. Category Search and Specification	2.83	2.33	1.33	2.08	0.83	2.42	1.75	1.67	1.08	1.00
29. Category Combination	2.58	1.67	1.42	1.08	0.67	1.25	1.25	0.92	0.83	0.42
30. Wisdom 1	2.75	1.25	1.08	1.50	0.58	1.92	2.17	2.92	1.08	1.50
31. Problem Solving Skills	2.33	1.25	1.08	1.17	0.58	1.42	1.17	2.08	1.25	1.58
32. Solution Characteristics	2.33	1.67	1.08	1.00	0.58	1.92	1.92	2.00	1.33	1.25
33. Problem Evaluation	2.50	1.00	0.92	0.92	0.58	1.50	1.42	1.42	1.00	1.25
34. Planning and Implementation	2.50	1.08	1.00	1.00	0.58	1.08	1.08	1.25	1.42	1.25
35. Leadership Knowledge	1.92	1.00	0.83	1.08	0.67	0.75	0.58	0.92	0.58	1.08
36. Wisdom 2	2.17	1.50	0.92	1.33	0.50	2.00	1.92	2.58	1.25	2.17
37. Leadership Problems Inventory	2.50	1.08	0.92	1.00	0.58	1.17	1.17	1.58	1.25	1.33
38. Army Radio Code Test	1.33	0.25	2.08	0.67	0.58	0.25	0.33	0.33	0.25	0.25
39. Superdit-Sound Memory	0.42	0.17	1.58	0.83	0.67	0.25	0.25	0.25	0.25	0.25
40. Superdit-Sound Memory with Interference	0.58	0.17	1.75	0.92	0.83	0.25	0.42	0.25	0.25	0.25
41. Superdit-Motor Programming	0.33	0.25	1.25	0.83	0.75	0.25	0.25	0.25	0.25	0.25

Note 0-2 = attribute not measured at all by predictor 3-5 = attribute measured partly by predictor 6-8 = attribute entirely measured by predictor

Table 1
Mean Extent of Measurement Ratings for Cognitive Instruments

Test or Scale	Attributes: Dependability 21	Initiative 22	Perseverance 23	Moral Courage 24	Motivating Others 25	Supervising 26	Swimming 27	Physical Flexibility 28	Physical Strength 29	Physical Endurance 30
1. ASVAB GS	0.33	1.00	0.50	0.33	0.33	1.08	0.00	0.08	0.00	0.08
2. ASVAB AR	0.33	0.83	0.33	0.17	0.25	0.75	0.00	0.00	0.00	0.00
3. ASVAB WK	0.42	0.92	0.42	0.25	0.75	1.08	0.00	0.00	0.00	0.00
4. ASVAB PC	0.50	0.92	0.42	0.25	0.83	1.25	0.00	0.00	0.00	0.00
5. ASVAB NO	0.25	0.92	0.75	0.17	0.42	0.58	0.00	0.17	0.00	0.08
6. ASVAB CS	0.25	0.67	0.75	0.17	0.25	0.42	0.00	0.25	0.00	0.08
7. ASVAB AS	0.25	0.83	0.50	0.17	0.17	0.25	0.00	0.67	0.58	0.25
8. ASVAB MK	0.42	1.00	0.58	0.17	0.33	0.83	0.00	0.08	0.00	0.00
9. ASVAB MC	0.25	0.75	0.42	0.17	0.33	0.50	0.08	0.25	0.25	0.17
10. ASVAB EI	0.33	0.67	0.50	0.67	0.25	0.25	0.00	0.17	0.08	0.08
11. Wonderlic	1.08	1.67	1.25	0.75	1.42	2.25	0.00	0.08	0.00	0.08
12. Project A Map Test	0.42	0.75	0.67	0.33	0.42	0.75	0.08	0.17	0.08	0.42
13. Project A Assembling Objects	0.42	0.58	1.00	0.33	0.42	0.67	0.00	0.25	0.17	0.42
14. TABE Language Composite	0.50	0.58	0.50	0.25	0.33	0.92	0.00	0.00	0.00	0.00
15. TABE Reading Composite	0.50	0.58	0.42	0.25	0.42	1.00	0.00	0.00	0.00	0.00
16. TABE Mathematics Composite	0.42	0.58	0.50	0.25	0.33	0.75	0.00	0.00	0.00	0.00
17. ABLE Vocabulary Composite	0.42	0.58	0.42	0.25	0.42	1.17	0.00	0.00	0.00	0.00
18. ABLE Reading Composite	0.50	0.75	0.42	0.25	0.58	1.33	0.00	0.00	0.00	0.00
19. ABLE Language Composite	0.50	0.67	0.42	0.25	0.42	1.08	0.00	0.00	0.00	0.00
20. ABLE Mathematics Composite	0.42	0.67	0.42	0.25	0.33	1.00	0.00	0.00	0.00	0.00
21. Project A PSA	0.42	0.50	0.58	0.17	0.25	0.58	0.08	0.25	0.00	0.08
22. Project A TID	0.33	0.42	0.42	0.17	0.17	0.33	0.08	0.17	0.00	0.00
23. Project A NM	0.42	0.42	0.58	0.17	0.25	0.50	0.00	0.00	0.00	0.00
24. Project A STM	0.17	0.33	0.25	0.08	0.17	0.33	0.00	0.00	0.00	0.08
25. DLAB	0.33	0.67	0.42	0.25	0.42	0.83	0.00	0.00	0.00	0.00
26. Problem Construction	1.00	1.83	1.25	0.67	1.00	2.08	0.00	0.00	0.00	0.00
27. Information Encoding	0.83	1.42	1.17	0.83	1.17	2.42	0.00	0.00	0.00	0.00
28. Category Search and Specification	1.17	1.42	1.25	1.75	1.33	2.83	0.00	0.00	0.00	0.00
29. Category Combination	0.33	0.75	1.17	0.17	0.50	1.75	0.00	0.00	0.00	0.00
30. Wisdom 1	1.17	0.83	0.83	2.83	1.42	2.17	0.00	0.00	0.00	0.00
31. Problem Solving Skills	0.92	1.33	0.67	1.67	0.83	2.17	0.00	0.00	0.00	0.00
32. Solution Characteristics	1.17	1.33	0.75	1.33	1.25	3.00	0.00	0.00	0.00	0.00
33. Problem Evaluation	0.92	1.00	0.75	1.50	0.92	2.42	0.00	0.00	0.00	0.00
34. Planning and Implementation	0.83	1.25	0.92	0.75	1.33	2.75	0.00	0.00	0.00	0.00
35. Leadership Knowledge	0.50	0.83	0.83	0.33	1.00	2.08	0.00	0.00	0.00	0.00
36. Wisdom 2	1.33	1.25	1.00	2.83	2.00	2.67	0.00	0.00	0.00	0.00
37. Leadership Problems Inventory	0.92	1.42	0.92	1.00	1.33	3.75	0.00	0.08	0.08	0.25
38. Army Radio Code Test	0.25	0.33	0.83	0.08	0.17	0.17	0.00	0.08	0.00	0.00
39. Superdit-Sound Memory	0.25	0.33	0.92	0.08	0.17	0.17	0.00	0.00	0.00	0.00
40. Superdit-Sound Memory with Interference	0.33	0.33	1.08	0.08	0.17	0.25	0.00	0.00	0.00	0.08
41. Superdit-Motor Programming	0.33	0.33	0.83	0.08	0.17	0.17	0.08	0.25	0.00	0.08

Note 0-2 = attribute not measured at all by predictor 3-5 = attribute measured partly by predictor 6-8 = attribute entirely measured by predictor

Table 1
Mean Extent of Measurement Ratings for Cognitive Instruments

Test or Scale	Attributes: Psychomotor Ability 31	Interest in Adventure 32	Interest in Skilled Trade 33	Interest in Other Cultures 34	Interest in People 35	Enterprising Interests 36	Leadership 37	Achievement and Effort 38	Personal Discipline 39	Physical Fitness and Military Bearing 40
1. ASVAB GS	0.17	1.25	1.58	0.83	0.42	0.50	1.67	1.75	0.50	0.33
2. ASVAB AR	0.50	0.08	1.58	0.33	0.25	0.58	1.67	1.83	0.58	0.33
3. ASVAB WK	0.17	0.17	0.58	0.92	1.00	1.17	2.00	1.83	0.67	0.25
4. ASVAB PC	0.17	0.25	0.67	1.17	1.00	1.08	2.17	1.92	0.75	0.25
5. ASVAB NO	1.00	0.17	1.33	0.25	0.25	0.42	0.92	1.33	0.58	0.42
6. ASVAB CS	1.25	0.08	1.08	0.33	0.25	0.50	0.75	1.08	0.75	0.25
7. ASVAB AS	2.00	2.50	5.50	0.42	0.25	1.08	0.67	1.58	0.83	0.83
8. ASVAB MK	0.50	0.17	2.00	0.25	0.17	0.67	1.50	1.83	0.75	0.42
9. ASVAB MC	1.00	1.58	4.25	0.25	0.25	0.67	1.00	1.50	0.58	0.42
10. ASVAB EI	0.83	1.08	4.17	0.25	0.17	0.67	0.83	1.42	0.50	0.83
11. Wonderlic	0.92	0.42	1.67	1.58	1.25	1.92	3.42	2.50	1.25	0.58
12. Project A Map Test	1.42	2.33	1.67	0.50	0.33	0.58	1.33	1.42	0.67	0.67
13. Project A Assembling Objects	2.00	1.67	2.67	0.42	0.50	0.42	0.92	1.50	0.58	0.33
14. TABE Language Composite	0.17	0.08	0.58	0.50	0.42	0.42	1.67	1.33	0.58	0.33
15. TABE Reading Composite	0.17	0.08	0.58	0.50	0.50	0.42	1.50	1.25	0.50	0.33
16. TABE Mathematics Composite	0.25	0.17	1.75	0.33	0.17	0.33	1.25	1.42	0.67	0.25
17. ABLE Vocabulary Composite	0.17	0.08	0.42	0.67	0.67	0.67	1.92	1.25	0.50	0.33
18. ABLE Reading Composite	0.17	0.08	0.75	0.75	0.67	0.50	1.92	1.42	0.50	0.33
19. ABLE Language Composite	0.17	0.17	0.42	0.50	0.33	0.33	1.67	1.42	0.58	0.33
20. ABLE Mathematics Composite	0.42	0.33	1.75	0.33	0.33	0.58	1.50	1.25	0.92	0.67
21. Project A PSA	2.67	0.42	1.25	0.08	0.08	0.17	0.83	1.00	0.75	0.42
22. Project A TID	2.58	0.75	1.25	0.08	0.08	0.17	0.75	0.92	0.83	0.42
23. Project A NM	1.50	0.17	1.00	0.08	0.00	0.33	0.75	1.25	0.75	0.33
24. Project A STM	1.08	0.17	0.83	0.25	0.25	0.42	1.08	1.00	0.50	0.33
25. DLAB	0.42	0.17	0.42	1.67	1.00	0.67	1.08	1.42	0.50	0.25
26. Problem Construction	0.08	0.50	0.75	1.08	0.83	1.25	3.42	2.50	1.50	0.42
27. Information Encoding	0.08	0.42	0.58	0.92	1.17	1.33	3.00	2.08	0.92	0.42
28. Category Search and Specification	0.08	0.33	0.50	0.83	1.58	1.50	3.42	2.00	1.58	0.58
29. Category Combination	0.08	0.42	0.42	1.08	1.00	1.17	2.67	1.75	1.00	0.33
30. Wisdom 1	0.00	0.33	0.25	1.33	2.08	1.33	2.83	1.50	2.17	0.25
31. Problem Solving Skills	0.00	0.25	0.25	1.08	1.50	1.33	3.50	1.92	1.08	0.33
32. Solution Characteristics	0.00	0.25	0.25	1.08	1.50	1.75	4.25	2.00	1.50	0.25
33. Problem Evaluation	0.00	0.25	0.25	1.33	1.42	1.50	3.17	1.67	1.25	0.33
34. Planning and Implementation	0.00	0.25	0.25	0.83	1.17	1.75	3.42	1.83	1.00	0.42
35. Leadership Knowledge	0.00	0.17	0.25	0.50	1.08	1.33	3.92	1.25	0.67	0.33
36. Wisdom 2	0.00	0.25	0.25	1.58	2.17	1.83	4.58	2.17	2.17	0.67
37. Leadership Problems Inventory	0.00	0.33	0.33	0.83	1.50	1.33	4.50	2.08	1.92	0.92
38. Army Radio Code Test	1.50	0.25	0.75	0.25	0.25	0.33	0.42	0.67	0.83	0.33
39. Superdit-Sound Memory	1.25	0.25	0.58	0.25	0.17	0.25	0.25	0.67	0.58	0.33
40. Superdit-Sound Memory with Interference	1.17	0.25	0.50	0.25	0.17	0.25	0.33	0.83	0.75	0.33
41. Superdit-Motor Programming	2.75	0.25	0.67	0.25	0.17	0.25	0.33	0.75	0.75	0.58

Note 0-2 = attribute not measured at all by predictor 3-5 = attribute measured partly by predictor 6-8 = attribute entirely measured by predictor

Table 1
Mean Extent of Measurement Ratings for Cognitive Instruments

Test or Scale	General Soldiering Proficiency 41	Infantry Proficiency 42	Combat Engineer Proficiency 43	Other Combat MOS Proficiency 44	Radio Teletype Operator Proficiency 45	Medical Care Specialist Proficiency 46	Other Non-Combat MOS Proficiency 47
	Attributes						
1. ASVAB GS	3.17	3.17	3.58	3.08	3.67	4.75	3.33
2. ASVAB AR	3.08	3.67	4.33	3.42	3.67	3.67	3.33
3. ASVAB WK	2.75	3.50	3.42	3.25	3.58	3.75	3.50
4. ASVAB PC	3.25	3.75	3.42	3.42	3.75	4.00	3.58
5. ASVAB NO	2.08	2.50	3.00	2.42	2.92	2.42	2.25
6. ASVAB CS	1.92	2.17	1.92	1.83	3.67	1.58	1.83
7. ASVAB AS	3.25	3.92	4.25	3.67	3.42	2.75	2.75
8. ASVAB MK	2.58	3.08	4.25	3.17	3.08	3.08	3.00
9. ASVAB MC	3.17	4.08	5.00	3.92	3.50	2.50	2.83
10. ASVAB EI	3.17	3.42	4.17	3.42	5.33	2.17	2.58
11. Wonderlic	3.75	3.83	4.17	4.00	4.50	4.50	4.17
12. Project A Map Test	3.58	3.75	3.58	3.33	2.67	2.08	2.42
13. Project A Assembling Objects	3.08	3.17	3.58	3.00	2.67	2.08	2.58
14. TABE Language Composite	2.58	2.17	2.17	2.17	3.08	2.67	2.33
15. TABE Reading Composite	2.75	2.42	2.58	2.58	3.25	3.17	2.83
16. TABE Mathematics Composite	3.33	3.17	3.92	3.25	3.08	3.08	3.00
17. ABLE Vocabulary Composite	2.83	2.67	2.83	2.50	2.92	2.75	2.50
18. ABLE Reading Composite	3.00	2.83	2.92	2.67	3.00	3.42	3.08
19. ABLE Language Composite	2.33	2.17	2.50	2.33	3.17	2.67	2.42
20. ABLE Mathematics Composite	3.25	3.17	4.00	3.25	3.17	3.08	3.17
21. Project A PSA	2.83	3.17	2.92	2.50	3.75	2.25	2.25
22. Project A TID	3.25	4.00	3.67	3.58	3.08	2.17	2.25
23. Project A NM	2.67	2.75	2.92	2.67	3.50	2.42	2.33
24. Project A STM	2.25	2.42	2.33	2.33	3.42	2.25	2.17
25. DLAB	2.00	2.08	1.92	2.00	3.25	2.00	2.17
26. Problem Construction	2.42	2.50	2.67	2.25	2.17	2.75	2.50
27. Information Encoding	2.17	2.17	2.33	2.17	1.75	2.42	2.25
28. Category Search and Specification	2.17	1.92	2.08	2.17	2.17	2.75	2.50
29. Category Combination	2.17	2.08	2.17	2.08	1.75	2.25	2.08
30. Wisdom 1	2.00	1.58	1.58	1.58	1.42	1.75	1.58
31. Problem Solving Skills	2.58	1.92	2.00	1.83	1.67	1.83	1.83
32. Solution Characteristics	2.25	2.00	2.17	1.92	2.17	2.25	1.92
33. Problem Evaluation	2.00	1.83	1.83	1.83	1.92	2.08	2.00
34. Planning and Implementation	2.25	2.17	2.25	2.08	1.75	2.00	1.92
35. Leadership Knowledge	1.92	1.83	1.50	1.58	1.42	1.50	1.75
36. Wisdom 2	1.92	1.75	1.50	1.50	1.25	1.83	1.58
37. Leadership Problems Inventory	2.50	2.25	2.00	2.00	1.92	2.33	2.17
38. Army Radio Code Test	1.75	1.67	1.50	1.58	4.83	1.00	1.42
39. Superdit-Sound Memory	1.33	1.25	1.25	1.42	4.25	0.83	1.17
40. Superdit-Sound Memory with Interference	1.33	1.25	1.25	1.42	4.42	0.92	1.17
41. Superdit-Motor Programming	1.42	1.42	1.42	1.58	4.58	1.08	1.17

Note 0-2 = attribute not measured at all by predictor 3-5 = attribute measured partly by predictor 6-8 = attribute entirely measured by predictor

Table 2
Mean Extent of Measurement Ratings for Noncognitive Instruments

	Attributes									
	Judgment	Planning	Adaptability	Creativity	Auditory Ability	Mechanical Ability	Spatial Ability	Perceptual Ability	Basic Math	Advanced Math
Test or Scale	1	2	3	4	5	6	7	8	9	10
1. ABI Academic Performance	2.91	2.18	1.73	1.45	0.73	0.82	1.09	0.73	3.27	3.27
2. ABI Formal Leadership	1.82	2.18	1.55	1.00	0.45	0.09	0.09	0.27	1.09	1.00
3. ABI Ruggedness	0.55	0.45	1.36	0.55	0.09	1.09	1.45	1.18	0.18	0.18
4. ABI Mechanical Activities	1.18	0.73	0.91	1.00	0.09	5.55	2.73	1.55	1.27	0.82
5. ABI Work Experience	1.00	1.18	0.91	0.55	0.36	0.27	0.36	0.64	0.82	0.55
6. ABI Home Economics	0.91	1.09	1.55	1.00	0.27	0.55	0.18	0.27	0.36	0.09
7. ABI Nondelinquency	2.00	0.91	0.73	0.36	0.09	0.09	0.09	0.09	0.73	0.55
8. ABI Team Sports/Group Orientation	0.45	0.36	1.27	0.55	0.09	0.45	0.27	0.45	0.18	0.09
9. ABI Work Skills	0.73	0.91	1.45	0.73	0.09	2.18	1.55	0.73	0.82	0.64
10. ABI Family/Community	0.64	0.45	3.55	0.73	0.18	0.36	0.36	0.55	0.27	0.18
11. ABI Cross Cultural Sensitivity	1.55	0.45	2.64	0.82	0.09	0.09	0.09	0.91	0.18	0.18
12. RBI Cognition Under Stress	4.36	2.82	3.36	1.55	0.09	0.09	0.09	0.45	0.36	0.27
13. RBI Mature Team Commitment	1.91	1.45	2.55	1.18	0.09	0.09	0.09	0.27	0.18	0.09
14. RBI Self Esteem	1.36	0.82	1.73	0.73	0.09	0.09	0.09	0.09	0.18	0.09
15. RBI Combat Motivation	0.73	0.36	1.18	0.27	0.09	0.27	0.18	0.55	0.18	0.09
16. RBI Need for Achievement	0.82	1.18	1.00	1.00	0.09	0.09	0.09	0.09	0.55	0.45
17. RBI Outdoor Orientation	0.27	0.27	1.00	0.45	0.36	0.27	0.82	1.09	0.18	0.09
18. RBI Physical Endurance	0.27	0.27	0.27	0.18	0.09	0.18	0.36	0.27	0.18	0.09
19. RBI Physical Strength	0.27	0.18	0.18	0.09	0.09	0.18	0.18	0.18	0.09	0.09
20. RBI Object Belief	0.73	0.18	0.45	0.18	0.09	0.09	0.09	0.09	0.09	0.09
21. FCABLE-Work Orientation	0.91	1.55	0.55	0.64	0.09	0.09	0.09	0.18	0.27	0.18
22. FCABLE-Dominance	0.82	0.91	0.55	0.45	0.09	0.09	0.09	0.09	0.18	0.09
23. FCABLE-Dependability	2.36	3.00	0.91	0.45	0.09	0.09	0.09	0.27	0.18	0.09
24. FCABLE-Agreeableness	0.45	0.45	2.18	0.36	0.09	0.09	0.09	0.09	0.18	0.09
25. FCABLE-Emotional Stability	2.36	1.27	2.64	0.64	0.09	0.09	0.09	0.09	0.18	0.09
26. AVOICE-Rugged/Outdoors	0.45	0.27	0.73	0.36	0.27	1.27	1.27	0.45	0.18	0.09
27. AVOICE-Audiovisual Arts	0.64	0.91	0.64	2.36	1.27	1.45	2.09	1.27	1.27	0.73
28. AVOICE-Interpersonal	0.91	0.91	0.82	0.27	0.09	0.18	0.18	0.36	0.45	0.36
29. AVOICE-Skilled Technical	0.91	0.73	0.64	0.73	0.27	2.18	1.27	1.00	2.73	2.27
30. AVOICE-Administrative	0.82	1.09	0.27	0.18	0.09	0.36	0.36	0.64	1.00	0.36
31. AVOICE-Food Service	0.36	0.36	0.36	0.27	0.09	0.18	0.18	0.18	0.45	0.18
32. AVOICE-Protective Services	1.09	0.45	0.55	0.36	0.27	0.36	0.36	1.27	0.55	0.27
33. AVOICE-Structural/Machines	0.55	0.55	0.45	0.27	0.09	3.09	1.55	0.64	0.73	0.27
34. JOB-High Expectations	0.73	0.91	0.55	0.45	0.09	0.27	0.09	0.09	0.18	0.09
35. JOB-Routine	0.18	0.45	0.73	1.09	0.09	0.09	0.18	0.27	0.27	0.09
36. JOB-Autonomy	1.09	0.91	1.09	1.27	0.09	0.09	0.09	0.09	0.36	0.09
37. JCQ-SF Scale	1.36	1.09	1.91	0.55	0.09	0.45	0.55	0.55	0.45	0.09
38. JCQ-Weapons	0.55	0.64	0.73	0.27	0.18	2.09	0.55	0.64	0.18	0.09
39. JCQ-Engineer	0.73	1.09	0.91	0.82	0.09	1.82	1.00	0.73	1.27	1.09
40. JCQ-Commo	0.55	0.82	0.82	0.73	1.09	1.27	0.27	0.45	0.73	0.27
41. JCQ-Medic	1.09	1.09	0.82	0.64	0.64	0.36	0.18	0.45	1.18	0.45
42. Organizational Identity	0.18	0.18	0.82	0.09	0.09	0.09	0.09	0.09	0.09	0.09
43. OSAI	1.27	0.64	3.00	0.45	0.09	0.09	0.09	0.18	0.09	0.09
44. SI Biodata	1.45	0.55	1.64	0.27	0.09	0.09	0.09	0.55	0.09	0.09

Note 0-2 = attribute not measured at all by predictor 3-5 = attribute measured partly by predictor 6-8 = attribute entirely measured by predictor

Table 2
Mean Extent of Measurement Ratings for Noncognitive Instruments

	Attributes									
Test or Scale	Reading Ability 11	Writing Ability 12	Language Ability 13	Communication 14	Non-Verbal Comm 15	Persuasiveness 16	Cultural/Interperson 17	Maturity 18	Autonomy 19	Team Playership 20
1. ABI Academic Performance	3.64	3.64	2.45	2.45	1.09	0.82	1.45	1.45	1.27	0.18
2. ABI Formal Leadership	1.27	1.18	0.64	3.18	2.55	3.64	2.09	2.18	2.00	3.64
3. ABI Ruggedness	0.18	0.18	0.18	0.55	0.55	0.45	0.36	0.82	2.00	1.45
4. ABI Mechanical Activities	0.36	0.27	0.18	0.45	0.45	0.55	0.36	0.64	1.00	0.45
5. ABI Work Experience	0.82	0.55	0.18	1.27	0.91	0.91	1.18	2.73	2.45	2.00
6. ABI Home Economics	0.36	0.27	0.18	0.91	0.91	0.91	0.91	1.82	3.64	1.36
7. ABI Nondelinquency	1.00	1.00	0.82	0.73	0.73	1.00	1.18	3.91	1.18	1.64
8. ABI Team Sports/Group Orientation	0.18	0.18	0.45	1.09	1.18	1.64	1.55	0.91	0.91	5.09
9. ABI Work Skills	0.73	0.73	0.55	0.91	0.55	0.82	1.27	1.91	1.55	1.91
10. ABI Family/Community	0.27	0.27	0.18	0.55	1.09	1.27	3.27	3.00	2.45	2.27
11. ABI Cross Cultural Sensitivity	0.27	0.36	0.82	1.27	2.36	2.55	6.09	2.73	1.27	1.45
12. RBI Cognition Under Stress	0.64	0.36	0.45	0.45	0.36	0.45	0.64	3.09	1.45	0.64
13. RBI Mature Team Commitment	0.18	0.18	0.36	1.18	1.09	1.82	2.09	2.91	1.00	5.55
14. RBI Self Esteem	0.18	0.18	0.09	0.91	0.55	1.82	1.91	3.36	4.09	2.73
15. RBI Combat Motivation	0.18	0.18	0.09	0.45	0.36	0.64	0.36	1.27	1.45	1.09
16. RBI Need for Achievement	0.45	0.45	0.36	0.27	0.09	0.45	0.45	1.27	2.45	0.91
17. RBI Outdoor Orientation	0.18	0.18	0.09	0.09	0.18	0.18	0.18	0.27	1.18	0.36
18. RBI Physical Endurance	0.18	0.18	0.09	0.09	0.18	0.09	0.09	0.27	0.64	0.27
19. RBI Physical Strength	0.09	0.09	0.09	0.09	0.09	0.45	0.27	0.09	0.36	0.09
20. RBI Object Belief	0.09	0.09	0.09	0.82	0.55	2.91	2.64	2.18	1.18	3.91
21. FCABLE-Work Orientation	0.45	0.45	0.18	0.82	0.64	1.09	0.64	2.00	2.09	2.27
22. FCABLE-Dominance	0.18	0.18	0.09	0.82	0.82	4.55	1.64	1.00	2.09	1.64
23. FCABLE-Dependability	0.18	0.18	0.09	0.36	0.36	1.09	1.09	3.73	1.91	2.27
24. FCABLE-Agreeableness	0.18	0.18	0.09	0.82	1.09	3.55	3.27	2.73	0.82	4.45
25. FCABLE-Emotional Stability	0.18	0.18	0.09	0.82	0.55	2.09	2.55	5.45	1.91	1.82
26. AVOICE-Rugged/Outdoors	0.18	0.18	0.09	0.09	0.09	0.18	0.36	0.27	2.00	0.91
27. AVOICE-Audiovisual Arts	0.55	0.73	1.00	1.00	0.82	0.36	0.73	0.18	0.45	0.09
28. AVOICE-Interpersonal	0.45	0.45	0.64	1.45	1.73	2.09	1.91	1.64	0.82	2.27
29. AVOICE-Skilled Technical	1.18	0.82	0.36	0.55	0.36	0.45	0.36	0.55	0.91	0.45
30. AVOICE-Administrative	1.00	0.73	0.27	0.64	0.64	0.27	0.27	0.55	0.64	0.55
31. AVOICE-Food Service	0.45	0.36	0.09	0.45	0.45	0.18	0.27	0.27	0.55	0.55
32. AVOICE-Protective Services	0.55	0.55	0.27	1.27	1.55	1.18	0.82	1.55	1.64	1.55
33. AVOICE-Structural/Machines	0.36	0.27	0.18	0.18	0.27	0.18	0.18	0.18	0.55	0.36
34. JOB-High Expectations	0.18	0.18	0.09	0.55	0.73	1.36	1.36	1.36	1.09	2.18
35. JOB-Routine	0.18	0.09	0.09	0.09	0.09	0.73	0.73	0.09	0.91	0.55
36. JOB-Autonomy	0.18	0.09	0.09	0.27	0.18	0.27	0.55	1.36	6.00	1.64
37. JCQ-SF Scale	0.18	0.18	0.55	1.45	1.18	1.55	2.73	1.36	1.64	2.64
38. JCQ-Weapons	0.18	0.18	0.09	0.27	0.36	0.27	0.27	0.36	0.36	0.36
39. JCQ-Engineer	0.55	0.45	0.09	0.36	0.36	0.18	0.09	0.36	0.45	0.45
40. JCQ-Commo	0.73	0.36	0.45	0.82	0.45	0.27	0.45	0.36	0.36	0.27
41. JCQ-Medic	1.18	1.09	0.36	0.91	1.09	0.45	0.91	1.45	0.82	1.00
42. Organizational Identity	0.09	0.09	0.09	0.09	0.09	0.64	0.82	0.73	0.18	3.00
43. OSAI	0.09	0.09	0.09	0.09	0.09	0.36	1.45	2.45	0.82	0.36
44. SI Biodata	0.09	0.09	0.18	1.09	1.91	3.27	3.64	2.18	0.36	2.36

Note 0-2 = attribute not measured at all by predictor 3-5 = attribute measured partly by predictor 6-8 = attribute entirely measured by predictor

Table 2
Mean Extent of Measurement Ratings for Noncognitive Instruments

	Attributes									
Test or Scale	Dependability 21	Initiative 22	Perseverance 23	Moral Courage 24	Motivating Others 25	Supervising 26	Swimming 27	Physical Flexibility 28	Physical Strength 29	Physical Endurance 30
1. ABI Academic Performance	1.09	2.18	2.27	0.27	0.36	1.18	0.18	0.00	0.00	0.00
2. ABI Formal Leadership	2.73	3.00	2.27	0.91	4.00	3.45	0.00	0.00	0.00	0.00
3. ABI Ruggedness	1.09	1.45	2.36	0.64	0.82	0.73	0.64	2.91	3.09	3.64
4. ABI Mechanical Activities	0.82	0.73	1.00	0.27	0.55	0.45	0.27	1.18	1.18	0.91
5. ABI Work Experience	2.91	3.18	2.45	0.73	0.64	1.00	0.09	0.45	0.36	0.64
6. ABI Home Economics	1.73	1.64	1.55	0.82	1.00	1.09	0.09	0.45	0.36	0.55
7. ABI Nondelinquency	4.45	1.64	1.55	3.36	1.00	0.91	0.09	0.09	0.09	0.09
8. ABI Team Sports/Group Orientation	2.09	1.27	1.64	0.73	1.73	0.82	0.73	1.64	1.64	1.64
9. ABI Work Skills	2.36	1.73	1.45	0.45	0.82	0.82	0.09	0.45	0.55	0.55
10. ABI Family/Community	1.73	0.91	2.00	1.36	1.09	1.00	0.27	0.18	0.18	0.18
11. ABI Cross Cultural Sensitivity	1.55	0.73	0.45	1.18	1.09	1.00	0.09	0.09	0.09	0.09
12. RBI Cognition Under Stress	1.36	0.55	1.09	1.18	0.55	0.91	0.09	0.09	0.09	0.36
13. RBI Mature Team Commitment	3.36	2.45	1.73	2.27	3.18	2.00	0.36	0.09	0.09	0.09
14. RBI Self Esteem	2.27	2.55	1.91	2.27	2.00	1.82	0.45	0.09	0.09	0.09
15. RBI Combat Motivation	1.36	1.73	1.45	2.18	1.00	0.91	0.27	0.55	1.18	1.18
16. RBI Need for Achievement	2.00	5.64	5.00	1.18	1.91	1.09	0.09	0.09	0.09	0.09
17. RBI Outdoor Orientation	0.36	0.36	1.00	0.36	0.18	0.18	0.82	1.91	1.91	2.09
18. RBI Physical Endurance	0.64	0.64	3.36	0.36	0.36	0.27	1.36	2.91	4.36	6.55
19. RBI Physical Strength	0.18	0.27	1.09	0.09	0.27	0.09	0.18	2.18	6.36	3.36
20. RBI Object Belief	1.64	0.64	0.36	2.00	2.45	1.82	0.09	0.09	0.09	0.09
21. FCABLE-Work Orientation	3.00	3.73	5.09	1.36	1.55	1.45	0.09	0.27	0.09	0.36
22. FCABLE-Dominance	1.09	2.09	1.36	0.82	3.64	3.27	0.09	0.18	0.18	0.18
23. FCABLE-Dependability	6.73	1.91	2.55	2.45	1.18	1.36	0.09	0.09	0.09	0.09
24. FCABLE-Agreeableness	1.91	1.00	1.09	0.82	2.18	1.82	0.09	0.09	0.09	0.09
25. FCABLE-Emotional Stability	2.82	1.18	1.73	2.36	0.82	1.27	0.45	0.09	0.09	0.36
26. AVOICE-Rugged/Outdoors	0.64	0.91	1.73	0.27	0.27	0.09	0.64	1.73	2.18	2.73
27. AVOICE-Audiovisual Arts	0.09	0.18	0.18	0.09	0.09	0.09	0.09	0.18	0.18	0.18
28. AVOICE-Interpersonal	1.09	0.82	0.36	1.00	1.73	2.00	0.09	0.09	0.09	0.09
29. AVOICE-Skilled Technical	0.73	1.18	0.91	0.27	0.36	0.27	0.09	0.18	0.09	0.09
30. AVOICE-Administrative	1.18	0.27	0.45	0.18	0.18	1.09	0.09	0.18	0.18	0.27
31. AVOICE-Food Service	0.64	0.27	0.45	0.18	0.18	0.18	0.09	0.18	0.27	0.36
32. AVOICE-Protective Services	1.55	0.82	1.18	1.64	0.82	0.73	0.18	1.27	1.73	1.45
33. AVOICE-Structural/Machines	0.45	0.27	0.45	0.18	0.18	0.18	0.09	0.55	1.09	1.00
34. JOB-High Expectations	1.73	2.18	1.73	1.00	1.64	1.91	0.09	0.09	0.09	0.09
35. JOB-Routine	1.27	1.00	1.00	0.09	0.09	0.45	0.09	0.09	0.09	0.09
36. JOB-Autonomy	0.64	1.91	1.18	0.64	0.09	0.09	0.36	0.09	0.09	0.09
37. JCQ-SF Scale	1.73	2.18	2.18	1.18	1.91	1.64	0.09	0.36	0.64	0.82
38. JCQ-Weapons	0.18	0.55	0.36	0.27	0.36	0.36	0.09	0.64	0.64	0.55
39. JCQ-Engineer	0.27	0.55	0.36	0.18	0.18	0.27	0.09	0.73	0.64	0.55
40. JCQ-Commo	0.45	0.36	0.27	0.18	0.18	0.18	0.09	0.36	0.27	0.27
41. JCQ-Medic	1.27	0.45	0.55	1.00	0.55	0.18	0.09	0.09	0.18	0.09
42. Organizational Identity	2.18	1.36	1.45	1.73	1.09	1.09	0.09	0.09	0.09	0.09
43. OSAI	0.73	0.55	1.55	0.73	0.09	0.18	0.36	0.27	0.27	0.64
44. SI Biodata	0.73	0.36	0.36	0.36	1.73	2.09	0.09	0.09	0.09	0.09

Note 0-2 = attribute not measured at all by predictor 3-5 = attribute measured partly by predictor 6-8 = attribute entirely measured by predictor

Table 2
Mean Extent of Measurement Ratings for Noncognitive Instruments

Test or Scale	Attributes: Psychomotor Ability 31	Interest in Adventure 32	Interest in Skilled Trade 33	Interest in Other Cultures 34	Interest in People 35	Enterprising Interests 36	Leadership 37	Achievement and Effort 38	Personal Discipline 39	Physical Fitness and Military Bearing 40
1. ABI Academic Performance	0.18	0.36	0.45	1.45	0.73	0.82	2.64	3.64	3.00	0.55
2. ABI Formal Leadership	0.09	0.27	0.18	1.09	2.82	4.09	5.36	4.00	2.55	1.00
3. ABI Ruggedness	1.45	6.55	1.64	0.82	0.82	0.64	0.91	1.27	1.73	3.91
4. ABI Mechanical Activities	2.27	1.55	5.36	0.36	0.45	0.36	0.64	0.91	0.91	0.82
5. ABI Work Experience	0.45	0.27	0.36	0.45	0.82	1.64	1.82	2.73	3.00	0.91
6. ABI Home Economics	0.82	0.27	0.64	0.36	1.09	0.55	0.64	0.82	1.91	0.55
7. ABI Nondelinquency	0.09	0.18	0.18	0.55	0.82	0.45	2.27	2.36	5.27	1.55
8. ABI Team Sports/Group Orientation	1.36	2.27	0.64	0.45	2.36	1.45	2.27	1.91	1.27	2.00
9. ABI Work Skills	0.82	2.18	3.18	0.91	1.09	1.18	1.36	2.18	1.55	0.82
10. ABI Family/Community	0.18	0.64	0.36	1.00	1.27	0.36	1.18	0.82	2.18	0.82
11. ABI Cross Cultural Sensitivity	0.09	0.73	0.18	5.91	4.00	0.45	1.18	0.45	0.73	0.36
12. RBI Cognition Under Stress	0.18	0.27	0.27	0.27	0.27	0.55	2.27	0.73	1.27	0.64
13. RBI Mature Team Commitment	0.09	0.18	0.27	0.73	2.55	2.00	4.36	2.64	2.09	0.36
14. RBI Self Esteem	0.09	0.18	0.18	0.18	0.55	1.27	2.82	2.00	1.55	1.18
15. RBI Combat Motivation	0.64	2.36	0.73	0.18	0.36	0.45	1.27	0.82	1.36	2.00
16. RBI Need for Achievement	0.09	0.18	0.27	0.27	0.45	3.00	2.82	5.82	1.91	1.00
17. RBI Outdoor Orientation	1.36	6.82	1.18	0.27	0.27	0.27	0.27	0.27	0.55	1.91
18. RBI Physical Endurance	1.73	2.18	0.27	0.18	0.18	0.18	0.27	0.91	1.27	3.73
19. RBI Physical Strength	1.27	1.36	0.73	0.27	0.27	0.45	0.36	0.36	0.36	3.00
20. RBI Object Belief	0.09	0.18	0.18	0.82	4.00	1.27	2.73	0.82	0.45	0.36
21. FCABLE-Work Orientation	0.09	0.18	0.27	0.18	0.36	1.64	2.27	4.64	3.27	0.91
22. FCABLE-Dominance	0.18	0.27	0.27	0.36	2.27	5.00	5.36	3.09	1.09	1.09
23. FCABLE-Dependability	0.09	0.18	0.27	0.36	0.82	0.27	1.91	3.00	3.82	0.91
24. FCABLE-Agreeableness	0.09	0.18	0.18	1.09	3.55	1.00	2.27	1.64	1.27	0.82
25. FCABLE-Emotional Stability	0.27	0.27	0.27	0.36	0.64	0.45	2.09	1.73	3.09	1.09
26. AVOICE-Rugged/Outdoors	1.64	6.64	0.91	0.27	0.27	0.27	0.55	0.73	1.09	2.45
27. AVOICE-Audiovisual Arts	1.36	0.09	2.09	0.45	0.18	0.18	0.18	0.45	0.27	0.18
28. AVOICE-Interpersonal	0.09	0.18	0.27	1.18	3.91	1.55	3.27	1.73	0.91	0.45
29. AVOICE-Skilled Technical	0.64	0.36	4.73	0.27	0.27	0.55	0.27	0.64	0.45	0.27
30. AVOICE-Administrative	0.73	0.09	0.45	0.18	0.36	0.55	0.18	0.36	0.82	0.36
31. AVOICE-Food Service	0.27	0.18	0.27	0.27	0.73	0.27	0.27	0.27	0.36	0.27
32. AVOICE-Protective Services	0.91	2.36	0.91	0.64	2.18	1.00	1.55	0.82	1.00	1.36
33. AVOICE-Structural/Machines	0.91	1.45	4.73	0.27	0.36	0.45	0.27	0.36	0.27	0.27
34. JOB-High Expectations	0.09	0.18	0.64	0.36	2.27	2.45	2.64	3.09	1.36	0.55
35. JOB-Routine	0.09	0.09	0.82	0.64	0.09	1.00	1.00	1.09	0.82	0.36
36. JOB-Autonomy	0.09	0.27	0.36	0.18	0.45	0.55	0.64	1.09	2.00	0.36
37. JCQ-SF Scale	0.55	2.45	1.91	3.36	2.55	1.91	1.64	2.27	1.27	1.27
38. JCQ-Weapons	0.82	1.64	1.45	0.18	0.27	0.18	0.36	0.36	0.45	0.55
39. JCQ-Engineer	0.82	1.18	2.09	0.18	0.18	0.27	0.18	0.27	0.36	0.45
40. JCQ-Commo	0.82	0.73	2.00	0.18	0.45	0.36	0.45	0.45	0.45	0.45
41. JCQ-Medic	0.18	0.55	0.55	0.45	1.55	0.73	0.55	0.55	0.45	0.55
42. Organizational Identity	0.09	0.64	0.36	0.55	1.00	0.55	1.73	2.09	2.00	1.09
43. OSAI	0.27	0.09	0.09	0.09	0.09	0.18	1.64	0.91	0.82	0.55
44. SI Biodata	0.09	0.09	0.09	1.64	3.18	1.36	3.09	1.00	0.82	0.64

Note 0-2 = attribute not measured at all by predictor 3-5 = attribute measured partly by predictor 6-8 = attribute entirely measured by predictor

Table 2
Mean Extent of Measurement Ratings for Noncognitive Instruments

Test or Scale	General Soldiering Proficiency 41	Infantry Proficiency 42	Combat Engineer Proficiency 43	Other Combat MOS Proficiency 44	Radio Teletype Operator Proficiency 45	Medical Care Specialist Proficiency 46	Other Non-Combat MOS Proficiency 47
	Attributes						
1. ABI Academic Performance	2.55	2.64	2.82	2.82	2.73	3.36	3.09
2. ABI Formal Leadership	1.64	1.55	1.64	1.45	1.45	1.73	1.64
3. ABI Ruggedness	3.18	3.27	2.73	2.82	1.82	1.64	1.73
4. ABI Mechanical Activities	2.55	2.36	3.64	2.73	3.09	1.45	2.27
5. ABI Work Experience	1.91	1.36	1.36	1.36	1.36	1.18	1.45
6. ABI Home Economics	1.45	0.64	0.55	0.55	0.45	0.82	0.82
7. ABI Nondelinquency	1.55	1.36	1.36	1.36	1.18	1.27	1.36
8. ABI Team Sports/Group Orientation	1.82	1.45	1.36	1.45	0.82	0.91	0.82
9. ABI Work Skills	2.09	1.91	2.00	1.91	1.91	1.82	1.91
10. ABI Family/Community	0.91	0.36	0.36	0.36	0.27	0.36	0.27
11. ABI Cross Cultural Sensitivity	0.82	0.55	0.45	0.55	0.55	1.09	0.45
12. RBI Cognition Under Stress	1.45	1.82	1.55	1.55	1.27	1.36	1.27
13. RBI Mature Team Commitment	1.64	1.45	1.09	1.09	0.91	0.91	0.91
14. RBI Self Esteem	1.36	1.36	1.36	1.36	1.27	1.27	1.36
15. RBI Combat Motivation	2.82	3.18	2.09	2.09	0.64	0.64	0.73
16. RBI Need for Achievement	1.91	1.64	1.64	1.64	1.55	1.55	1.64
17. RBI Outdoor Orientation	2.45	2.27	2.09	1.91	0.91	0.91	0.91
18. RBI Physical Endurance	2.27	2.27	1.91	1.91	0.73	0.91	0.73
19. RBI Physical Strength	2.00	2.27	1.91	1.91	0.64	0.64	0.64
20. RBI Object Belief	0.82	0.55	0.55	0.55	0.55	1.00	0.73
21. FCABLE-Work Orientation	2.45	2.18	2.18	2.18	2.18	2.18	2.09
22. FCABLE-Dominance	1.27	1.09	1.09	1.09	1.00	0.91	1.00
23. FCABLE-Dependability	1.82	1.73	1.73	1.73	1.73	1.55	1.64
24. FCABLE-Agreeableness	0.82	0.73	0.73	0.73	0.64	1.36	0.82
25. FCABLE-Emotional Stability	1.64	1.64	1.64	1.64	1.64	2.00	1.73
26. AVOICE-Rugged/Outdoors	2.82	3.09	2.73	3.00	1.00	0.55	0.82
27. AVOICE-Audiovisual Arts	0.18	0.36	0.64	0.36	0.45	0.73	0.73
28. AVOICE-Interpersonal	0.91	0.82	0.73	0.73	0.64	3.64	1.64
29. AVOICE-Skilled Technical	1.73	1.55	3.00	2.27	3.45	2.00	2.36
30. AVOICE-Administrative	0.73	0.55	0.64	0.55	1.00	0.55	2.36
31. AVOICE-Food Service	0.45	0.36	0.36	0.36	0.36	0.45	1.00
32. AVOICE-Protective Services	1.27	1.91	0.91	1.36	0.36	0.64	1.55
33. AVOICE-Structural/Machines	1.45	1.55	2.36	1.45	1.64	0.82	1.55
34. JOB-High Expectations	1.45	1.09	1.09	1.09	0.82	1.36	1.36
35. JOB-Routine	0.91	1.09	1.09	1.09	1.45	0.91	1.27
36. JOB-Autonomy	0.82	0.64	0.64	0.64	0.91	0.91	1.09
37. JCQ-SF Scale	3.00	2.45	1.73	2.00	1.27	1.36	1.18
38. JCQ-Weapons	2.45	4.00	2.09	2.55	0.73	0.64	0.91
39. JCQ-Engineer	1.27	1.27	5.09	2.27	0.82	0.73	1.00
40. JCQ-Commo	1.18	1.00	0.82	1.09	4.64	0.73	1.27
41. JCQ-Medic	1.18	0.91	0.82	0.91	0.64	5.27	0.82
42. Organizational Identity	1.18	1.00	1.00	1.00	0.82	1.00	1.09
43. OSAI	1.27	1.27	1.09	1.09	0.82	1.00	0.82
44. SI Biodata	0.91	0.55	0.55	0.55	0.55	0.82	1.00

Note. 0-2 = attribute not measured at all by predictor 3-5 = attribute measured partly by predictor 6-8 = attribute entirely measured by predictor

Table 3
Mean Extent of Measurement Ratings for Physical and Psychomotor Instruments

Test or Scale	Attributes: Judgment 1	Planning 2	Adaptability 3	Creativity 4	Auditory Ability 5	Mechanical Ability 6	Spatial Ability 7	Perceptual Ability 8	Basic Math 9	Advanced Math 10
1. Project A Target Tracking 1	0.50	0.40	0.90	0.40	0.00	1.30	3.30	2.70	0.50	0.30
2. Project A Target Tracking 2	0.50	0.20	0.80	0.40	0.00	1.50	3.40	2.80	0.50	0.30
3. Project A Target Shoot Test	0.90	0.50	0.80	0.50	0.00	1.40	3.00	2.70	0.30	0.30
4. Cannon Shoot Test	0.90	0.80	1.10	0.70	0.00	1.40	3.60	2.90	0.30	0.30
5. Army Physical Fitness Test	0.20	0.10	0.10	0.10	0.00	0.10	0.10	0.10	0.10	0.10
6. SFAS Physical Endurance Composite	0.30	0.30	0.50	0.10	0.00	0.20	0.40	0.50	0.20	0.10
7. SFAS Physical Fitness Composite	0.20	0.20	0.30	0.10	0.00	0.20	0.20	0.30	0.20	0.10
8. SFAS Swim Test	0.20	0.20	0.30	0.10	0.00	0.20	0.20	0.30	0.20	0.10

Note: 0-2 = attribute not measured at all by predictor 3-5 = attribute measured partly by predictor 6-8 = attribute entirely measured by predictor

Table 3
Mean Extent of Measurement Ratings for Physical and Psychomotor Instruments

Test or Scale	Reading Ability 11	Writing Ability 12	Language Ability 13	Communication 14	Non-Verbal Comm 15	Persuasiveness 16	Cultural/Interperson 17	Maturity 18	Autonomy 19	Team Playership 20
	Attributes									
1. Project A Target Tracking 1	0.40	0.30	0.10	0.10	0.10	0.10	0.10	0.30	0.40	0.20
2. Project A Target Tracking 2	0.40	0.30	0.10	0.10	0.10	0.10	0.10	0.40	0.40	0.20
3. Project A Target Shoot Test	0.40	0.30	0.10	0.10	0.10	0.10	0.10	0.20	0.40	0.20
4. Cannon Shoot Test	0.40	0.20	0.20	0.10	0.10	0.10	0.10	0.20	0.40	0.20
5. Army Physical Fitness Test	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.70	0.50	0.10
6. SFAS Physical Endurance Composite	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.90	0.70	0.10
7. SFAS Physical Fitness Composite	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.70	0.50	0.10
8. SFAS Swim Test	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.70	0.60	0.10

Note: 0-2 = attribute not measured at all by predictor 3-5 = attribute measured partly by predictor 6 -8 = attribute entirely measured

Table 3
Mean Extent of Measurement Ratings for Physical and Psychomotor Instruments

	Attributes									
Test or Scale	Dependability 21	Initiative 22	Perseverance 23	Moral Courage 24	Motivating Others 25	Supervising 26	Swimming 27	Physical Flexibility 28	Physical Strength 29	Physical Endurance 30
1. Project A Target Tracking 1	0.20	0.40	0.50	0.20	0.20	0.10	0.10	0.90	0.10	0.30
2. Project A Target Tracking 2	0.20	0.50	0.60	0.20	0.20	0.10	0.10	0.90	0.10	0.30
3. Project A Target Shoot Test	0.20	0.40	0.50	0.20	0.20	0.10	0.10	0.90	0.10	0.30
4. Cannon Shoot Test	0.20	0.40	0.50	0.20	0.20	0.10	0.10	0.60	0.10	0.20
5. Army Physical Fitness Test	0.50	0.90	2.70	0.20	0.30	0.10	2.20	3.30	6.00	5.10
6. SFAS Physical Endurance Composite	0.80	1.00	4.90	0.30	0.50	0.10	1.70	3.80	6.20	7.30
7. SFAS Physical Fitness Composite	0.60	0.80	4.00	0.20	0.40	0.10	1.40	2.60	4.10	6.20
8. SFAS Swim Test	0.50	0.60	3.20	0.20	0.30	0.10	6.80	2.30	4.40	5.10

Note: 0-2 = attribute not measured at all by predictor 3-5 = attribute measured partly by predictor 6 -8 = attribute entirely measured

Table 3
Mean Extent of Measurement Ratings for Physical and Psychomotor Instruments

Test or Scale	Attributes: Psychomotor Ability 31	Interest in Adventure 32	Interest in Skilled Trade 33	Interest in Other Cultures 34	Interest in People 35	Enterprising Interests 36	Leadership 37	Achievement and Effort 38	Personal Discipline 39	Physical Fitness and Military Bearing 40
1. Project A Target Tracking 1	6.70	1.00	0.80	0.20	0.20	0.20	0.20	0.80	0.50	0.40
2. Project A Target Tracking 2	6.90	0.90	0.80	0.20	0.20	0.20	0.20	0.80	0.50	0.40
3. Project A Target Shoot Test	6.60	1.10	0.70	0.20	0.20	0.20	0.20	0.80	0.50	0.40
4. Cannon Shoot Test	6.00	1.00	0.70	0.20	0.20	0.20	0.20	0.60	0.50	0.40
5. Army Physical Fitness Test	1.00	1.90	0.70	0.10	0.10	0.30	0.40	1.60	1.30	5.90
6. SFAS Physical Endurance Composite	1.10	2.10	0.60	0.10	0.10	0.40	0.60	2.00	1.40	6.10
7. SFAS Physical Fitness Composite	1.10	1.70	0.50	0.10	0.10	0.30	0.40	1.60	1.20	5.70
8. SFAS Swim Test	0.80	1.60	0.30	0.30	0.10	0.40	0.40	1.40	1.10	4.20

Note: 0-2 = attribute not measured at all by predictor 3-5 = attribute measured partly by predictor 6 -8 = attribute entirely measured

Table 3
Mean Extent of Measurement Ratings for Physical and Psychomotor Instruments

	Attributes						
	General Soldiering Proficiency	Infantry Proficiency	Combat Engineer Proficiency	Other Combat MOS Proficiency	Radio Teletype Operator Proficiency	Medical Care Specialist Proficiency	Other Non-Combat MOS Proficiency
Test or Scale	**41**	**42**	**43**	**44**	**45**	**46**	**47**
1. Project A Target Tracking 1	3.40	3.20	3.00	2.70	2.10	1.80	1.90
2. Project A Target Tracking 2	3.30	3.20	3.10	2.70	2.10	1.80	1.90
3. Project A Target Shoot Test	3.30	3.20	2.90	2.60	2.10	1.80	1.90
4. Cannon Shoot Test	3.00	3.00	2.80	2.50	2.00	2.00	1.90
5. Army Physical Fitness Test	2.40	2.50	1.80	1.60	0.90	1.10	1.20
6. SFAS Physical Endurance Composite	2.50	2.60	1.90	1.80	0.90	1.10	1.10
7. SFAS Physical Fitness Composite	2.40	2.50	1.70	1.60	0.90	1.10	1.10
8. SFAS Swim Test	1.50	1.70	1.10	1.00	0.60	0.80	0.80

Note: 0-2 = attribute not measured at all by predictor 3-5 = attribute measured partly by predictor 6 -8 = attribute entirely measured

Table 4
Mean Extent of Measurement Ratings for Performance Measures

	Attributes									
	Judgment	Planning	Adaptability	Creativity	Auditory Ability	Mechanical Ability	Spatial Ability	Perceptual Ability	Basic Math	Advanced Math
Test or Scale	1	2	3	4	5	6	7	8	9	10
1. Self Development Test	2.83	2.33	1.58	1.25	0.00	1.17	1.17	0.75	0.83	0.50
2. SFAS Military Orienteering	2.58	1.83	2.50	1.17	0.25	0.25	4.50	3.33	0.83	0.58
3. SFAS Peer Rankings	2.67	2.50	2.00	1.42	0.00	0.25	0.17	0.33	0.08	0.17
4. SFAS Situation Reaction Exercises	4.42	4.17	2.92	2.75	0.17	0.58	0.83	0.92	0.25	0.17
5. Number of Awards and Certificates	1.67	1.08	1.25	0.92	0.08	0.33	0.33	0.42	0.50	0.00
6. Number of Article 15s and Flag Actions	1.67	0.75	0.67	0.25	0.08	0.08	0.08	0.08	0.00	0.00
7. Promotion Rate	1.75	1.33	1.25	0.83	0.08	0.42	0.25	0.33	0.25	0.00
8. Work and Training Portfolio	1.17	1.67	1.25	0.75	0.17	0.58	0.75	0.67	0.25	0.42
9. Language Training Record	0.50	0.75	0.33	0.08	0.50	0.25	0.33	0.58	0.17	0.17
10. Army Wide-Performance Rating Scales	1.33	1.00	1.08	0.58	0.08	0.83	0.92	0.83	0.25	0.08
11. Teaching Role Play	1.92	3.92	1.42	3.42	0.33	0.42	0.42	0.75	0.25	0.08
12. Cultural Adaptability Role Play	3.17	1.33	3.17	1.25	0.58	0.17	0.33	1.50	0.00	0.00
13. Structural Interview	3.00	2.83	2.67	2.25	0.33	0.25	0.25	0.58	0.17	0.17
14. Low Fidelity Situational Simulation	4.08	1.83	1.92	1.58	0.00	0.08	0.17	0.42	0.00	0.00
15. High Fidelity Situational Simulation	4.17	1.75	2.17	1.83	0.50	0.00	0.42	1.17	0.00	0.00
16. NCO Role Plays	3.17	2.08	1.67	1.42	0.17	0.00	0.25	0.67	0.33	0.00

Note: 0-2 = attribute not measured at all by predictor 3-5 = attribute measured partly by predictor 6 -8 = attribute entirely measured

Table 4
Mean Extent of Measurement Ratings for Performance Measures

	Attributes									
Test or Scale	Reading Ability 11	Writing Ability 12	Language Ability 13	Communication 14	Non-Verbal Comm 15	Persuasiveness 16	Cultural/Interperson 17	Maturity 18	Autonomy 19	Team Playership 20
1. Self Development Test	2.08	1.17	0.50	0.50	0.00	0.83	0.50	1.00	0.92	1.17
2. SFAS Military Orienteering	0.33	0.00	0.00	0.00	0.00	0.00	0.00	2.17	1.83	0.17
3. SFAS Peer Rankings	0.17	0.00	0.08	1.83	1.50	2.67	1.42	2.42	0.92	5.25
4. SFAS Situation Reaction Exercises	0.50	0.33	0.08	3.25	1.75	2.33	1.00	2.92	0.92	4.50
5. Number of Awards and Certificates	0.25	0.17	0.17	1.17	0.42	1.08	0.67	1.75	0.83	2.00
6. Number of Article 15s and Flag Actions	0.08	0.08	0.00	0.42	0.50	0.92	0.75	2.25	0.67	1.50
7. Promotion Rate	0.67	0.83	0.58	1.00	0.75	1.00	0.83	2.08	1.25	1.83
8. Work and Training Portfolio	0.75	1.00	0.83	1.33	0.33	0.67	0.75	1.42	0.92	1.42
9. Language Training Record	1.08	0.58	5.58	1.58	0.25	0.33	0.83	0.25	0.08	0.00
10. Army Wide-Performance Rating Scales	0.50	0.33	0.00	1.83	0.50	1.25	0.50	1.92	1.42	1.92
11. Teaching Role Play	1.08	0.67	0.08	5.25	4.08	2.17	1.00	1.75	0.58	1.17
12. Cultural Adaptability Role Play	0.33	0.42	2.67	4.25	4.75	5.00	5.58	2.75	1.08	0.83
13. Structural Interview	0.17	0.17	0.58	5.08	4.50	3.33	2.58	2.42	2.00	2.42
14. Low Fidelity Situational Simulation	1.75	0.25	0.08	1.83	0.33	2.33	2.67	2.25	0.67	2.83
15. High Fidelity Situational Simulation	0.67	0.08	0.08	2.17	2.50	2.83	2.83	2.17	0.92	2.75
16. NCO Role Plays	0.17	0.08	0.33	4.83	4.33	3.58	1.75	2.58	1.08	1.50

Note: 0-2 = attribute not measured at all by predictor 3-5 = attribute measured partly by predictor 6 -8 = attribute entirely measured

Table 4
Mean Extent of Measurement Ratings for Performance Measures

	Attributes									
Test or Scale	Dependability 21	Initiative 22	Perseverance 23	Moral Courage 24	Motivating Others 25	Supervising 26	Swimming 27	Physical Flexibility 28	Physical Strength 29	Physical Endurance 30
1. Self Development Test	1.50	1.75	1.00	0.50	1.67	3.00	0.00	0.00	0.00	0.08
2. SFAS Military Orienteering	0.92	2.50	4.17	0.58	0.08	0.00	0.00	1.42	3.08	4.58
3. SFAS Peer Rankings	3.58	2.92	2.67	2.00	4.50	3.58	0.00	0.25	0.67	0.83
4. SFAS Situation Reaction Exercises	3.83	3.83	3.17	1.58	3.92	3.58	0.08	1.00	2.00	2.75
5. Number of Awards and Certificates	2.92	3.42	2.33	1.33	2.08	2.08	0.00	0.25	1.08	1.42
6. Number of Article 15s and Flag Actions	3.25	1.25	1.25	2.42	0.92	1.33	0.00	0.08	0.50	0.50
7. Promotion Rate	3.42	3.42	2.67	1.50	2.08	2.58	0.00	0.17	0.25	0.33
8. Work and Training Portfolio	1.67	1.83	1.50	0.75	1.08	1.25	0.00	0.17	0.17	0.25
9. Language Training Record	0.50	1.17	0.58	0.00	0.00	0.33	0.00	0.00	0.00	0.00
10. Army Wide-Performance Rating Scales	2.83	2.92	2.75	1.42	2.67	4.42	0.00	0.42	1.50	1.58
11. Teaching Role Play	1.00	1.75	1.00	0.25	3.92	2.67	0.17	0.00	0.33	0.42
12. Cultural Adaptability Role Play	1.25	1.17	0.83	0.92	0.67	0.75	0.00	0.00	0.08	0.08
13. Structural Interview	2.42	3.08	2.75	2.08	2.42	1.83	0.00	0.17	0.33	0.33
14. Low Fidelity Situational Simulation	1.33	1.67	1.17	2.17	2.92	3.08	0.00	0.00	0.08	0.08
15. High Fidelity Situational Simulation	1.25	1.67	1.17	2.08	2.83	3.25	0.00	0.00	0.00	0.00
16. NCO Role Plays	1.17	1.75	0.83	1.33	3.58	4.33	0.00	0.00	0.00	0.00

Note: 0-2 = attribute not measured at all by predictor 3-5 = attribute measured partly by predictor 6 -8 = attribute entirely measured

Table 4
Mean Extent of Measurement Ratings for Performance Measures

	Attributes									
Test or Scale	Psychomotor Ability 31	Interest in Adventure 32	Interest in Skilled Trade 33	Interest in Other Cultures 34	Interest in People 35	Enterprising Interests 36	Leadership 37	Achievement and Effort 38	Personal Discipline 39	Physical Fitness and Military Bearing 40
1. Self Development Test	0.08	0.58	0.58	0.58	0.67	0.75	4.33	3.17	1.67	1.00
2. SFAS Military Orienteering	1.83	2.58	0.25	0.00	0.00	0.08	0.67	2.75	1.92	3.50
3. SFAS Peer Rankings	0.25	0.67	0.17	0.25	1.83	2.00	5.67	3.00	2.92	2.50
4. SFAS Situation Reaction Exercises	1.00	1.17	0.25	0.25	1.92	2.17	4.92	3.42	2.08	3.17
5. Number of Awards and Certificates	0.67	0.50	0.25	0.17	0.33	1.00	3.75	4.58	2.75	2.42
6. Number of Article 15s and Flag Actions	0.25	0.17	0.17	0.17	0.17	0.33	3.00	2.92	5.17	2.58
7. Promotion Rate	0.25	0.25	0.17	0.33	0.75	1.08	3.58	4.00	4.25	3.00
8. Work and Training Portfolio	0.42	1.00	0.83	0.75	1.00	1.42	3.25	3.17	1.75	1.75
9. Language Training Record	0.00	0.00	0.00	1.67	1.25	0.25	0.42	1.42	0.67	0.25
10. Army Wide-Performance Rating Scales	0.92	0.33	0.08	0.00	1.42	1.00	4.67	4.33	5.33	5.58
11. Teaching Role Play	0.42	0.17	0.17	0.08	2.08	1.33	3.25	2.17	1.08	0.75
12. Cultural Adaptability Role Play	0.25	0.00	0.00	3.92	3.17	0.83	1.58	1.42	1.42	0.50
13. Structural Interview	0.17	0.92	0.92	1.17	1.42	1.33	3.33	3.08	3.17	1.83
14. Low Fidelity Situational Simulation	0.00	0.25	0.17	1.25	1.33	0.75	4.00	2.50	1.75	0.58
15. High Fidelity Situational Simulation	0.00	0.25	0.17	1.42	1.50	1.08	4.17	2.50	1.58	0.50
16. NCO Role Plays	0.42	0.08	0.00	0.50	1.83	1.00	4.75	2.25	1.75	0.67

Note: 0-2 = attribute not measured at all by predictor 3-5 = attribute measured partly by predictor 6 -8 = attribute entirely measured

Table 4
Mean Extent of Measurement Ratings for Performance Measures

	Attributes						
Test or Scale	General Soldiering Proficiency 41	Infantry Proficiency 42	Combat Engineer Proficiency 43	Other Combat MOS Proficiency 44	Radio Teletype Operator Proficiency 45	Medical Care Specialist Proficiency 46	Other Non-Combat MOS Proficiency 47
1. Self Development Test	3.75	4.33	4.08	4.17	4.00	3.58	3.58
2. SFAS Military Orienteering	4.17	3.83	2.83	2.83	0.92	0.58	0.92
3. SFAS Peer Rankings	2.83	2.25	2.08	2.17	1.58	1.50	1.83
4. SFAS Situation Reaction Exercises	3.58	2.83	2.33	2.25	1.50	1.50	1.67
5. Number of Awards and Certificates	2.83	2.58	2.33	2.33	2.08	1.92	2.00
6. Number of Article 15s and Flag Actions	1.50	1.08	0.92	0.92	0.83	0.83	0.83
7. Promotion Rate	2.75	2.33	2.17	2.17	2.08	1.67	1.75
8. Work and Training Portfolio	3.58	3.25	2.83	2.83	2.67	2.67	2.67
9. Language Training Record	0.25	0.25	0.17	0.25	0.17	0.17	0.17
10. Army Wide-Performance Rating Scales	4.50	3.25	2.92	3.00	2.67	2.58	2.58
11. Teaching Role Play	2.50	1.75	1.25	1.25	1.25	1.33	1.33
12. Cultural Adaptability Role Play	0.75	0.75	0.58	0.58	0.58	0.50	0.50
13. Structural Interview	1.42	1.33	1.33	1.33	1.25	1.25	1.33
14. Low Fidelity Situational Simulation	1.42	1.08	1.00	0.92	0.83	0.83	0.83
15. High Fidelity Situational Simulation	1.42	1.08	1.00	0.92	0.83	0.83	0.83
16. NCO Role Plays	1.42	0.75	0.75	0.75	0.67	0.83	0.83

Note: 0-2 = attribute not measured at all by predictor 3-5 = attribute measured partly by predictor 6 -8 = attribute entirely measured

We calculated Intraclass Correlations (ICCs) for the ratings for each predictor category to assess the reliability of the judgments within that category. The appropriate ICC for this model is based on a three-way ANOVA (Peterson, Owens-Kurtz, et al., 1990). The results of these analyses are shown in Table 5. The reliabilities for the ratings for all four categories are high, ranging from .89 to .96, even though the four categories may differ in their opportunities for disagreement (e.g., the cognitive and noncognitive categories could be expected to have fewer zero ratings than the physical/ psychomotor and performance categories). The number of raters varied across the four rating exercises (ranging from 9 to 12); therefore, we chose a common N of 20 and adjusted the ICCs to allow comparisons between them. When adjusted to this common number of raters, the ICCs were not practically different (e.g., ranging from .93 to .98). We are, therefore, confident in the quality of all of the judgments.

Table 5

Intraclass Correlation Coefficients (ICCs) for Expert Judgments

Number of Raters	Rating Exercise			
	Cognitive Measures (K=42) (N=12)	Non-Cognitive Measures (K=44) (N=11)	Performance Measures (K=8) (N=12)	Physical/ Psychomotor (K=16) (N=9)
N	.93	.90	.89	.96
20	.95	.94	.93	.98

Notes: N = actual number of raters
K = number of measures
The ICCs adjusted to the full N reflect the overall level of reliability of the observed ratings. Given that ICCs are influenced by sample size, a common N (N=20) was chosen to allow comparison of the ICCs across the rating exercises where N varies.

To further summarize the mean ratings, we compiled a listing of the top-ranked measures (based on their mean extent of measurement ratings) for each of the 47 attributes. These listings are included in Appendix H. Referring to these listings enables the reader to bypass the step of scrutinizing the data in Tables 1- 4, to quickly identify the rank ordering of tests and scales that are likely to be good measures of each specific attribute. We listed the ten highest rated measures for each attribute; measures rated lass than 3.0 were dropped.

As an example of the outcome of the expert judgments, we identified the highest ranking predictors for the ten most important SF attributes (based on job analysis data). Those attributes, along with their grand mean (across all MOS) on a five-point importance scale were:

Attribute Name and Number	Importance (grand mean)
Team Playership (#20)	4.54
Maturity (#18)	4.47
Dependability (#21)	4.45
Judgment and Decision Making (#1)	4.45
Adaptability (#3)	4.36
Cultural/Interpersonal Adaptability (#17)	4.29
Physical Endurance (#30)	4.27
Initiative (#22)	4.23
Perseverence (#23)	4.20
Autonomy (#19)	4.19

For the two cognitive ability attributes, Judgment and Decision Making and Adaptability, the highest ranking likely predictors were general intelligence measures (e.g., Wonderlic) and measures of planning and problem-solving (e.g., Problem Solving Skills, Planning and Implementation, Solution Characteristics, Category Search and Specification). For the seven highly ranked interpersonal/motivation/character attributes (attribute numbers 17, 18, 19, 20, 21, 22, and 23 listed above), biodata measures from the ABI, RBI, and FCABLE were the common likely measures. SFAS measures such as Peer Rankings and Situation Reaction Exercises were also listed for Team Playership and Dependability. Perseverance and the tenth attribute Physical Endurance had similar listings of likely predictors; these measures included SFAS Physical Endurance Composite, RBI Physical Endurance, SFAS Military Orienteering, SFAS Swim Test, etc. Please keep in mind that this is only an example, not a complete summary of the ratings.

Identify Useful Measures. The three main objectives for the identification of useful predictors were to ensure:

(1) that predictors selected would be of high quality by using the experts' extent of measurement judgments,

(2) the feasibility of the predictors by selecting predictors that are readily available (i.e., operational, experimental, published) as well as high in quality, and

(3) the comprehensiveness of the total set of predictors by including predictors to measure all the job analysis attributes that appear to be measurable.

We used the means tables to identify the most likely predictor measure(s) for each of the 47 SF attributes.[1] We used a set of decision rules for identifying predictor measures and systematically reviewed the results of Tables 1 through 4. The decision rules we used to specify the most promising predictor measure(s) for each attribute are:

[1]We found there was not a lot of agreement across judges on one best bet; therefore, we did not use the "best bet" information in identifying promising measures.

(1) Identify measures rated 3.0 or greater for the attribute.

(2) Determine whether the attribute is measured either by existing operational measures or by experimental measures that are readily available.

If it is, then determine whether these measures might not be adequate for another reason (e.g., need to assess a different population than assessed by the measure, need to place a measure at a point in the selection process where the readily available measures cannot be used).

If they are not adequate, go to (3). Otherwise, use the readily available measures.

(3) Consider which proposed or other measures would be useful to measure attributes not measured by the readily available measures.

The application of these decision rules led to a set of recommended predictors. They are described in Chapter 3.

Collection of Expert Judgments about Criterion Measures

The purpose of the criterion expert judgment exercise was to map potential criterion measures against SF performance categories to examine the extent to which criterion measures cover the performance domain. These potential criterion measures included measures that can be (or already are) collected at any point in the SF career progression. Many measures are currently collected early in the SF career progression, e.g., during assessment (SFAS) phases and training (SFQC) phases. These measures may serve as criteria for the variables that are used in pre-screening candidates for SFAS. These same measures may be archivally obtained and treated as predictors of later success in SF -- for those who are selected. Thus, we set out to ensure that we included all potential measures in the criterion domain, ranging from measures collected early in the assessment/selection process (often employed as predictors) to measures more traditionally viewed as criteria, such as work samples and ratings of on-the-job performance.

We completed the expert judgment exercise in the following four steps:

- Gather written documents and interview researchers
- Prepare materials for the exercise
- Collect expert judgments
- Analyze the data and interpret the results

Gather Information and Conduct Interviews. In this step, the main activity was to collect information about a wide variety of measures that either (1) currently were used as criterion measures, or (2) could potentially serve as criterion measures for the SF performance categories. There are many measures that are currently in use or under development by the Army that we considered to be good candidates for measuring

performance in the various SF performance categories. We identified these tests by first identifying criterion measures that project members had worked on (e.g., in military job performance measurement projects). Then we asked Army test and measurement experts for their input. We conducted interviews with a number of researchers from ARI, Army TRADOC, HumRRO, and AIR.

To begin each interview, we described the Roadmap project purpose and background. We asked the researchers to describe their projects that included potential criterion measures. In some cases, we were already familiar with the interviewee's research, and therefore probed for certain descriptive and psychometric information about specific measures. We asked interviewees to inform us of any relevant work done by others (that we were unaware of). We also asked for copies of technical or other reports that we could use to locate specific descriptive details for the criterion measures.

We reviewed codebooks for the SFQC and SFAS databases to identify variables that would be useful as criteria in a validation study. We found several measures to include in the expert judgment exercise; these variables included the end status of SFAS (graduated/did not graduate), peer rankings of leadership potential (during SFAS), and several SFQC training status variables, land navigation scores, and peer rankings.

At the end of the information gathering stage, we had a set of potential measures that represented a variety of measurement methods, including archival data, job performance ratings, peer rankings, hands-on performance tests, and written tests. While a number of the instruments have been used operationally in SF (e.g., Land Navigation Field Exam, End-of-Training Written School Knowledge Tests), there are others that have been developed recently for either SF (e.g., SF-Common Behaviorally-Anchored Rating Scales) or for the conventional Army (e.g., Hands-On (Common Task)) Performance Tests. It is important to note that some of these measures could be used as either predictor or criterion measures, depending on what point in time they would be collected and how they would be used.

Prepare Materials. We consolidated all of the information from the interviews, technical reports, and other materials, onto a description form for each measure (an example is provided in Figure 6). We listed the following information (except where the information was not available from any source):

- Short description of measure

- Psychometric properties:
  - Scoring
  - Relevance
  - Comprehensiveness
  - Discriminability
  - Practicality/feasibility
  - Susceptibility to contamination
  - Correlations with other variables
  - Variables that predict it best

Operational Measure	1

Measure: SFAS Graduation

Short Description of Measure: SFAS is a three week assessment program in which SF candidates participate in a series of rigorous physical exercises such as ruckmarches, obstacle course, and runs. Participants are deprived of sleep and put under extreme physical stress in a series of team events that require planning, teamwork, and physical endurance.

SFAS Graduation is thus a measure of completion of the SF assessment and selection program, the first major step in becoming a member of SF. This variable is recorded in the SFAS database.

Psychometrics:

Scoring: This variable has values of 0 for "NO" (did not graduate) and 1 for "YES."

Note that additional categorical variables exist that have more detail than just "Yes/No," e.g., SFAS Final Status (HISTORY) and Reason Dropped from SFAS (RESULT).

Relevance:: Relevant to the assessment/selection domain; this measure of performance in the assessment/selection program is collected before any job experience is gained

Comprehensiveness: This is a summary level measure of performance, summarizing behavior on many individual-level physical activities/tests and team-level exercises (e.g., situation-reaction and military orienteering events).

Discriminability: The dichotomous scoring does not provide as much information as a continuous distribution would; it also provides less information than the HISTORY and RESULT variables do about the final outcome of the SFAS performance.

Practicality/feasibility: No extra time is required to develop this measure; it is available in archival records.

Susceptibility to contamination: There are reasons beyond the individual's performance/control (e.g., medical and involuntary drop reasons) that can account for not graduating.

Correlations with other Variables: Analyses to correlate SFAS Graduation with other variables have not been run yet because the process of building the SFdata base is not complete.

Variables that predict it best: SFAS decision-makers were found to rely heavily on ruck march scores in making the graduation decision (Teplitzky, 1991). Validation analyses have not been completed; these will be run when the database is completed.

Figure 6
Example Criterion Measure Description Form

We assigned one of the following labels to each measure to indicate its level of development:

- Proposed -- instruments under consideration for development for Special Forces.
- Experimental -- instruments that have been developed and field tested but are not currently in use.
- Operational -- instruments that are currently in use.

A list of the criterion measures appears in Figure 7. The complete set of description forms is included in Appendix I. Appendix J contains the bibliography compiled for these criterion measure description forms.

We planned to ask the experts to rate the extent to which each measure measures the SF job performance categories defined in the job analysis. Unlike the predictor judgments, these judgments require knowledge of SF jobs as well as knowledge of measurement. We designated expert judges as those who had participated in the SF Job Analysis study and were therefore very knowledgeable about SF jobs. Four individuals (two at HumRRO and two at AIR) were asked to participate.

1. SFAS Graduation	2. SFAS Peer Rank	3. Q-Course Honors
4. Q-Course Final Status	5. Q-Course Retrained	6. Q-Course Tries at Bragg
7. Q-Course Peer Rank	8. Q-Course LN Written	9. Q-C Land Nav Field Exam
10. Promotion Rate	11. # of Disciplinary Act.	12. # of Awards, Memoranda,..
13. NCOER and OER	14. Common Tasks Hands-On	15. MOS Task Hands-On
16. Training Performance Test	17. MOS-Specific Ratings	18. SF Common Ratings
19. SF Knowledge Test	20. MOS Knowledge Test	21. Task Performance Ratings
22. Situational Judgment Test	23. Job Simulations	24. Training Knowledge Test
25. Q-Course Rating Scales	26. Cadre Ratings (Robin Sage)	27. Self Development Test
28. Language Proficiency Test	29. Language School Grades	30. JRTC Rating
31. Language School Ratings	32. Client Ratings	

JRTC=Joint-Readiness Training Center

Figure 7. Measures of Special Forces Training and On-the-Job Proficiency (Criteria)

We developed an expert judgment task to gather ratings of how well each Published, Experimental, and Proposed measure measures each SF performance category. The purpose of this expert judgment task was to assess the adequacy of the coverage of the criterion domain. We developed a spreadsheet for each judge to use to directly enter judgments into a data base. Each judge received instructions for completing the exercise (these materials are included in Appendix K).

The instructions asked that experts review the definitions of the 26 performance categories and scan the set of 32 measures. Judges were to carefully read the information provided for each measure, then rate each measure to indicate the extent to which it measures each performance category, using the following scale.

0----------1----------2----------3----------4----------5----------6----------7----------8

This performance category is not at all measurable by the criterion measure (it is almost useless)	This performance category is measured partly by the criterion measure (it is of some use)	This performance category is entirely measured by the criterion measure (it is highly useful)

Collect Data. All four designated judges completed ratings to link the SF performance categories with the criterion measures. The rating exercise required approximately three to four hours for each judge to complete. The judges completed their ratings on November 9 and 10, 1994.

Analyze and Interpret Data. We merged the already-entered data from the four judges and checked the file for errors. We calculated means and standard deviations for the ratings of each performance category and criterion measure combination. Higher means indicate a stronger relationship between the measures and the performance categories. Table 6 displays the means for the linkages between the measures and the SF performance categories.

We calculated the Intraclass Correlation (ICC) for the ratings for the measures to assess the reliability of the judgments. The appropriate model is a three-way ANOVA. The intraclass correlation for four raters was .83, showing very good agreement (.55 for one rater).

The inspection of the criterion expert judgments had two primary objectives:

(1) to determine whether criterion measures are available to sufficiently cover the SF performance categories for both training and on-the-job performance, and
(2) to identify criterion measures that could be used to buttress criterion measurement for specific studies and purposes.

Our first step was to sort the measures into two categories: (1) training measures and (2) on-the-job performance measures. Then we examined the mapping of the training measures against the performance categories to identify areas that were not well measured by existing variables. We repeated this step for the job performance measures. We found that for most validation purposes existing criterion measures would be sufficient; they measure most of the performance categories to some extent. We also identified measures that should be used to buttress existing measures for specific validation purposes. The results of this analysis appear in Chapter 3.

Table 6

Mean Judgments of the Relevance of Measures to SF Performance Categories

	Measures										
	SFAS Graduation	Peer Rankings - SFAS	SFQC HONORS	Final Training Status	Retrained - Different MOS	Total Tries - Ft. Bragg	Peer Rankings - SFQC	SFQC Land Nav Field	SFQC Land Nav Written	Promotion Rate	Disciplinary Actions
Performance Categories	1	2	3	4	5	6	7	8	9	10	11
A. Teaching Others	0.25	0.50	0.50	0.50	0.00	0.25	0.75	0.00	0.00	0.50	0.00
B. Building and Maintaining Effective Relationships with Indigenous Popula	0.00	1.75	0.75	1.25	0.00	0.25	2.00	0.00	0.00	0.75	0.25
C. Handling Interpersonal Situations	1.75	3.75	0.75	1.25	0.00	0.25	3.25	0.00	0.00	0.75	3.50
D. Using and Enhancing Language Skills	0.00	0.25	0.25	0.50	0.00	0.25	0.25	0.00	0.00	0.00	0.00
E. Contributing to the Team Effort	2.00	4.75	1.75	0.75	0.00	0.50	4.75	0.25	0.00	2.25	2.00
F. Showing Initiative and Extra Effort	2.75	3.25	4.50	3.75	1.75	2.00	3.50	0.50	1.25	4.00	2.75
G. Displaying Honesty and Integrity	1.00	2.75	1.25	1.25	0.50	0.75	2.50	0.00	0.00	1.50	4.25
H. Planning and Preparing for Missions	2.75	1.75	2.00	1.50	0.25	0.50	2.00	0.25	0.50	2.00	0.00
I. Decision Making	1.75	1.75	1.00	1.75	0.50	0.75	2.25	0.50	0.75	1.25	0.50
J. Confronting Physical and Environmental Challenges	4.75	2.50	1.75	2.00	0.00	1.50	2.25	0.25	1.75	1.75	0.00
K. Navigating in the Field	2.75	1.75	1.50	2.00	0.00	1.25	1.75	4.00	5.25	1.00	0.00
L. Troubleshooting and Solving Problems	2.50	1.75	1.50	1.25	0.50	0.75	1.75	0.50	0.75	1.25	0.25
M. Being Safety Conscious	0.50	0.50	0.75	2.00	0.00	0.50	0.25	0.00	0.50	0.00	0.50
N. Administering First Aid and Treating Casualties	0.25	0.50	0.75	1.25	0.00	0.25	0.50	0.00	0.25	0.00	0.00
O. Managing Administrative Duties	0.00	0.00	1.00	1.25	0.00	0.25	0.00	0.00	0.00	1.25	0.25
P. Operating and Maintaining Direct-Fire Weapons	0.25	0.25	1.75	2.00	1.75	1.75	0.50	0.00	0.25	1.25	0.25
Q. Employing Indirect Fire Weapons and Techniques	0.25	0.25	1.75	2.00	1.75	1.75	0.50	0.00	0.00	1.25	0.25
R. Employing Demolitions Techniques	0.25	0.25	1.75	2.00	1.75	1.75	0.50	0.00	0.00	1.25	0.25
S. Constructing for Mission-Related Requirements	0.75	1.00	1.75	2.00	1.75	1.75	0.50	0.00	0.00	1.25	0.25
T. Following Communication Procedures and Policies	0.25	0.25	1.75	2.00	1.75	1.75	0.50	0.00	0.00	1.25	0.25
U. Assembling and Operating Commo Equipment	0.25	0.25	1.75	2.00	1.75	1.75	0.50	0.00	0.00	1.25	0.25
V. Evaluating and Treating Medical Conditions and Injuries	0.25	0.25	1.75	2.00	1.75	1.75	0.50	0.00	0.00	1.25	0.25
W. Determining and Administering Medications and Dosages	0.25	0.25	1.75	2.00	1.75	1.75	0.50	0.00	0.00	1.25	0.25
X. Ensuring Standards of Health-Related Facilities	0.00	0.00	1.50	2.00	1.50	1.75	0.00	0.00	0.00	1.25	0.00
Y. Considering Subordinates	2.00	3.75	2.00	2.00	1.00	1.00	4.00	0.00	0.00	2.75	2.25
Z. Providing Direction	1.25	3.50	2.00	2.00	1.00	1.00	4.00	0.00	0.00	3.00	2.50

Table 6

Mean Judgments of the Relevance of Measures to SF Performance Categories

	Measures										
Performance Categories	Awards, Memorandum, Cert	NCOER/OER	HO Common Perform. Tests	HO MOS-Specific Perf. Tests	End-of-Training (MOS) Perform.	MOS-Specific BARS	SF-Common BARS	Written Common JK Tests	Written MOS-Specific Tests	Task Perform. Ratings	Situational Judgment Test
	12	13	14	15	16	17	18	19	20	21	22
A. Teaching Others	1.00	0.75	2.50	1.00	1.00	0.25	4.25	3.25	0.25	2.75	1.75
B. Building and Maintaining Effective Relationships with Indigenous Popula	0.50	0.75	0.75	0.50	0.50	0.25	4.25	1.75	0.25	2.00	2.75
C. Handling Interpersonal Situations	2.50	1.25	1.00	0.75	0.75	0.25	4.25	2.25	0.25	2.00	3.50
D. Using and Enhancing Language Skills	0.50	0.00	1.00	0.50	0.50	0.25	4.25	1.75	0.25	2.00	0.75
E. Contributing to the Team Effort	2.50	2.00	0.25	0.50	0.50	0.25	4.25	1.50	0.25	1.50	2.25
F. Showing Initiative and Extra Effort	4.00	3.00	0.00	0.50	0.50	0.25	4.25	1.50	0.25	1.50	1.50
G. Displaying Honesty and Integrity	3.25	2.75	0.50	0.50	0.50	0.25	4.25	1.50	0.25	1.50	2.75
H. Planning and Preparing for Missions	0.75	1.00	1.00	0.50	0.50	0.25	4.25	2.50	0.25	1.75	3.50
I. Decision Making	1.00	1.25	0.50	0.50	0.50	0.25	4.25	2.25	0.25	1.50	2.50
J. Confronting Physical and Environmental Challenges	0.50	0.50	1.25	0.50	0.50	0.25	4.25	2.50	0.25	2.25	0.00
K. Navigating in the Field	0.25	0.25	3.75	0.50	0.50	0.25	4.00	4.00	0.25	3.00	1.00
L. Troubleshooting and Solving Problems	0.75	1.00	2.75	1.00	1.00	0.25	4.25	3.50	0.25	3.00	1.00
M. Being Safety Conscious	0.50	0.25	3.00	0.50	0.50	0.25	4.25	3.50	0.25	2.50	0.25
N. Administering First Aid and Treating Casualties	0.25	0.00	2.50	1.00	1.00	0.25	4.25	3.50	0.25	3.00	1.25
O. Managing Administrative Duties	0.75	1.00	0.50	3.50	3.50	3.00	1.50	0.75	3.25	3.00	0.50
P. Operating and Maintaining Direct-Fire Weapons	0.75	0.00	1.00	4.00	4.00	4.25	0.25	0.25	4.00	3.25	0.25
Q. Employing Indirect Fire Weapons and Techniques	0.75	0.00	1.00	4.00	4.00	4.00	0.25	0.25	4.00	3.25	0.00
R. Employing Demolitions Techniques	0.75	0.00	1.00	4.00	4.00	4.00	0.25	0.25	3.75	3.25	0.25
S. Constructing for Mission-Related Requirements	0.75	0.00	1.00	4.00	4.00	4.25	0.25	0.25	3.75	3.25	0.25
T. Following Communication Procedures and Policies	0.75	0.00	1.00	4.00	4.00	4.25	0.25	0.25	3.75	3.25	0.50
U. Assembling and Operating Commo Equipment	0.75	0.00	1.00	4.00	4.00	4.25	0.25	0.25	3.75	3.25	0.00
V. Evaluating and Treating Medical Conditions and Injuries	1.00	0.00	1.00	4.00	4.00	4.25	0.25	0.25	3.75	3.25	0.75
W. Determining and Administering Medications and Dosages	1.00	0.00	1.00	4.00	4.00	4.25	0.25	0.25	3.75	3.25	0.75
X. Ensuring Standards of Health-Related Facilities	0.75	0.00	1.00	4.00	4.00	4.25	0.25	0.25	3.75	3.25	0.00
Y. Considering Subordinates	2.00	1.75	0.25	2.50	2.50	4.75	0.25	0.25	2.75	3.00	3.75
Z. Providing Direction	2.25	2.50	1.00	2.75	2.75	4.75	0.25	0.25	2.75	3.00	3.75

Table 6

Mean Judgments of the Relevance of Measures to SF Performance Categories

	Measures									
Performance Categories	SF Simulations 23	End-of-Training Written Tests 24	SFQC Peer/Instructor Ratings 25	Cadre Rtgs - Sage Role Play 26	Self Development Test 27	Defense Lang. Proficiency Test 28	Language School Grades 29	JRTC Performance 30	Language School Instructor Rtgs 31	"Client" Ratings 32
A. Teaching Others	0.50	1.00	1.00	2.00	1.50	0.00	0.00	1.75	0.00	3.50
B. Building and Maintaining Effective Relationships with Indigenous Po	0.75	0.50	1.75	3.50	0.75	0.00	0.00	0.50	0.00	4.25
C. Handling Interpersonal Situations	1.25	1.00	3.50	3.50	1.50	0.00	0.00	2.00	0.25	3.50
D. Using and Enhancing Language Skills	0.25	0.50	0.75	2.00	0.00	4.50	4.50	0.25	3.75	3.25
E. Contributing to the Team Effort	0.75	0.25	4.25	4.00	1.25	0.00	0.00	3.75	0.00	2.00
F. Showing Initiative and Extra Effort	0.75	0.00	3.50	4.00	1.00	0.50	0.75	3.50	3.00	3.00
G. Displaying Honesty and Integrity	0.50	0.75	3.00	2.75	1.00	0.00	0.00	2.00	0.00	2.50
H. Planning and Preparing for Missions	4.00	1.25	2.75	4.25	2.75	0.00	0.00	4.50	0.00	2.00
I. Decision Making	4.50	0.75	3.00	4.25	3.00	0.00	0.00	4.50	0.00	1.25
J. Confronting Physical and Environmental Challenges	0.00	0.00	3.00	4.50	0.75	0.00	0.00	1.50	0.00	1.25
K. Navigating in the Field	0.25	1.25	3.00	4.25	1.50	0.00	0.00	2.00	0.00	2.00
L. Troubleshooting and Solving Problems	2.25	0.50	3.00	4.25	2.00	0.00	0.00	3.75	0.00	1.50
M. Being Safety Conscious	0.25	0.50	1.50	3.00	1.25	0.00	0.00	1.00	0.00	1.25
N. Administering First Aid and Treating Casualties	0.00	0.75	0.75	1.50	2.00	0.00	0.00	0.75	0.00	1.50
O. Managing Administrative Duties	0.75	2.00	0.75	1.50	3.25	0.00	0.00	1.75	0.00	1.50
P. Operating and Maintaining Direct-Fire Weapons	0.00	3.25	2.00	2.50	3.50	0.00	0.00	1.00	0.00	1.50
Q. Employing Indirect Fire Weapons and Techniques	0.00	3.25	2.00	1.75	3.50	0.00	0.00	1.00	0.00	1.50
R. Employing Demolitions Techniques	0.00	3.25	2.00	2.50	3.50	0.00	0.00	1.00	0.00	1.50
S. Constructing for Mission-Related Requirements	0.00	3.25	2.00	2.50	3.50	0.00	0.00	1.00	0.00	1.50
T. Following Communication Procedures and Policies	0.25	3.25	2.00	2.50	3.50	0.00	0.00	1.00	0.00	1.50
U. Assembling and Operating Commo Equipment	0.00	3.25	2.00	2.50	3.50	0.00	0.00	1.00	0.00	1.50
V. Evaluating and Treating Medical Conditions and Injuries	0.75	3.25	2.00	2.50	3.50	0.00	0.00	1.00	0.00	1.50
W. Determining and Administering Medications and Dosages	0.75	3.25	2.00	2.50	3.50	0.00	0.00	1.00	0.00	1.50
X. Ensuring Standards of Health-Related Facilities	0.00	3.25	2.00	2.50	3.50	0.00	0.00	1.00	0.00	1.50
Y. Considering Subordinates	0.75	2.00	3.50	3.75	3.50	0.00	0.00	2.25	0.00	2.00
Z. Providing Direction	1.50	2.00	3.50	3.75	3.75	0.00	0.00	3.00	0.00	2.25

Collection of Expert Judgments about the Importance of SF Performance Categories for SF Missions

We constructed a survey to gather judgments about the importance of the 26 performance categories for SF's primary missions. The impetus for this was the common comment by SF personnel (when they looked at the performance categories) that the importance of each category varies with mission. Three steps were required to develop and conduct this expert judgment task:

- Prepare materials for the exercise
- Collect expert judgments
- Analyze the data and interpret the results

Prepare Materials. We prepared a brief background description and step-by-step instructions for completing the rating exercise. We used a five point rating scale which ranged from 1 ("unimportant") to 5 ("extremely important"). We instructed the expert raters to read through the list of descriptions of 21 performance categories. For the purpose of this exercise, we collapsed the 11 MOS-relevant skill performance categories into 6 more general skill area dimensions (e.g., "Medic Skills", "Engineering Skills"), which yielded 21, rather than the usual 26, performance categories. The instructions asked raters to rate the importance of each performance category for each of the five primary SF missions: Foreign Internal Defense (FID), Unconventional Warfare (UW), Direct Action (DA), Counterterrorism (CT), and Special Reconnaisance (SR). The materials used to collect these judgments are given in Appendix L.

Collect Data. We obtained permission from the head of the Special Operations Proponency Office (SOPO) to survey the population of that office (N = 6) to make mission performance judgments. We scheduled a date (November 30, 1994) to meet with them to explain the purpose and procedures of the exercise.

Six expert judges from SOPO participated in the expert judgment exercise. Their ranks were: one lieutenant colonel, one major, one captain, two warrant officers, and one sergeant major. We met with the group on November 30 and explained the procedure.

Analyze and Interpret Data. Table 7 displays the mean importance ratings for each of the performance categories and each SF mission. Figure 8 graphically depicts the mean ratings for the first performance categories A-K. We selected these performance categories to illustrate that there are differences in mean importance ratings between two subgroups of missions. Interpersonal, cultural, and language performance areas (performance categories A - D) are very important for FID and UW missions but not for DA, CT, and SR missions. All five missions have more similar patterns of means for the remaining performance categories.

To assess the reliability of the judgments, we computed an intraclass correlation for the ratings. The appropriate model is a three-way ANOVA. The result of this analysis was an ICC of .89 (.57 for one rater), a very respectable level for six judges.

Table 7

Mean Judgments of the Importance of Performance Categories for Missions

	Primary Special Forces Missions				
Performance Categories	Foreign Internal Defense	Unconvention -al Warfare	Direct Action	Counter- terrorism	Special Recon.
A. Teaching Others	5.00	5.00	1.67	2.00	1.50
B. Relations with Indigenous People	5.00	5.00	1.67	2.83	1.67
C. Interpersonal Situations	4.83	4.83	1.83	2.83	1.83
D. Enhancing Language Skills	5.00	5.00	1.83	2.17	2.00
E. Team Effort and Morale	4.50	4.83	5.00	5.00	4.83
F. Initiative and Extra Effort	5.00	5.00	4.67	4.67	4.83
G. Honesty and Integrity	4.83	4.83	4.17	4.33	4.83
H. Planning and Preparing	4.83	5.00	5.00	4.83	4.83
I. Decision Making	4.83	5.00	5.00	4.83	4.83
J. Physical and Environmental Challenges	4.17	4.67	4.33	3.83	4.67
K. Navigating in the Field	4.33	5.00	5.00	4.33	5.00
L. Troubleshooting and Solving Problems	5.00	5.00	4.67	4.67	4.50
M. Safety Conscious	4.50	4.17	3.83	4.00	3.83
N. First Aid, Treating Casualties	4.33	4.67	4.67	4.50	4.83
O. Administrative Duties	3.83	4.17	2.00	2.17	2.17
P. Weapons Skills	5.00	5.00	4.67	4.83	3.67
Q. Engineering Skills	4.50	4.67	4.17	4.00	3.00
R. Communications Skills	4.33	4.83	4.50	4.00	5.00
S. Medic Skills	4.67	4.67	4.33	4.33	4.17
T. Team Leader Skills	4.83	4.83	4.83	4.83	4.83
U. Intelligence Skills	4.33	5.00	4.33	4.50	5.00

As shown in Table 7 and on the graph in Figure 8, all of the SF performance categories are important for FID and UW missions. DA, SR, and CT have a different profile; these missions do not require (A) Teaching Others, (B) Building and Maintaining Relationships with Indigenous People, (C) Handling Interpersonal Situations, and (D) Using Language Skills.

Conclusions

In summary, careful systematic procedures based on prior research methods were used to collect judgments. As in prior efforts, the judgments showed high levels of interrater agreement, and provided important information for constructing the Roadmap which is described in Chapter 3.

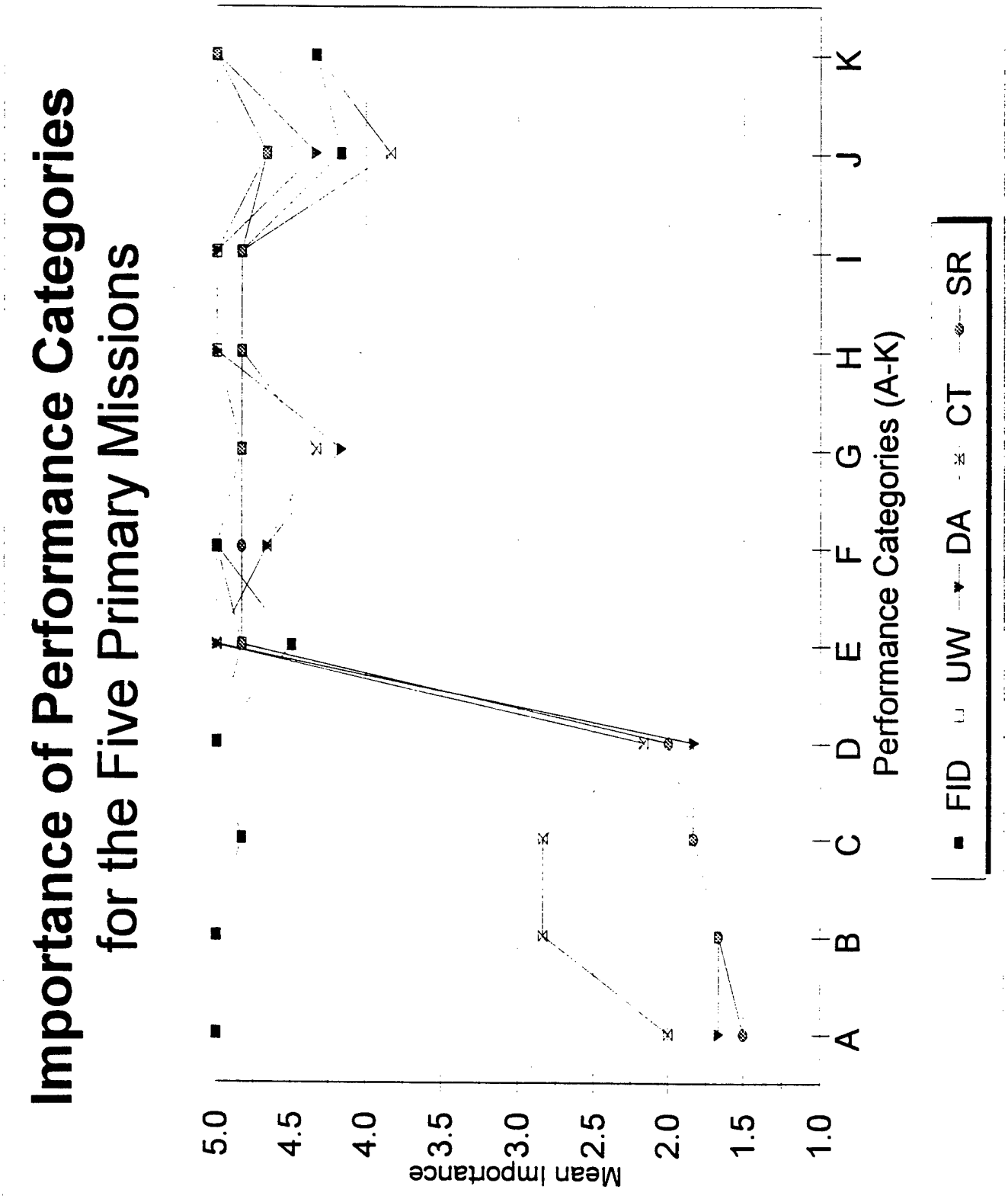

Figure 8. Importance of Performance Categories A-K for the Five Primary Missions

CHAPTER III
ORGANIZATION OF EXPERT JUDGMENTS AND OBSERVATIONS INTO A ROADMAP FOR SELECTION AND CLASSIFICATION RESEARCH

Once the expert judgments were in place, our primary objective was to organize the expert judgment data along with observations and other information from USAJFKSWCS and SF into an agenda--a Roadmap--for SF selection and classification research. This involved four steps:

(1) Gathering information about current and future SF selection directions based on anticipated SF missions,

(2) Using the expert judgment data to identify sets of predictors likely to be useful for SF selection,

(3) Using the expert judgment data to assess the sufficiency of existing criterion measures and identify additional measures, and

(4) Organizing information into projects that will lead to enhancement of SF selection and classification.

Development of Information about Future Trends

We conducted one-on-one interviews with nine officers from the 1st Special Warfare Training Group (SWTG), Special Operations Proponency Office (SOPO), Directorate of Training and Doctrine (DOTD), the 3rd Special Forces Group Airborne (SFG[A]), and the 7th SFG[A] to gather information about trends in SF missions and future directions from key decision-makers in SF and USAJFKSWCS. We asked them questions like *How do you expect SF missions to change in the next decade?, How would those changes affect SF selection and classification? and What changes are already planned for SF selection and training?* Additionally, we read publications focusing on likely mission changes and emphasis (e.g., Boyatt, 1994).

We learned that key decision-makers in SF and USAJFKSWCS expect that their largest primary mission, foreign internal defense (FID), will continue to be the major focus of SF and that other types of missions involving cross-cultural interactions such as humanitarian aid and coalition warfare will grow. Missions without a cross-cultural emphasis such as direct action are expected to diminish (perhaps being handled by the Rangers in the future). Attributes identified during the job analysis as being relevant to building relationships with indigenous people are therefore expected to be even more important to future SF missions.

Some other relevant findings from our discussions were:

- The number of students selected for SFAS and the Q-Course will decrease in the next few years because the staffing requirements of the SFG[A] have been met.

- SF decision makers, particularly those in the field, want selection tools that will identify individuals likely to perform well on the job.

- SWTG decision makers are interested in assessing the validity of current SFAS and Q-Course efforts.

- SWTG decision makers need data and information to make changes in the sequence and content of selection and training activities.

- SF decision makers are interested in client satisfaction but there are a number of obstacles to measuring it. For example, some host nations would be threatened or offended by efforts to involve them in evaluating SF teams, and country team personnel may not have sufficient opportunity to observe team performance.

Development of Predictor Sets

Four principles guided the identification of measures likely to be useful for predicting job performance:

(1) **Quality**--The measures selected for the Roadmap should be of high quality based on expert judgment.

(2) **Feasibility**--Wherever possible, the measures selected for the Roadmap should require minimal development cost.

(3) **Comprehensiveness**--As a whole, the measures selected should measure as many of the attributes needed for successful performance in Special Forces as possible.

(4) **Priority**--Attributes related to the job performance category *B. Building effective relationships with indigenous people* should be covered by at least one selected measure because this performance category is important for future SF missions. Based on data from the job analysis those attributes include: Judgment and Reasoning Ability, Adaptability, Language Ability, Communication, Non-Verbal Communication, Persuasiveness, Maturity, Dependability, Initiative, Motivating Others, Supervising, and Interest in People and in Other Cultures.

Using those principles as decision rules, we examined the expert judgments and identified sets of tests, measures, and scales likely to be useful for SF. This involved three steps:

Step 1: Examining the expert judgment data one-attribute-at-a-time and selecting feasible (readily available) measures that were also rated as useful by experts. Here, available refers to extant measures and forms; data may not necessarily exist for SF personnel on the measures. These measures formed Predictor Set 1.

Step 2: Identifying attributes that were not sufficiently covered by Predictor Set 1. Those attributes were: communication ability, non-verbal communication, intercultural adaptability, and conventional Army task experience and proficiency.

Step 3: Developing predictor sets that include experimental and yet undeveloped measures to cover attributes identified in Step 2. This resulted in Predictor Sets 2 and 3.

The three predictor sets appear in Figure 9. Detailed descriptions of all of the measures in the predictor sets are provided in Appendices B-E. Those descriptions provide research histories and results and define constructs measured by each instrument.

Predictor Set 1--Currently Available, Useful Measures. The measures in Predictor Set 1 are expected to measure a host of job analysis attributes, particularly leadership, temperament, interest, and perceptual and analytic abilities needed for SF jobs. Moreover, most of the cognitive attributes are covered by archival cognitive measures and many non-cognitive attributes are covered by extant measures developed by the Army Research Institute.

The *SF Biographical, Interest, and Temperament Survey* (SFBITS) is an instrument that can be aggregated from scales on the Army Biodata Inventory, Ranger Biodata Inventory, Forced-Choice Assessment of Background and Life Experiences, Army Vocational Interest Career Examination, Job Orientation Blank, and Organizational Identity items. Those instruments have substantial research support and address many attributes that are not already well-measured by archival variables. Specific steps for the aggregation of scales from those instruments to form SFBITS are provided in Appendix M. SFBITS could be used to screen applicants for entry into SFAS, particularly in future years when the number of applicants accepted for SFAS is reduced due to reduced staffing requirements in the SFG[A]. Of course, it could be administered during SFAS and considered along with other SFAS scores in the SFAS graduation decision.

There are two versions of the *Situational Judgment Test* (SJT) either of which could be used. One version was developed during the Army's Project A to serve as a criterion measure of NCO performance. It was administered to thousands of NCOs in the conventional Army. Another version is nearing completion in another Army project-Expanding the Concept of Quality in Personnel (ECQUIP). The ECQUIP SJT is named

Predictor Set 1--Currently Available, Useful Measures

Paper-and-Pencil or Computer-Administered:

SF Biographical, Interest, and Temperament Survey
Situational Judgment Test
Leadership Problems Inventory
Assembling Objects

Computer-Administered:

Target Identification
Target Tracking I and II

Archival Variables:

Armed Services Vocational Aptitude Battery (ASVAB)
Wonderlic
Defense Language Aptitude Battery
SFAS Peer Rankings
SFAS Situation Reaction Exercise Scores
SFAS Swim Test Score
SFAS Physical Endurance Composite
SFAS Physical Fitness Composite
SFAS Military Orienteering Composite
Army Physical Fitness Test
Honors Received in the Q-Course
Total Number of Tries in Ft. Bragg Training (Q-Course)

Predictor Set 2--Content Valid Role Plays

Teaching Role Play
Cultural Adaptability Role Play
Other possibilities: Structured Interview and NCO Role Play

Predictor Set 3--Measures of conventional Army experience and proficiency

Work and Education History Survey--e.g., language fluency, language courses taken, number of awards and certificates, promotion rate, prior MOS, prior training
Conventional Army Self Development Test Scores
Job Compatibility Questionnaire (with five scales--SF, Weapons, Engineering, Communication, and Medic)
Army-Wide NCO Performance Rating Scales

Figure 9. Predictor Sets

the Army Leadership Questionnaire (ALQ) and is expected to measure self-efficacy as well as leadership and other temperament constructs. The *Leadership Problems Inventory* (LPI) is also an ECQUIP product. The LPI is expected to measure the ability to set priorities among leadership and supervisory problems.

Assembling Objects is a spatial ability test developed in the Army's Project A and later used as a part of the Enhanced Computer-Administered Test (ECAT) battery. It exists in both paper-and-pencil and computer-administered forms. Similarly, *Target Identification*, and *Target Tracking I and II* were included in the ECAT battery. Target Identification measures perceptual speed and accuracy while the tracking tests are measures of psychomotor ability.

Some of the variables in Predictor Set 1 are already used by SWC for selection into SFAS or for graduation from SFAS. The archival variables are ones that are currently available in SFAS and Q-Course data bases. They are measures of physical abilities, cognitive abilities, and leadership.

How might Predictor Set 1 be used? With exception of the SFAS and Q-Course variables, the Predictor Set 1 measures could be used either for pre-SFAS screening or during SFAS as a part of the graduation decision. It is possible that USAJFKSWCS will desire more stringent pre-SFAS screening if the applicant pool remains large and the SF staffing requirements level off or decline (i.e., if fewer people are needed). Predictor Set 1 variables could also reduce attrition from SFAS. Recall that SFAS includes endurance, physical fitness, and military orienteering exercises. Research suggests that individuals who perform poorly on spatial tasks also perform poorly on land navigation (Busciglio, Teplitzky, & Welborn, 1991). Also, some of the scales from the Army Biodata Inventory and Ranger Biodata Inventory are likely to predict physical endurance and fitness (see Appendix M). Of course, Predictor Set 1 could be administered during SFAS and considered along with other SFAS scores in the SFAS graduation decision.

Why are there so many measures in Predictor Set 1? As a whole, Predictor Set 1 is designed to cover as much of the predictor domain as possible with existing measures. There is some overlap among the predictors. For example, the FCABLE Dependability scale (which will be a part of SFBITS) is expected to measure the attribute Dependability which was rated as very important for performing effectively in the performance category "Building Effective Relationships with Indigenous Populations" in the job analysis (Russell et al., 1994). To a lesser extent, SFAS Peer Rankings are also expected to measure Dependability. Both are included in Predictor Set 1 because they are two different methods of measuring the attribute, and together they are likely to be better than either measure alone. Also, in any experimental battery there is some uncertainty about how instruments will fare during validation. Reasonable duplication in important areas is wise.

We expect that Predictor Set 1 will take about four hours to administer, perhaps a little less time if it is fully automated. Administering all the predictors in Predictor Set 1 and gathering the archival records from the SFAS data base would provide a wealth of

information about the correlation among measures and allow for refinement of the battery.

Predictor Set 2--Content Valid Job Samples. Predictor Set 2 includes exercises expected to measure communication, non-verbal communication, cultural adaptability, and a number of temperament attributes. The job samples are not currently available. Ideas for developing them appear in their predictor descriptions in Appendix E. The *Teaching Role Play,* for example, would involve developing a list of 20 or so simple, basic soldiering tasks (e.g., knot tying, first aid). SFAS students would be allowed to select a task to teach and would be given a box of materials to use in preparing for their teaching sessions. About three days later each student would spend 15-20 minutes teaching a small group of candidates his selected task. Cadre members would be trained to evaluate communication and interpersonal skills and would observe the session. It would probably be best to administer the job samples during SFAS.

Predictor Set 3--Measures of Conventional Army Task Experience, Proficiency, and Preference. One of the important recent findings from the Army's Building the Career Force Project is that performance during the first tour predicts second tour performance (Campbell, Johnson, & Fellows, 1994; Campbell, Peterson, & Johnson, in press). Taking that one step further, NCO performance should predict training and on-the-job performance in SF jobs. Previous job performance and experience in the conventional Army provides a wealth of virtually untapped information. Indeed, Campbell and his colleagues found that measures of past performance provided *good incremental validity over the Armed Services Vocational Aptitude Battery* for predicting virtually all aspects of performance. The *Work and Education History Survey* (WEHS) would contain items to document background and experience, some of which are already collected by USAJFKSWCS. WEHS could eventually be a weighted application blank with weight given to specific types of experiences. The *Self Development Test* (SDT) has replaced the old Skill Qualification Test as a measure of MOS proficiency and would likely predict proficiency in SF MOS. Here, USAJFKSWCS would not administer an SDT; instead individual's scores on the SDT for their MOS would be collected by self-report in the WEHS. A *Job Compatibility Questionnaire* (Villanova, Bernardin, Johnson, & Dahmus, 1994) would measure preferences for specific types of work activities. MOS-specific scales would be composed of job activities specific to each MOS and would be expected to facilitate MOS assignment. The job analysis data would serve as a starting point, and the remaining development steps for this instrument would not be highly labor intensive. Finally, peer and supervisor ratings on *Army-Wide NCO Rating Scales* developed during Project A for the assessment of NCO performance should be good measures of NCO leadership and effort.

Development of Criterion Sets

As mentioned in Chapter 2, our primary goals in reviewing the criterion expert judgments that appear in Table 6 were:

(1) to determine whether criterion measures were available to sufficiently cover the SF performance categories for both training and on-the-job performance, and

(2) to identify criterion measures that could be used to buttress criterion measurement for specific studies and purposes.

The guiding principle in the examination of the criterion data was to ensure adequate coverage of the criterion domain. As many researchers have noted, the criterion is often neglected. Criteria are often selected simply because they are convenient (e.g., administrative indices or grades) with little regard to what they measure and the comprehensiveness with which they cover important areas of job performance. In the expert judgment exercise reported in Chapter 2, we mapped criteria against the 26 job performance areas defined in the job analysis. That mapping enabled us to examine the adequacy with which available criterion measures cover the SF performance domain.

Figure 10 lists different types of criterion measures, and detailed descriptions of the criterion variables appear in Appendix I. Consistent with predictor discussions, the term "available" means that an instrument has been developed and in some cases data exist in a data file (but not necessarily so). "Supplemental" training criteria require development; neither forms nor archival data exist.

Training Criteria. We began by considering only the training criteria. Available training criteria are listed as Criterion Set 1 in Figure 10. Based on the expert judgment data presented in Table 6, three variables that exist in the Q-Course data base appeared to be somewhat useful for measuring initiative and effort as well as proficiency in MOS-specific performance categories: *Final Training Status*, *Q-Course Honors*, and *Total Number of Tries in Ft. Bragg Training*. *Q-Course Peer Ranking* is expected to measure job performance categories having to do with teamwork and leadership. The two land navigation test scores are relevant to one of the performance categories, *Navigating in the Field*. *MOS Course Grades* are expected to measure MOS-specific skills for NCOs and officers, and *Language School Grades* should measure language proficiency. The remaining two criteria are ones we did not learn about until after completion of the expert judgment exercise. We expect *Intercultural Communication Course Grades* to be relevant to at least two performance categories that involve interpersonal relationships with others. The *Robin Sage ISOGATE Test* is administered during the Robin Sage exercise and is a measure of mission planning and implementation knowledge. In all, these variables tap most of the SF performance categories (i.e., the domain of performance) and would be sufficient for validating Predictor Sets 1, 2, and 3.

The supplemental training criteria listed under Criterion Set 2 are not currently available. We recommend that the Army buttress the existing training criteria (Criterion Set 1) with for-research-only measures that are aligned with the goals and purposes of the study. For example, if the purpose of the study is to evaluate predictors of MOS proficiency, the Army should consider supplementing the operational *MOS Course Grades* with new *Hands-on or Written MOS Proficiency Measures* (see Criterion Set 2) to strengthen measurement of MOS-specific proficiency criteria and to allow for greater performance variation than might be observed in the operational measures. If the purpose of the study is to evaluate predictors of interpersonal and intercultural adeptness, *Peer Rankings* and *Intercultural Communication Course Grades* should be

Criterion Set 1--Available, Useful Training Criteria

Final Training Status
Q-Course Honors
Total Number of Tries in Ft. Bragg Training
Q-Course Peer Ranking
Q-Course Land Navigation Written Test Score
Q-Course Land Navigation Field Test Score
MOS Course Grades
Language School Grades
Intercultural Communication Course Grades
Robin Sage ISOGATE Test Score

Criterion Set 2--Supplemental Training Criteria

For-research-only peer and cadre ratings during MOS training
For-research-only peer, cadre, guerilla chief, and guerilla ratings of Robin Sage performance.
For-research-only briefback ratings for officers.
Hands-on or written MOS proficiency test

Criterion Set 3--Available Job Performance Criteria

Self Development Test Scores
Language Proficiency Scores
Number of Awards, Memoranda, and Certificates
Number of Disciplinary Actions
Promotion Rate
SF-Common Performance Rating Scales
MOS-Specific Performance Rating Scales

Criterion Set 4--Supplemental Job Performance Criteria

Cultural Situation Judgment Test
Automated mission planning simulation
or
Joint-Readiness Training Center (JRTC) observer ratings of individual performance.
Hands-on or written MOS proficiency test

Figure 10. Criterion Sets

supplemented with for-research-only peer and cadre ratings. ARI is in the process of developing a rating form that could be used for this purpose and the MOS-specific and SF-common performance rating scales developed during the job analysis as measures of on-the-job proficiency could be tailored to training performance without much difficulty.

It is important to note here that USAJFKSWCS maintains extensive raw data on students in the SFQC trainee folders. The folders contain behavioral observations, spot reports, ratings, outcomes, and comments that would provide a rich source of information for the development of supplemental criteria. For example, cadre members keep notes and behavioral observations during the Robin Sage exercise. In turn, those notes are placed in the candidate's after action folder. Development of any additional Robin Sage rating or measurement tools should begin with a content analysis of those written observations. A different approach would be to develop a scoring protocol based on SME judgments and review and score observations noted in the folders for SFQC graduates in previous years on relevant performance dimensions. That way scores would be available for individuals who are already in the field.

On-The-Job Performance Criteria. Available job performance criteria formed Criterion Set 3 shown in Figure 10. Recall that available means that forms exist; data files containing the scores do not. The *Self Development Test* (SDT) has replaced the old Skill Qualification Test as a measure of MOS proficiency. Here, USAJFKSWCS would not administer an SDT; instead individual's scores on the SDT for their MOS would be collected by self-report. *Language Proficiency Scores* are obvious measures of language proficiency. As indicated in Table 6, *Number of Awards, Memoranda, and Certificates, Number of Disciplinary Actions, and Promotion Rate* are expected to measure performance areas such as showing initiative and effort, displaying honesty and integrity, and handling interpersonal situations. As with the SDT, language proficiency scores, Number of Awards, Memoranda, and Certificates, Number of Disciplinary Actions, and Promotion Rate would be collected through self-report. We have classified these variables as "available" because the Personnel File Form used in Project A would serve as a draft instrument that could be revised with little time commitment. Those data are also available in the Enlisted Master File (EMF), but research suggests that self-report is as accurate as the EMF and easier to collect (Campbell & Zook, 1990).

For-research-only peer and supervisor ratings on the *SF-Common Performance Rating Scales* and *MOS-Specific Performance Rating Scales* that were developed during the SF job analysis should be good measures of most of the performance categories. The SF-Common Rating scales address performance areas that are common to all positions on the team such as contributing to the team effort, showing initiative and extra effort, displaying honest and integrity. The MOS-Specific rating scales should lend support to other MOS-Specific performance criteria. These scales are available, but no data on SF personnel has yet been collected.

As with the training criteria, the Army should consider supplementing the existing job performance criteria with additional measures depending upon the purposes of the study (Criterion Set 4). The existing criteria are not very strong in the measurement of these performance categories: "Teaching Others", "Building Relationships with Indigenous

Populations," "Mission Planning," "Decision Making," and MOS proficiency. Possible suitable measures are a *Cultural Situation Judgment Test*, *Automated Mission Planning Simulation* or *Joint-Readiness Training Center (JRTC) Ratings* and *Hands-On or Written MOS Proficiency Tests*. (Descriptions of these measures are provided in Appendix I.) *Cultural Situation Judgment Tests* have been used by intercultural communication trainers for years under the name "Cultural Assimilator" (Fiedler, Mitchell, & Triandis, 1971). They present situations based on critical incidents in another culture and ask the respondent to analyze the situation and identify an appropriate approach. An *Automated Mission Planning Simulation* would present a complex mission planning scenario and query respondents about their actions and approach as the problem unfolds. Respondents would make decisions at various levels and the results of those decisions would affect what options were made available to them at each decision point. *JRTC Ratings* would be an alternate measure of mission planning and decision making. The current emphases in exercises run at the JRTC are on training and the team level. Teams perform several cycles of mission planning, isolation, preparation, execution, and after-action review; during this process, they are observed by Observer-Controllers (OCs). Teams are given feedback at the team, not the individual, level. The OCs record qualitative information in "gray books" and everything is videotaped (Dyer, 1994). *JRTC Ratings* would require the development of rating materials for OCs to use in evaluating individual performance.

Development of the Roadmap

We examined the validation requirements for each of the predictor sets and re-examined our discussions with SF and USAJFKSWCS decision-makers to identify projects. The resulting Roadmap is composed of eight projects designed to enhance SF selection and classification. Five of the eight projects are predictor validation steps and the remaining three projects involve the development of tools and information to facilitate decision making at USAJFKSWCS. The eight projects are:

Validation Projects:

Project 1	Concurrent Criterion-Related Validation of Readily Available Predictor Measures Against On the Job Performance.
Project 2	Development and Implementation of Content Valid Job Sample Tests (Role Plays)
Project 3	Validation of Measures of Conventional Army Task Proficiency, Experience and Preference Against Training Performance
Project 4	Validation of Training Performance Against On The Job Performance
Project 5	Predictive Validation of All Predictors Against On The Job Performance

Projects to Develop Decision-Making Tools and Other Information:

Project 6 Development of a Selection and Training Decision Simulator

Project 7 Review of New Measures of Leader Problem Solving Performance

Project 8 Training Performance Study

Figure 11 summarizes the requirements, strengths, and deficiencies of each of the five validation projects.

Project 1: Concurrent Criterion-Related Validation of Readily Available Predictor Measures Against On the Job Performance. Project 1 takes advantage of ARI research on conventional Army jobs to compose a test battery (Predictor Set 1) that measures many attributes that are important for SF job performance. The measures are not content based and would need to be validated in a criterion-related validation strategy. Moreover, documentation of the validity of the paper-and-pencil instruments against on-the-job performance (as opposed to training performance) is a must to ensure their credibility. A concurrent design would allow for quick turn-around of results and would be reasonable given the type of instruments to be used. Criterion Set 3, Available Job Performance Measures, should be used as criteria. As mentioned, we expect Predictor Set 1 to take about 4 hours to administer in its entirety. Rating scales can be administered in one hour or less.

Project 1 would result in a rich data base documenting the relationships among archival SFAS measures, new predictors, and on-the-job performance. Since many of the measures have been pilot-tested on applicants or SFAS candidates, these data would allow comparison across subject populations to estimate the generalizability of results from one population to another. The database from Project 1 could be queried to address a variety of research issues relevant to variables from both the archival data bases (SFAS and Q-Course), none of which have been examined against on-the-job performance in the past. In sum, the data base from this project could be used to build on and fine-tune the existing selection system.

Project 2: Development and Implementation of Content Valid Job Sample Tests (Role Plays). Project 2 could also be conducted to get a product into the field quickly. Together with the paper-and-pencil instruments from Project 1, the role plays should strengthen the measurement of interpersonal skills. The only real drawback is that they could be labor intensive to administer operationally, and it would be important to develop a training program for cadre members. It is possible that two days of testing time would be needed for the job samples.

Project 3: Validation of Measures of Conventional Army Task Proficiency, Experience and Preference Against Training Performance. As mentioned earlier, Predictor Set 3 is expected to predict training proficiency in MOS and language school. The *Work and Education History Survey* (WEHS) (including self-reported Self

Project	Validation Requirements	Strengths	Deficiencies
Project 1--Concurrent, Criterion-Related Validation of Readily Available Measures Against On-the-Job Performance Assemble the SF Biographical, Interest, and Temperament Survey. Collect Predictor Set 1 and Criterion Set 3 data on approximately 80 SF NCOs and 20 SF Officers from each of the five SFG[A]. Participants would need to be people for whom SFAS & Q-Course data exist.	Criterion-related validation is needed. Validation against job performance preferred to validation against training criteria. Concurrent design would be reasonable.	Low development cost. Measures many attributes that are important for job performance. Easy to administer.	Relies solely on experimental items for measurement of intercultural adaptability. Other attributes such as communication ability are not covered.
Project 2--Development and Implementation of Content Valid Job Sample Tests (Role Plays) Develop role plays with SME input. Develop role play manuals, cadre training, & rating materials. Conduct a small sample tryout. Conduct a large-scale pilot test on SFAS students. Implement.	Content validation is sufficient.	Could be implemented quickly. Would measure attributes that are important for job performance and not measured by other instruments.	Development needed. Labor intensive to administer operationally.
Project 3--Validation of Measures of Conventional Army Task Proficiency, Experience, and Preference Against Training Performance. Develop and pilot test proficiency measures (Predictor Set 3). Collect data from SFAS applicants. Validate against existing training criteria (Criterion Set 1).	Criterion-related validation is needed. Criteria must include MOS-specific measures of training proficiency.	Should enhance the fit between individuals and SF, SF MOS, and SF group assignments. Facilitates prediction of technical job proficiency. Buttresses prediction of leadership.	Many logistics to work out, particularly with SF recruiting. Would need to address concerns about accuracy and bias in ratings.
Project 4--Validation of Training Performance Against On The Job Performance Prepare data base of training data (Criterion Set 2). Collect data during Robin Sage, perhaps using ROTC students as guerillas. After graduates have been in the field for at least one year, collect job performance criteria (Set 4).	Criterion-related validation against job performance is needed.	Most of the measures are available now.	Results would not be available until Q-Course graduates have had the opportunity to perform in the field for at least one year. Additional criterion development would be needed.
Project 5--Predictive Validation of All Predictors Against On-the-Job Performance Continue development and enhancement of data bases. Add new variables as projects occur. When at least 500 soldiers who have complete data have been in the field for at least one year, collect Criterion Set 4 data.	Ultimately, predictive validation of all measures against on-the-job performance is desirable.	Allows a bottom-line assessment of the whole selection and training system against on-the-job performance.	Requires extensive data base maintenance. Results of efforts would not become available for several years.

Figure 11. Requirements, Strengths, and Deficiencies of the Five Roadmap Validation Projects

Development Test Scores) and the *Job Compatibility Questionnaire* (JCQ) would need to developed for this project, but development of those measures would be relatively low in cost compared to the job sample exercises. Existing forms and job analysis data are available to form the draft instruments. The *Army-Wide NCO Performance Rating Scales* are available.

Administration of the WEHS and JCT would be simple and straightforward. Both of them are paper-and-pencil instruments that could be administered during or prior to SFAS. If peer and supervisor ratings from the conventional Army were collected as a part of this effort there would be a number of logistics to work out in getting ratings made and returned for processing. Any concerns about accuracy and bias in ratings would need to be addressed to ensure the ratings have credibility. However, this technology of gathering performance ratings from a variety of peers, supervisors, and subordinates, also called 360-degree feedback, is now widely used in industry. Some of the procedures used in industry would probably help address these issues.

Project 4: Validation of Training Performance Against On The Job Performance. One of the projects suggested by USAJFKSWCS decision makers was to assess current training against job performance measures. This would be a predictive validation study where training measures are collected for a few classes and after sufficient numbers of students have been in their field assignments for at least a year, criterion measures would be collected. There are two drawbacks. First, no results would become available for about two years. Second, the training variables used would need to be enhanced to ensure that they reflect the full range of training experiences. At least, the raw information available in personnel file folders would need to be analyzed. We would recommend using supplemental training criteria as the predictors as well as existing measures (Criterion Set 2) and validating them against supplemental on-the-job performance measures (Criterion Set 4).

Project 5: Predictive Validation of All Predictors Against On The Job Performance. Ultimately predictive validation of all measures against on-the-job performance is the true test of a selection system. Economically, the only way that could be accomplished for SF would be to develop and maintain data bases with complete predictor and criterion information and wait until sufficient numbers of individuals reach the field. Range restriction will be a problem if selection instruments such as role plays are used to make decisions, and it would be wise to accumulate enough data to allow for range restriction corrections before implementing measures.

Projects 6-8 are not validation projects. They involve the development of tools and information to facilitate decision making at USAJFKSWCS. They are based on needs and issues that emerged from interviews and from a final review of the expert judgment data.

Project 6: Development of a Selection and Training Decision Simulator. During our discussions with SWTG decision makers, we learned that they are interested in examining the potential impact of their decisions about sequencing selection and training activities. This project would result in a piece of software that would allow SWTG

decision makers to simulate the potential impact of changes in the sequence of selection and training activities (e.g., what happens if we screen on spatial ability before SFAS?). A database would need to be developed based on the covariances of the measures, current pass rates, and other statistics. The database could be developed now to reflect only the measures contained in the current selection system or could be developed after the implementation of new selection measures. A user-friendly software interface or querying procedure would also be needed.

Project 7: Review of New Measures of Leader Problem Solving Performance. A number of tests in development are expected to be highly useful for measuring SF leadership, creativity, and judgment and reasoning attributes (based on the expert judgments). It is possible that those measures would be particularly important for SF officer selection. But, those measures which are under development in an Army officer leadership study will not be available for about two or three more years. At that time, ARI should assess the usefulness of those measures for SF jobs.

Project 8: Training Performance Study. As mentioned earlier, SF and USAJFKSWCS decision makers are interested in client satisfaction, but client satisfaction is particularly difficult to measure for SF. The training performance study proposed as Project 8 would provide an estimate of training gains that host nations can expect from SF involvement. It involves developing a procedure for measuring training gains of individuals trained by SF soldiers. Such a procedure would result in (a) feedback to teams on their training accomplishments and (b) information SF could use to illustrate its training accomplishments to its clients. Here, ARI would develop and administer basic soldiering (move, shoot, communicate) hands-on tests to personnel playing the role of guerillas immediately prior to the Robin Sage exercise. Guerillas would be re-tested at the close of Robin Sage. It would be highly desirable to pre- and post-test a group of Army personnel comparable to the guerillas as a control group.

Recommendations

Projects 1 and 2, *Concurrent Criterion-Related Validation of Readily Available Predictor Measures Against On the Job Performance* and *Development and Implementation of Content Valid Job Sample Tests*, are designed to supplement SF selection and classification with measures of leadership, temperament, and communication and analytic skills. Both projects would provide useful highly useful measures that address many of the SF attributes identified in the job analysis. Based on our understanding of SF and USAJFKSWCS needs and priorities, we recommend that Projects 1 and 2 be conducted concurrently and as soon as possible. As shown in Figure 12, Project 1 will take about 8-12 months, and Project 2 will be shorter, perhaps 6-10 months (to the completion of the draft report). Those two projects together would provide strong measures in areas that are currently not well addressed in the selection system.

After the completion of Projects 1 and 2, it would be reasonable to conduct Projects 3 and 4. Project 3, *Validation of Measures of Conventional Army Task Proficiency, Experience and Preference Against Training Performance*, addresses the fit between individuals and SF jobs and could be conducted within a year's time. Project 4,

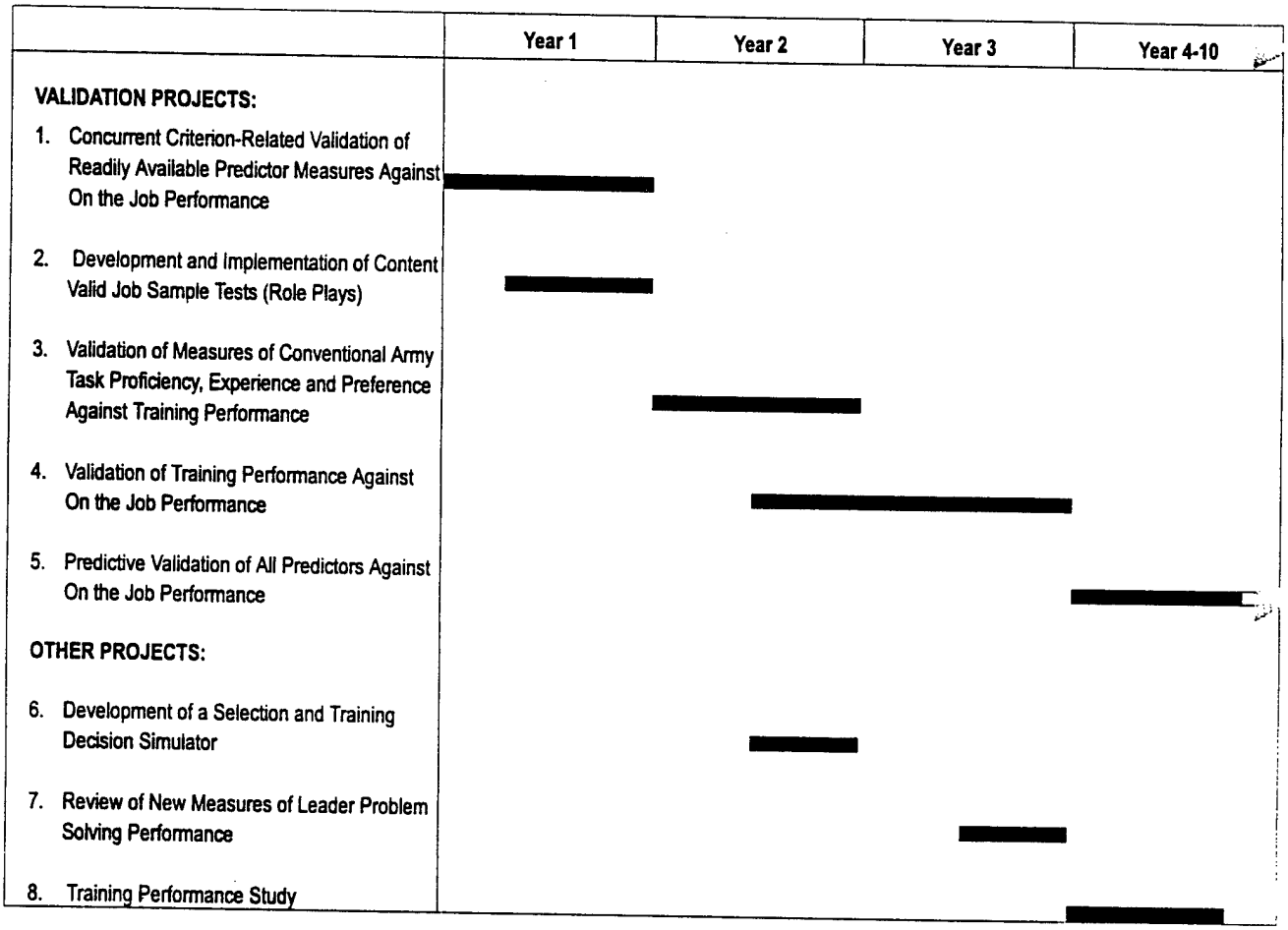

Figure 12. Roadmap Project Timeline

Validation of Training Performance Against On The Job Performance, is of interest to USAJFKSWCS. It would evaluate the usefulness of training data for predicting job performance. Clearly, Projects 3 and 4 build on each other because Project 3 necessitates training criteria, and in Project 4 those criteria become predictors of on the job performance. It would be most efficient to begin Project 3 and then start Project 4 several months into Project 3.

Similarly, Projects 3 and 4 build up to Project 5, *Predictive Validation of All Predictors Against On The Job Performance*. But before starting predictive validation it would be wise to conduct Project 7, *Review of New Measures of Leader Problem Solving Performance*. Recall that leader problem solving measures which are in development in ARI projects could be highly useful to SF, particularly for measure officer attributes. It will be important to consider their potential usefulness again in two or three years--before beginning the predictive validation project.

Projects 6 and 8 could be conducted at any point in time. The *Development of a Selection and Training Decision Simulator* (Project 6) would result in a piece of software that would allow SWTG decision makers to analyze the potential impact of changes in the sequence of selection and training activities. The eighth project, *Training Performance Study*, involves developing a procedure for measuring training gains of individuals trained by SF soldiers. Such a procedure would result in (a) feedback to teams on their training accomplishments and (b) information SF could use to illustrate its training accomplishments to its clients.

Finally, it bears mention that there are two ways to enhance the economic feasibility of the research. First, since the ECQUIP project focuses on NCO leadership and NCOs are the applicant population for enlisted jobs, it makes sense to couple data collections for the two projects wherever possible. Perhaps role plays or SFBITs could be pilot tested along with ECQUIP measures. Second, borrowing the Enhanced Computer-Assisted Test (ECAT) platform, equipment, and software could streamline data collection, minimize test printing and scanning costs, and make database preparation more efficient. The ECAT battery contains several of the tests recommended in Predictor Set 1, and it would be relatively easy to program the remaining tests--*SFBITS*, *SJT*, and *LPI*. All of three of them are verbal (not graphic) tests with straightforward scoring protocols that could easily be programmed in C like the rest of the ECAT battery.

References

Berkhouse, R. G. (1963). Research on combat selection and Special Forces manpower problems--Status report. Arlington, VA: Army Behavioral Evaluation Laboratory.

Berkhouse, R. G., & Cook, K. G. (1961). Development of preliminary screening measures for Special Forces trainees (Research Memorandum 61-7). Arlington, VA: Army Behavioral Evaluation Laboratory.

Berkhouse, R. G., Mendelson, M. A., & Cook, K. G. (1961). Development of performance measures of individual proficiency in Special Forces (Research Memorandum 61-6). Arlington, VA: Army Behavioral Evaluation Laboratory.

Boyatt, M. D. (1994), October). Unconventional operations forces of special operations. Special Warfare, 7(4), 10-17.

Brooks, J. E. (1991). A research needs analysis for U.S. Army Special Operations Forces: Interim report (ARI Research Report No. 1600). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Burke, W. P., & Dyer, F. N. (1984). Physical fitness predictors of success and injury in Ranger training (ARI Research Report No. 1366). Alexandria, VA: U.S. Army Research Institute. (AD A160 288)

Busciglio, H. H., Teplitzky, M. L., & Welborn, C. (1991). Project A spatial tests and military orienteering performance in the Special Forces Assessment and Selection Program (ARI Technical Report No. 921). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A233 432)

Campbell, C. H., Ford, P., Rumsey, M. G., Pulakos, E. D., Borman, W. C., Felker, D. B., De Vera, M. V., & Riegelhaupt, B. J. (1990). Development of multiple job performance measures in a representative sample of jobs. Personnel Psychology, 43(2), 277-300.

Campbell, J. P., Johnson, J., & Fellows, M. (1994, April). The correlation of performance with performance: Criterion dynamics. Paper presented at the ninth annual conference of the Society for Industrial and Organizational Psychology, Inc., Nashville, TN.

Campbell, J. P., Peterson, N. G., Johnson, J. (in press). The prediction of future performance from current performance and from training performance. 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, 1993 fiscal year. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Campbell, J. P., & Zook, L. M. (Eds.). (1990). Improving the selection, classification, and utilization of Army enlisted personnel: Final report on Project A (ARI Research Report No. 1597). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A242 921)

DeMatteo, J. S., White, L. A., Teplitzky, M. L., & Sachs, S. A. (1991). Relationship between temperament constructs and selection for Special Forces training. Paper presented at the 33rd annual conference of the Military Testing Association, San Antonio, TX.

Department of the Army. (1979). Job and task analysis handbook (TRADOC PAM 351-4(t)). Fort Monroe, VA: Department of the Army Training and Doctrine Command.

Department of the Army. (1989a). Mission training plan for the Special Forces company: Special reconnaissance (ARTEP 31-807-31-MTP). Washington, DC: Headquarters Department of the Army.

Department of the Army. (1989b). Mission training plan for the Special Forces company: Direct action (ARTEP 31-807-32-MTP). Washington, DC: Headquarters Department of the Army.

Department of the Army. (1990). Doctrine for Special Forces Operations (FM 31-20). Washington, DC: Headquarters Department of the Army.

Department of the Army. (1991). Military Qualification standards II Special Forces Branch (18) Company Grade Officer's Manual (STP 31-18II-MQS). Washington, DC: Headquarters Department of the Army.

Diana, M. Teplitzky, M. L., & Zazanis, M. M. (1994). Special Forces Qualification Course longitudinal attrition statistics: FY89-FY91 SFAS attendees. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Fiedler, F. E., Mitchell, T., & Triandis, H. C. (1971). The culture assimilator: An approach to cross-cultural training. Journal of Applied Psychology, 55, 95-102.

Flanagan, J. C. (1954). The critical incident technique. Psychological Bulletin, 51, 327-358.

Fricke, J. R. (1990, Winter). The Special Forces Q-Course. Special Warfare, pp. 4-11.

Gorden, F. (1994). Selection and classification for the Army of the Future. In M. G. Rumsey, C. B. Walker, J. H. Harris (Eds.), Personnel selection and classification. Hillsdale, NJ: Erlbaum.

Marder, M., & Medland, F. F. (1964). Construction of experimental aptitude battery for Special Forces enlisted personnel (APRO Research Memorandum 64-9). Washington, DC: U.S. Army Personnel Research Office.

Olmstead, J. A., Caviness, J. A., Powers, T. R., Maxey, J. L., & Cleary, F. K. (1972). Selection and training for small independent action forces: Final report (HumRRO-TR-72-2). Alexandria, VA: Human Resources Research Organization.

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 No. 739). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Peterson, N. G., & Bownas, D. A. (1982). Skills, task structure, and performance acquisition. In M. D. Dunnette, & E. A. Fleishman (Eds.), Human performance and productivity (Vol. 1). Hillsdale, NJ: Erlbaum.

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., Owens-Kurtz, C., Hoffman, R. G., Arabian, J. M., & Whetzel, D. L. (1990). Army Synthetic Validity Project: Report of Phase II results, Volume I (ARI Technical Report No. 892). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A226 355)

Peterson, N. G., Smith, D., Hoffman, R. G., Pulakos, E. D., Reynolds, D., Potts, B. C., Oppler, S. H., & Whetzel, D. L. (1993). Expanding the concept of quality in personnel: Base period final report. Washington, DC: American Institutes for Research.

Pleban, R. J., Allentoff, H. L., & Thompson, T. J. (1989). Preliminary assessment of selected predictors of Special Forces Qualification Course success (ARI Research Report No. 1539). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Pleban, R. J., Thompson, T. J., Valentine, P. J., Dewey, G. I., Allentoff, H. L., & Wesolowski, M. A. (1988). Selection and assessment of Special Forces Qualification Course candidates: Preliminary issues (ARI Research Note 87-30). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Pulakos, E. D., & Borman, W. C. (1985). Development and field test of the Army wide rating scales and the rater orientation and training program (Report No. 715). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Russell, T. L., Crafts, J. L., Tagliareni, F. A., McCloy, R. A., & Barkley, P. (1994). Job analysis of Special Forces jobs (HumRRO FR-PRD-94-19). Alexandria, VA: Human Resources Research Organization.

Schmidt, F. L., Hunter, J. E., Croll, P. R., & McKenzie, R. C. (1983). Estimation of employment test validities by expert judgment. Journal of Applied Psychology, 68, 590-601.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin, 86, 420-428.

Smith, P. C., & Kendall, C. M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. Journal of Applied Psychology, 47, 149-155.

Society of Industrial and Organizational Psychology, Inc. (1987). Principles for the validation and use of personnel selection procedures (3rd ed.). College Park, MD: Author.

Teplitzky, M. L. (1991). Physical performance predictors of success in Special Forces Assessment and Selection (ARI Research Report No. 1606). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Velky, J. L. (1990, Winter). Special Forces assessment and selection. Special Warfare, 3(1), 12-15.

Villanova, P., Bernardin, H. J.,Johnson, D. L., & Dahmus, S. A. (1994). The validity of a measure of job compatibility in the prediction of job performance and turnover of motion picture theater personnnel. Personnel Psychology, 44, 703-742.

Wing, H., Peterson, N. G., & Hoffman, R. G. (1984, August). Expert judgments of predictor-criterion validity relationships. Paper presented at the 92nd Annual American Psychological Association Convention.

Wise, L. L., Peterson, N. G., Hoffman, R. G., Campbell, J. P., & Arabian, J. M. (1990). Amry synthetic validation project: Phase III results (RS-WP-90). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Appendix A
Special Forces Assessment and Selection Data Analysis

The Special Forces Assessment and Selection (SFAS) data base includes hundreds of individual variables that are very specific, too many to include in the expert judgment exercises as individual variables. We, therefore, conducted some preliminary analyses to reduce the variable set using data sets from 1990-1993.

The primary goal of the analyses was to identify a reasonable set of variables for inclusion in the expert judgment exercise. Also, it is important to consider the frame of reference for the ROADMAP project. The use of SFAS scores we were most concerned with is in a validation study that includes individuals who graduated in SFAS in different years. We needed variables that could be used consistently across courses and years. We focused on the situation reaction exercise variables.

Situation Reaction Exercises

Overview. The situation reaction exercises are a series of job simulations wherein teams of cohorts in SFAS are assigned missions. Individuals are rated by SFAS cadre. In 1993, the Army Research Institute implemented a major SFAS assessor training program. Raters use a three point scale to make their ratings. But they only complete the ratings if they observe behaviors they believe reflect either Outstanding "3" or Unsatisfactory "1" performance.

The variables in the SFAS data base are counts of the number of outstanding and unsatisfactory ratings on several dimensions:

Common Dimensions:

- Motivation
- Responsibility
- Stability
- Intelligence
- Physical Fitness
- Trustworthiness
- Teamwork

Leadership Dimensions:

- Communication
- Judgment
- Influence
- Decisiveness

Each SFAS candidate serves as a team leader on at least one SR, and he is rated on the dimensions during his service as team leader. Cadre members may also rate the candidate when he is not in a leadership role; those ratings are tallied in separate variables in the SFAS data base.

Analyses. We decided to try out two methods of constructing scores. The first method included only the ratings made while an individual was the team leader. The second method included all of the ratings made for an individual, regardless of the exercise. Both methods involved subtracting the number of unsatisfactory ratings from the number of outstanding ratings to form an overall score for each dimension as the first step (e.g., motivation=number of outstanding motivation ratings - number of satisfactory motivation ratings).

After forming the overall scores, we computed correlation matrices of the overall scores for 1989, 1991, 1992, and 1993 SFAS data. We factored SR data from 1989, 1991, 1992, and 1993 SFAS classes to learn more about the empirical relationships among the variables. We used principal factoring with varimax rotation and selected two factor solutions for all the years (although we did examine three factor solutions for '92 and '93). We found that the overall scores which included all of the ratings, regardless of exercise were more interpretable than those that did not. For 1989, the correlation matrix for the leadership-only ratings was indeterminant. We, therefore, chose to form our composites based on the tallies across all the exercises.

Results. Figure A-1 shows the results of the factor analyses for 1989, 1991, 1992, and 1993 SFAS data. Each of the factor solutions explains about 30 to 40 percent of the total common variance. That is relatively low, but it is also to be expected given that these are operational data assembled across several courses for each year. There are several sources of error in this type of operational data (e.g., change of cadre members across courses) making it all that more important to find aggregate composite scores that take advantage of the reliable score variance.

There is some evidence that the SR ratings could be measuring two or three underlying factors: (1) effort and dependability, (2) judgment, and (3) physical fitness. The major findings include:

Principal Factor Analysis of 1989 SR Data			
Factor 1	Factor 2	Communality	Variable
79*	17	652	Teamwork
78*	28	684	Motivation
65*	11	432	Physical Fitness
41*	34*	287	Responsibility
30*	10	102	Stability
15*	14*	041	Trustworthiness
13	65*	443	Decisiveness
37	59*	485	Influence
15	58*	355	Judgment
18	57*	362	Communication
12	50*	262	Intelligence
29%	8%		Percent of Common Variance in the Unrotated Factor Solution

Principal Factor Analysis of 1991 SR Data			
Factor 1	Factor 2	Communality	Variable
91*	-18	856	Motivation
86*	-22	791	Teamwork
84*	-15	734	Physical Fitness
69*	29	562	Responsibility
48*	-15	248	Communication
41*	-34*	290	Influence
41*	10	178	Intelligence
21*	01	046	Stability
15*	09	030	Trustworthiness
-01	89*	791	Judgment
-02	82*	677	Decisiveness
32%	15%		Percent of Common Variance in the Unrotated Factor Solution

Figure A-1. Factor Analyses of Situation Reaction Exercise Overall Scores

Principal Factor Analysis of 1992 SR Data			
Factor 1	Factor 2	Communality	Variable
86*	34	858	Motivation
80*	06	642	Teamwork
76*	21	615	Responsibility
36*	-10	142	Decisiveness
36*	-23	184	Intelligence
35*	04	127	Influence
18*	-02	033	Stability
10*	-05	013	Communication
40	82*	834	Physical Fitness
-22	70*	537	Trustworthiness
02	70*	484	Judgment
26%	14%		Percent of Common Variance in the Unrotated Factor Solution

Principal Factor Analysis of 1993 SR Data			
Factor 1	Factor 2	Communality	Variable
90*	-09	858	Motivation
55*	16	642	Teamwork
50*	24	615	Responsibility
49*	18	142	Decisiveness
34*	28*	184	Intelligence
20*	03	127	Influence
20*	02	033	Stability
16*	06	013	Communication
-51*	69*	834	Physical Fitness
31	51*	537	Trustworthiness
07	22*	484	Judgment
20%	9%		Percent of Common Variance in the Unrotated Factor Solution

Figure A-1. Factor Analyses of Situation Reaction Exercise Overall Scores (Continued)

- Factor 1, Effort and Dependability. Four dimensions consistently defined the first factor across all years: Responsibility, Motivation, Teamwork, and Stability, although Stability had very low communalities and loadings (low variance).

- Factor 2, Judgment. Judgment consistently defined the second factor across all years.

- Physical Fitness. Physical fitness loaded on factor 1 for 3 years and on factor 2 for one year. It could be either combined with the variables in Factor 1 or left on its own. We recommend leaving it separate is reasonable since it represents a domain of abilities somewhat distinct from the others in factor 1.

- Communication and Intelligence loaded with Judgment in the 1989 data and on factor 1 for all other years. They had very low to moderate communalities across the years. We recommend pooling them with the Factor 1 data.

- Three out of four years Decisiveness loaded with Judgment, in two instances with very strong loadings. We recommend pooling Decisiveness with Judgment.

- The Trustworthiness variable appears to mean different things across the years. For two years it had split loadings across the factors and very low communalities. One year it loaded strongly on factor 1 and another year it loaded strongly on factor 2. We recommend dropping Trustworthiness from analyses across years.

- Influence also shifts over years. We recommend dropping it from analyses across years.

Appendix B
Cognitive Measures

Operational: Armed Services Vocational Aptitude Battery (ASVAB) Introduction

The Armed Services Vocational Aptitude Battery (ASVAB), the Services' primary selection and classification tool, is a highly useful general purpose cognitive predictor. ASVAB scores (i.e., subtest scores, composites, or the ASVAB general factor scores) are valid predictors of training performance (Earles & Ree, 1992; Ree & Earles, 1991, 1992; Welsh, Kucinkas, & Curran, 1990). The ASVAB predicts training success in a host of schools, for a variety of jobs, and in all the Services (Welsh et al., 1990). Recent research has demonstrated the usefulness of the ASVAB for predicting job performance; ASVAB scores are good predictors of both first- and second-tour job performance (McHenry, Hough, Toquam, Hanson, & Ashworth, 1990; Oppler, Peterson, & Russell, 1993; Peterson & Rosse, 1992).

Short Description of Test:

The ASVAB that has been administered since 1980 includes ten subtests, eight of which are power tests and two of which are speeded (i.e., CS and NO) (Welsh et al., 1990).

Subtest	# of items	Testing time
General Science (GS)	25	11
Arithmetic Reasoning (AR)	30	36
Word Knowledge (WK)	35	11
Paragraph Comprehension (PC)	15	13
Numerical Operations (NO)	50	3
Coding Speed (CS)	84	7
Auto & Shop Information (AS)	25	11
Mathematics Knowledge (MK)	25	24
Mechanical Comprehension (MC)	25	19
Electronics Information (EI)	20	9

The ASVAB is currently administered via paper and pencil, but a computer adaptive version has been used experimentally and is under consideration for implementation. It is also possible that a spatial test (probably Assembling Objects) will be added to the ASVAB before the year 2000.

Subtest Intercorrelations: 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., 1990a), 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.

1. Operational: ASVAB General Science (GS)

Construct Measured:

Knowledge of the physical and biological sciences.

Short Description of Test:

This test asks basic questions about biological and physical sciences. It has 25 items and an 11 minute time limit.

Sample: A rose is a kind of:

a. animal.
b. bird.
c. flower.
d. fish.

Psychometrics:

Scoring: The score is the number correct. Number correct scores are standardized to T-Scores

Subgroup Differences: Effect sizes (standardized mean differences) for gender and race/ethnic differences are shown below (Russell, Reynolds & Campbell, 1994). Positive effect sizes indicate that male or white means are higher, while negative effect sizes indicate that Female, Black or Hispanic means are higher.

	Male/Female	W/B	W/H
General Science (GS)	0.36	1.24	1.00

Reliability:

	Alt. Forms	Int. Consis.
General Science (GS)	.83	.86

Validity Evidence: 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. (1990b) meta-analyzed available subtest validities for ASVAB forms that are currently in use (N was greater than 52,000). The corrected-for-range-restriction validity was .64 for GS for predicting school grades. The standard deviation of the average validity for GS was relatively large suggesting differences across studies, Services, and/or jobs in absolute levels of validity.

Ree and Earles (1992) 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. (1990b). The average validity for GS was .66.

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. GS (r=.50) was one of the best predictors.

2. Operational: ASVAB Arithmetic Reasoning (AR)

Construct Measured:

Arithmetic Reasoning

Short Description of Test:

Contains 30 word problems emphasizing mathematical reasoning. It has a time limit of 36 minutes.

Sample: A student bought a sandwich for 80 cents, milk for 20 cents, and pie for 30 cents. How much did the meal cost?
a. $1.00
b. $1.20
c. $1.30
d. $1.40

Psychometrics:

Scoring: The score is the number correct. Number correct scores are standardized to T-Scores

Subgroup Differences: Effect sizes (standardized mean differences) for gender and race/ethnic differences are shown below (Russell, Reynolds & Campbell, 1994). Positive effect sizes indicate that male or white means are higher, while negative effect sizes indicate that Female, Black or Hispanic means are higher.

Subtest	Male/Female	W/B	W/H
Arithmetic Reasoning (AR)	0.28	1.16	0.85

Reliability:

	Alt. Forms	Int. Consis.
Arithmetic Reasoning (AR)	.87	.91

Validity Evidence: 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. (1990b) meta-analyzed available subtest validities for ASVAB forms that are currently in use (N was greater than 52,000). The corrected-for-range-restriction validity was .64 for AR for predicting school grades.

Ree and Earles (1992) 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. (1990b). The average validity for AR was .68.

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, AR (r=.44).

3. Operational: ASVAB Word Knowledge (WK)

Construct Measured:

Vocabulary--ability to understand the meaning of words.

Short Description of Test:

Respondents are asked to select an alternative word whose meaning is most nearly the same as the meaning of a work underlined in a phrase. There are 35 items and an 11 minute time limit.

Sample: It was a small table.
a. sturdy
b. round
c. little
d. cheap

Psychometrics:

Scoring: The score is the number correct. Number correct scores are standardized to T-Scores

Subgroup Differences: Effect sizes (standardized mean differences) for gender and race/ethnic differences are shown below (Russell, Reynolds & Campbell, 1994). Positive effect sizes indicate that male or white means are higher, while negative effect sizes indicate that Female, Black or Hispanic means are higher.

Subtest	Male/Female	W/B	W/H
Word Knowledge (WK)	-.01	1.29	1.00

Reliability:

	Alt. Forms	Int. Consis.
Word Knowledge (WK)	.88	.92

Validity Evidence: 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. (1990b) meta-analyzed available subtest validities for ASVAB forms that are currently in use (N was greater than 52,000). The corrected-for-range-restriction validity was .63 for WK for predicting school grades.

Ree and Earles (1992) 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. (1990b). The average validity for WK was .66.

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, WK (r=.46).

4. Operational: ASVAB Paragraph Comprehension (PC)

Construct Measured:

Paragraph Comprehension

Short Description of Test:

This test requires the examinee to read a paragraph and answer questions about it. Fifteen items are presented with a 13 minute time limit.

Sample: The duty of the lighthouse keeper is to keep the light burning no matter what happens, so that ships will be warned of the presence of dangerous rocks. If a shipwreck should occur near the lighthouse, even though he would like to aid in the rescue of its crew and passengers, the lighthouse keeper must
a. stay at his light.
b. rush to their aid.
c. turn out the light.
d. quickly sound the siren.

Psychometrics:

Scoring: The score is the number correct. Number correct scores are standardized to T-Scores

Subgroup Differences: Effect sizes (standardized mean differences) for gender and race/ethnic differences are shown below (Russell, Reynolds & Campbell, 1994). Positive effect sizes indicate that male or white means are higher, while negative effect sizes indicate that Female, Black or Hispanic means are higher.

Subtest	Male/Female	W/B	W/H
Paragraph Comprehension (PC)	-.19	1.07	.89

Reliability:

	Alt. Forms	Int. Consis.
Paragraph Comprehension (PC)	.72	.81

Validity Evidence: 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. (1990b) meta-analyzed available subtest validities for ASVAB forms that are currently in use (N was greater than 52,000). The corrected-for-range-restriction validity was .64 for PC for predicting school grades.

Ree and Earles (1992) 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. (1990b). The average validity for PC was .62.

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, PC (r=.40).

5. Operational: ASVAB Numerical Operations (NO)

Construct Measured:

Ability to work basic math problems quickly. [It typically loads with Coding Speed in factor solutions, so speededness is an important aspect of the test.]

Short Description of Test:

NO is a speeded test of four arithmetic operations (addition, subtraction, multiplication, & division). It contains 50 items and has a three minute time limit.

Sample: 3 X 3
a. 1
b. 6
c. 9
d. 12

Psychometrics:

Scoring: The score is the number correct. Number correct scores are standardized to T-Scores

Subgroup Differences: Effect sizes (standardized mean differences) for gender and race/ethnic differences are shown below (Russell, Reynolds & Campbell, 1994). Positive effect sizes indicate that male or white means are higher, while negative effect sizes indicate that Female, Black or Hispanic means are higher.

Subtest	Male/Female	W/B	W/H
Numerical Operations (NO)	-.19	.94	.70

Reliability:

	Alt. Forms
Numerical Operations (NO)	.70

Validity Evidence: NO typically yields validities that are somewhat lower than those for most of the other ASVAB subtests. Welsh et al. (1990b) meta-analyzed available subtest validities for ASVAB forms that are currently in use (N was greater than 52,000). The corrected-for-range-restriction validity was .49 for NO for predicting school grades.

Ree and Earles (1992) 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. (1990b). The average validity for NO was .51.

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, NO (r=.29).

6. Operational: ASVAB Coding Speed (CS)

Construct Measured:

Coding Speed

Short Description of Test:

This test provided a reference list of 100 words matched with four-digit code numbers. The respondent is to select the correct code number for each of 84 words administered under speeded conditions. The respondent is required to use the key at the top of the page which lists the code words with the associated code numbers, and then they are to review the sample words and find the alternative which lists the correct code number.

CS is speeded; it contains 84 items and has a 7 minute time limit.

Psychometrics:

Scoring: The score is the number correct. Number correct scores are standardized to T-Scores

Subgroup Differences: Effect sizes (standardized mean differences) for gender and race/ethnic differences are shown below (Russell, Reynolds & Campbell, 1994). Positive effect sizes indicate that male or white means are higher, while negative effect sizes indicate that Female, Black or Hispanic means are higher.

Subtest	Male/Female	W/B	W/H
Coding Speed (CS)	-.42	.96	.60

Reliability:

	Alt. Forms
Coding Speed (CS)	.73

Validity Evidence: CS has consistently yielded the lowest validities of the ASVAB subtests. In the Welsh et al. (1990b) meta-analysis, the corrected-for-range-restriction validity was .44 for CS for predicting school grades. In the Ree and Earles (1992) study, the average corrected-for-range-restriction validity for predicting final school grades for CS was .47. Maier and Mayberry (1989) reported ASVAB subtest validities (corrected-for-range-restriction) for predicting hands-on performance in the infantry rifleman job, CS (r=.26).

7. Operational: ASVAB Auto/Shop Information (AS)

Construct Measured:

Knowledge of auto mechanics, shop practices, and tool functions.

Short Description of Test:

AS contains 25 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.

Psychometrics:

Scoring: The score is the number correct. Number correct scores are standardized to T-Scores

Subgroup Differences: Effect sizes (standardized mean differences) for gender and race/ethnic differences are shown below (Russell, Reynolds & Campbell, 1994). Positive effect sizes indicate that male or white means are higher, while negative effect sizes indicate that Female, Black or Hispanic means are higher.

Subtest	Male/Female	W/B	W/H
Auto Shop Information (AS)	1.25	1.23	.82

Reliability:

	Alt. Forms	Int. Consis.
Auto Shop Information (AS)	.83	.87

Validity Evidence: 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. (1990b) meta-analyzed available subtest validities for ASVAB forms that are currently in use (N was greater than 52,000). The corrected-for-range-restriction validity was .49 for AS for predicting school grades.

Ree and Earles (1992) 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. (1990b). The average validity for AS was .52.

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, AS (r=.50).

8. Operational: ASVAB Mathematics Knowledge (MK)

Construct Measured:

Knowledge of algebra, geometry and fractions.

Short Description of Test:

This test contains 25 items and has a 24 minute time limit.

Sample: The area of a rectangle 2 feet by 3 feet is equal to
a. 2 square feet.
b. 4 square feet.
c. 6 square feet.
d. 8 square feet.

Psychometrics:

Scoring: The score is the number correct. Number correct scores are standardized to T-Scores

Subgroup Differences: Effect sizes (standardized mean differences) for gender and race/ethnic differences are shown below (Russell, Reynolds & Campbell, 1994). Positive effect sizes indicate that male or white means are higher, while negative effect sizes indicate that Female, Black or Hispanic means are higher.

Subtest	Male/Female	W/B	W/H
Math Knowledge (MK)	.14	.88	.73

Reliability:

	Alt. Forms	Int. Consis.
Math Knowledge (MK)	.84	.87

Validity Evidence: 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. (1990b) meta-analyzed available subtest validities for ASVAB forms that are currently in use (N was greater than 52,000). The corrected-for-range-restriction validity was .63 for MK for predicting school grades.

Ree and Earles (1992) 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. (1990b). The average validity for MK was .65.

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, MK (r=.38).

9. Operational: ASVAB Mechanical Comprehension (MC)

Construct Measured:

Ability to understand mechanical principles, such as gears, levers, pulleys, and hydraulics.

Short Description of Test:

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 objects would feel the coldest if all are the same temperature. MC contains 25 items and has a 19 minute time limit.

Psychometrics:

Scoring: The score is the number correct. Number correct scores are standardized to T-Scores

Subgroup Differences: Effect sizes (standardized mean differences) for gender and race/ethnic differences are shown below (Russell, Reynolds & Campbell, 1994). Positive effect sizes indicate that male or white means are higher, while negative effect sizes indicate that Female, Black or Hispanic means are higher.

Subtest	Male/Female	W/B	W/H
Mechanical Comprehension	.83	1.20	.83

Reliability:

	Alt. Forms	Int. Consis.
Mechanical Comprehension	.78	.85

Validity Evidence: 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. (1990b) meta-analyzed available subtest validities for ASVAB forms that are currently in use (N was greater than 52,000). The corrected-for-range-restriction validity was .58 for MC for predicting school grades.

Ree and Earles (1992) 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. (1990b). The average validity for MC was .59.

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, MC (r=.51).

10. Operational: ASVAB Electronics Information (EI)

Construct Measured:

Knowledge of electronics and radio principles.

Short Description of Test:

EI presents items either verbally or pictorially and evaluates the applicants' electronics knowledge. It contains 20 items with a 9 minute time limit.

Sample: What does the abbreviation AC stand for?
a. additional charge
b. alternating coil
c. alternating current
d. ampere current

Psychometrics:

Scoring: The score is the number correct. Number correct scores are standardized to T-Scores

Subgroup Differences: Effect sizes (standardized mean differences) for gender and race/ethnic differences are shown below (Russell, Reynolds & Campbell, 1994). Positive effect sizes indicate that male or white means are higher, while negative effect sizes indicate that Female, Black or Hispanic means are higher.

Subtest	Male/Female	W/B	W/H
Electronics Information (EI)	.78	1.22	.92

Reliability:

	Alt. Forms	Int. Consis.
Electronics Information (EI)	.72	.81

Validity Evidence: 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. (1990b) meta-analyzed available subtest validities for ASVAB forms that are currently in use (N was greater than 52,000). The corrected-for-range-restriction validity was .60 for EI for predicting school grades.

Ree and Earles (1992) 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. (1990b). The average validity for EI was .61.

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, EI (r=.52).

11. Published: Wonderlic Personnel Test and Scholastic Level Exam

Construct Measured:

General Cognitive Ability

Short Description of Test:

The Wonderlic measures the level at which individuals learn and understand instructions, and solve problems, by asking a series of questions including word comparisons, disarranged sentences, sentence parallelism, direction following, number comparisons, number series, analysis of geometric figures and story problems requiring either math or logic solutions. The test questions are arranged in order of difficulty, with the most difficult items appearing at the end of the test (Wonderlic Users Guide, 1992).

Number of Items: 50 Time Limit: 12 minutes
Apparatus: Paper and pencil (available in computerized format)

Psychometrics:

Scoring: The score is the total number of correct answers.

Correlations with other constructs: There is substantial evidence that the Wonderlic is a good measure of g. The Wonderlic Users Guide cites correlations with the Weschler Adult Intelligence Scale ranging from .75 to .96 for the WAIS-R and from .85 to .91 for the WAIS. It also correlates with the Otis-Lennon Mental Ability Test .83 to .99 (N=22 to 561).

Subgroup Differences: Blacks tend to score about 1 Sd below whites; Hispanics score about .84 Sd below whites, across three studies from 1970-1992. Females' scores are typically comparable to males' scores.

Reliability: The User's Guide cites internal consistency reliability estimates of .88 - .94 (McKelvie, 1989); KR-20 r=.88. Test-Retest reliabilities range from .82 to .94.

Jobs used for in the past: Minimum passing scores, or cut scores have been calculated for 71 occupations which vary from maid/matron, security guard, receptionist to general manager, and chemist. The U.S. Department of Labor also provides listings of occupational titles and job descriptions along with a measure of job complexity which has been found to correlate with scores on the Wonderlic for 134 job titles (except for the strength scale).

Validity Evidence: There are hundreds of studies analyzing the predictive validity of the Wonderlic, Hunter and Hunter (1984) summarized this research in a meta-analysis showing validities of .63 with ability, .33 with college grades, .33 with biodata, .27 with education. Other validity studies looking at positions in business settings, engineering, professional positions and vocational training programs, validities ranged from .26 to .67.

The Wonderlic has also been found to correlate with success in the SFQC .29 (N=293) (Pleban, Allentoff, & Thompson, 1989). They found that the Wonderlic was significantly correlated with SFQC criteria (N=188-282): Map Exam (r=.52); Land Nav Field Training Exercise (r=.28); Patrolling written exam (r=.31).

12. Experimental: Project A MAP Test (MP)

Construct Measured:

Spatial Orientation--This test measures one's ability to "appreciate one's location relative to land marks in the environment" (Peterson et. al., 1987).

Short Description of Test:

Subjects are given a map with various landmarks such as a campsite, a forest, a lake, and so on. Several items refer to each map, within each item, subjects are provided compass directions by indicating the direction from one landmark to another (e.g., "the forest is North of the campsite") and they are informed of their present location on the map. Given this information, the subject must determine which direction to take to reach another landmark.

Number of Items: 20 Time Limit: 12 minutes
Apparatus: Paper and pencil

Psychometrics:

Scoring: The score is the total number of correct answers.

Correlations with other constructs: The Map test correlates with Assembling Objects r=.50, .52; Object Rotation r=.39, .42; MAZE r=.44, .42; Orientation r=.53, .54; Reasoning r=.52, .51 all N's=9332, 6941 respectively. Factor analytic research including the Map Test suggests that it represents a first order Orientation factor and loads highly on a second order general spatial factor. Busciglio & Teplitzky (1994) found the MAP correlated with MAZE r=.48; Orientation r=.52; and the Wonderlic r=.66 with N=232.

Subgroup Differences: Whites tend to score 1 SD higher than blacks (large sample effect sizes range from .98 to 1.08). Whites score .4 to .6 SD higher than Hispanics. Males tend to score higher than females with effect sizes (standardized mean differences) between .28 to .30.

Reliability: Cronbach's alpha: .88 (N=6754); .89 (N=9332); .90 (N=290). Test-Retest Reliability: .78 (N=499); .84 (N=97).

Practice Effects: Test performance on spatial ability tests is to some degree malleable; test scores improve with practice (Lohman, 1988). However, the gains are not substantially larger than those observed for tests of other abilities (Russell et al., 1994). There is also some evidence that gains from practice are larger for speeded tests than for power tests (Dunnette, Corpe, & Toquam, 1987). Gains from practice on the Map test have been low in two studies. With a one week interval between testing sessions (N=100), subjects' scores went up .08 sd from testing 1 to testing 2 (Peterson, 1987). With one month between testing sessions (N=473) subjects' scores again went up .08 sd from testing 1 to testing 2 (Toquam, Peterson, Rosse, Ashworth, Hanson, & Hallam, 1986).

Continued Next Page

Validity Evidence: In Project A, McHenry et al. (1990) combined six Project A spatial tests to form one composite score. The spatial score yielded modest incremental validity (beyond that afforded by the ASVAB) for predicting technical proficiency in Army enlisted MOS and hands-on performance. Similar results were obtained for a longitudinal validation sample.

Busciglio & Teplitzky (1994) found that the MAP test is predictive of performance in the SFQC land navigation exercises, adding unique variance over other variables in predicting success in this exercise. They found the MAP test to be the best single predictor of first time land navigation success ($F=7.97$, $p<.01$).

Busciglio & Teplitzky (1990) also found the MAP test to predict success in SFAS military orienteering exercises. They found that high MAP scores are related to higher ratings and less time needed to complete the military orienteering exercises. Ratings on the Early (Task I Day and Night and Task II Day) and Later (Task II Night, Task III and Task IV) Orienteering scores are correlated with the Map Test .31, .23; $p<.0001$ respectively. The time for the Early Orienteering scores is related to the Map test -.24 $p<.0001$.

13. Experimental: Project A Assembling Objects (AO)

Construct Measured:

General Spatial Ability--Spatial Visualization

Short Description of Test:

Subjects visualize how an object will look when its part are put together or assembled according to instructions. In part one, the items in the picture are labeled with letters and the subject must visually put the parts together according to the letters. In part two, pieces in the pictures fit together like a puzzle. Subjects must determine which figure from 4 alternatives is the correct shape when the parts are all put together.

Number of Items: 36 Time Limit: 18 minutes
Apparatus: Paper and pencil (Computer administered version is also available)

Psychometrics:

Scoring: The score is the total number of correct answers.

Correlations with other constructs: Assembling Objects correlates with Object Rotation r=.41, .46; MAZE r=.51, .51; Orientation r=.46, .50; Reasoning r=.56, .56 Map test r=.50; .52 all N's=9332, 6941 respectively (Peterson, Russell et al., 1990). Teplitzky found that Assembling Objects correlated with MAP r=.43; and the Wonderlic r=.32 with N=483.

Factor analytic research suggests that Assembling Objects is a good marker test for general spatial ability (Russell, Humphreys, Peterson, & Rosse, 1992).

Subgroup Differences: Gender differences tend to be rather small with effect sizes ranging from -.02 to .08 in large samples (Peterson, Russell et al. 1990). Whites tend to score higher than African Americans with effect sizes ranging from .78 to .83. Whites tend to score higher than Hispanics with effect sizes .15, .24, and .25 (Peterson, Russell et al. 1990).

Reliability: Cronbach alphas of .88 (N=6754); .90 (N=9332); .92 (N=290). Test-Retest Reliability: .70 (N=499); .74 (N=97).

Practice and Coaching Effects: Test performance on spatial ability tests is to some degree malleable; test scores improve with practice (Lohman, 1988). However, the gains are not substantially larger than those observed for tests of other abilities (Russell et al., 1994). There is also some evidence that gains from practice are larger for speeded tests than for power tests (Dunnette, Corpe, & Toquam, 1987). Gains from practice on the Assembling Objects test have been low in two studies. With a one week interval between testing sessions (N=100), subjects' scores went up .08 sd from testing 1 to testing 2 (Peterson, 1987). With one month between testing sessions (N=473) subjects' scores again went up .06 sd from testing 1 to testing 2 (Toquam, Peterson, Rosse, Ashworth, Hanson, & Hallam, 1986). Busciglio and Palmer (1992) studied the effects of practice and coaching on three spatial tests, one of which was Assembling Objects. Practice effects were significant for all three tests. The effects of coaching on Assembling Objects were negligible.

Continued Next Page

Validity Evidence: In Project A, McHenry et al. (1990) combined six Project A spatial tests to form one composite score. The spatial score yielded modest incremental validity (beyond that afforded by the ASVAB) for predicting technical proficiency in Army enlisted MOS and hands-on performance. Similar results were obtained for a longitudinal validation sample.

Mayberry and Hiatt (1990) administered the ASVAB Form 6 Space Perception, ECAT Figural Reasoning, 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.

Published: Test of Adult Basic Education (TABE) Introduction

Construct Measured:

Educational achievement in language, reading and mathematics.

Short Description of Test:

The TABE has 7 subtests assessing three major areas: Language, Math, and Reading. It has four achievement levels, the highest of which is oriented to those individuals with about 8.6 to 12.9 years of school.

Subtest:	Number of Items/Minutes	Shortened Form Items/Minutes
Reading		
Vocabulary	30 / 17	15 / 8
Comprehension	40 / 37	15 / 14
Mathematics		
Math. Computation	48 / 43	15 / 14
Concepts & Appl.	40 / 37	15 / 14
Language		
Mechanics	30 / 15	15 / 7
Expression	45 / 41	15 / 14
Spelling	30 / 13	---

Apparatus: Paper and pencil

Correlations with other constructs: Total TABE with the GED r=.64; with Air Force Reading Abilities Test - Vocabulary r=.57; Comprehension r=.50
The TABE was correlated with GED subtest scores - the two tests were taken within 6 weeks of each other (N=678): correlation between TABE Reading and GED Social Studies (r=.63), Science (r=.60), Reading (r=.64); TABE Mathematics and GED Mathematics (r=.64); TABE Language with GED Writing (r=.55); the total TABE battery with average GED (r=.70).

Jobs used for in the past: The TABE has been used as an overall Reading Grade level variable (but is not stored in a permanent database) (Grafton, personal communication). This variable is used by the education division to determine who needs additional educational training.

Validity Evidence: The only available validity justification is content validity. The test was developed based on curriculum guides, textbooks, and instructional programs.

14. Published: TABE Language Composite

Short Test Description:

The TABE Language Composite subsumes three tests:

Language Mechanics - This test contains 30 items that measure skills in the mechanics of capitalization and punctuation. Editing skills are measured in the context of passages presented in various formats.

Language Expression - This test contains 45 items that measure skills in language usage and sentence structure. The items measure skills in the use of various parts of speech, formation and organization of sentences and paragraphs, and writing for clarity. All items in the test are based on rules of written standard English.

Spelling - This test contain 30 items that measure applications of spelling rules for consonants, vowels, and various structural forms. Items are presented in the context of sentences with a missing word. The subject identifies the correct spelling of the word that would complete the sentence.

Psychometrics:

Correlations with other constructs: The TABE Language scale is correlated with GED Writing (r=.55).

Internal Consistency Reliability:

Reports of KR20 statistics (Technical Report) for TABE forms 5 and (6)

Subtest:	KR20
Mechanics	.77 (.76)
Expression	.85 (.85)
Spelling	.84 (.82)

Validity Evidence: The only available validity justification is content validity. The test was developed based on curriculum guides, textbooks, and instructional programs.

15. Published: TABE Reading Composite

Short Test Description:

The Table Reading Composite subsumes two subtests:

Reading Vocabulary - The test contains 30 items that measure same meaning words, opposite-meaning words, multi-meaning words, the meaning of affixes, and words in context.

Reading Comprehension - This test contains 40 items that measure comprehension of reading passages. Items test ability to extract details, analyze characters, identify main ideas, and interpret events described in passages. Items also test ability to differentiate various forms of writing and various writing techniques.

Psychometrics:

Correlations with other constructs: The TABE was correlated with GED subtest scores - the two tests were taken within 6 weeks of each other (N=678): correlation between TABE Reading and GED Social Studies (r=.63), GED Science (r=.60), and GED Reading (r=.64).

Internal Consistency Reliability:

Reports of KR20 statistics for TABE forms 5 and (6)

Subtest:	KR20
Vocabulary	.87 (.86)
Comprehension	.87 (.89)

Validity Evidence: The only available validity justification is content validity. The test was developed based on curriculum guides, textbooks, and instructional programs.

16. Published: TABE Mathematics Composite

Short Test Description:

The TABE Mathematics composite contains the following:

Mathematics Computation - This test contains 48 items that measure understanding of the operations of addition, subtraction, multiplication, and division. Depending on the level of the test, content includes whole numbers, decimals, fractions, algebraic expressions, percents, and exponents.

Mathematics Concepts & Applications - This test contains 40 items that measure understanding of mathematics concepts. Specific skills include numeration, number sentences, number theory, problem solving, measurement, and geometry.

Psychometrics:

Correlations with other constructs: The TABE was correlated with GED subtest scores - the two tests were taken within 6 weeks of each other (N=678): TABE Mathematics and GED Mathematics were correlated (r=.64).

Internal Consistency Reliability:

Reports of KR20 statistics (Technical Report) for TABE forms 5 and (6)

Subtest:	KR20
Math. Computation	.91 (.91)
Concepts & Appl.	.84 (.83)

Validity Evidence: The only available validity justification is content validity. The test was developed based on curriculum guides, textbooks, and instructional programs.

Published: Adult Basic Learning Exam (ABLE) Introduction

Construct Measured:

Educational achievement in reading, mathematics and language arts. Educational achievement of adults with about 12 years of school.

Short Description of Test:

The ABLE is a multiple choice format test, and is available in parallel forms. Test level 3 is appropriate for audiences who have had at least 8 years of school (ABLE, Norms Booklet, 1986).

Number of Items:		Time Limit:
Reading Comp.	48	35 min.
Vocabulary	32	20 min
Spelling	30	
Language	30	
Numerical Operations	40	
Problem Solving	40	(total test 175-215 min)

Apparatus: Paper and pencil

Psychometrics:

Scoring: The score is the total number of questions answered correctly.

Reliability: The Mental Measurements Yearbook (1992) reports that internal consistency estimates range between .8-.9.

Validity Evidence: The primary source of validity is content validity of the items and test as compared to stated objectives.
Intercorrelations with the Stanford Achievement Test series for Level III are all about .80. Specifically, Vocabulary correlates .80; Reading Comprehension .80; Spelling .80; and Total Mathematics .81 (Norms Booklet, 1986).

17. Published: ABLE Vocabulary Composite

Short Description of Test:

The ABLE Vocabulary Composite subsumes 1 subtest:

Vocabulary - This test is designed to tap into the individual's understanding and knowledge of words which are typically used by adults at work and in daily activities. The test is a multiple choice format with 32 items. The subject reads a sentence and is required to fill in the last word of the sentence, given 3 words to choose from to complete the sentence.

Psychometrics:

Scoring: The score is the total number of questions answered correctly.

Reliability: The Mental Measurements Yearbook (1992) reports that internal consistency estimates range between .8-.9. The ABLE, Norms Booklet, (1986) reports the following reliability estimates:

	KR21	
	Form E	Form F
Vocabulary	.82	.83

Validity Evidence: The primary source of validity is the content validity of the items and content of the test as compared to stated objectives.
Intercorrelations with the Stanford Achievement Test series for Level III are all about .80. Specifically, Vocabulary's correlation is .80 (Norms Booklet, 1986).

18. Published: ABLE Reading Comprehension Composite

Short Description of Test:

The ABLE Reading Comprehension Composite subsumes 1 subtest:

Reading Comprehension - This test is designed to measure the subject's ability to understand written information. The subject is presented with information (educational or functional - signs, advertisements in nature) to read. Then the subject is to answer questions about the information. Questions tap the individual's ability to understand the explicit message of the information, as well as to draw inferences and conclusions from the information.

Psychometrics:

Scoring: The score is the total number of questions answered correctly.

Reliability: The Mental Measurements Yearbook (1992) reports that internal consistency estimates range between .8-.9. The ABLE, Norms Booklet, (1986) reports the following reliability estimates:

	KR21	
	Form E	Form F
Reading Comp.	.90	.91

Validity Evidence: The primary source of validity is the content validity of the items and content of the test as compared to stated objectives.
Intercorrelations with the Stanford Achievement Test series for Level III are all about .80. Specifically, Reading Comprehension's correlation is .80 (Norms Booklet, 1986).

19. Published: ABLE Language Composite

Short Description of Test:

The ABLE Language Composite subsumes 2 subtest:

Spelling - This test is designed to measure the individual's level of written communication skills. The subject is presented with four words and must identify the word which is misspelled. There are 30 test items.

Language - This test has two parts 1. Capitalization and Punctuation; and 2. Applied Grammar. 1. Capitalization and punctuation taps the individual's use of capital letters, and punctuation such as commas, periods, colons. The subject reads a sentence that has words, or groups of words that are underlined. The subject must identify if there is a mistake in the use of capitals or punctuation. 2. Applied grammar taps usage of verbs, adjectives, pronouns, etc. The subject is required to read a sentence that has a blank in it, and choose from among four alternatives the correct word to fill the blank.

Psychometrics:

Scoring: The score is the total number of questions answered correctly.

Reliability: The Mental Measurements Yearbook (1992) reports that internal consistency estimates range between .8-.9. The ABLE, Norms Booklet, (1986) reports the following reliability estimates:

	KR21	
	Form E	Form F
Spelling	.89	.89
Language	.88	.88
Total Language	.94	.95

Validity Evidence: The primary source of validity is the content validity of the items and content of the test as compared to stated objectives.
Intercorrelations with the Stanford Achievement Test series for Level III are all about .80. Specifically, Spelling's correlation is .80 (Norms Booklet, 1986).

20. Published: ABLE Mathematics Composite

Short Description of Test:

The ABLE Mathematics Composite subsumes 2 subtest:

Number Operations - This test is designed to tap the individual's ability to read/write numbers; interpret fractions; operate on ratios, proportions and percentages; and to work with equations. The subject is required to calculate answers to number problems using mathematical operations. The test is comprised of 40 items and the subject must choose an anwser from 4 number/answer alternatives, or option 5 which is an answer "not given" option.

Problem Solving - The subject is required to solve 40 problems which are typical problems adults encounter. The test measures the individual's ability to develop an answer, to record and retrieve information, to measure, and to use geometric concepts. The test also includes items that tap the individual's ability to verify statistics and estimate outcomes.

Psychometrics:

Scoring: The score is the total number of questions answered correctly.

Reliability: The Mental Measurements Yearbook (1992) reports that internal consistency estimates range between .8-.9. The ABLE, Norms Booklet, (1986) reports the following reliability estimates:

	KR21	
	Form E	Form F
Numerical Operations	.90	.91
Problem Solving	.90	.89
Total Mathematics	.94	.94

Jobs used for in the past: Francis Grafton suggests using the ABLE for a basic Reading score. However, at this point this information is not collected.

Validity Evidence: The primary source of validity is the content validity of the items and content of the test as compared to stated objectives.
Intercorrelations with the Stanford Achievement Test series for Level III are all about .80. Specifically, the correlation of Total Mathematics is .81 (Norms Booklet, 1986).

21. Experimental: Project A Perceptual Speed and Accuracy (PSA)

Construct Measured:

The ability to perceive visual information quickly and accurately and to perform simple processing tasks.

Short Description of Test:

The respondent makes a rapid comparison of two visual stimuli presented simultaneously to determine if they are the same or different (e.g., //$*$ vs. //$/*). Stimuli presented include alpha, numeric, symbolic, and a mix of the previous three. The character length of stimuli is varied on three levels: 2, 5, and 9 characters. The Employee Aptitude Survey visual skills and abilities, and ASVAB coding speed were used as marker tests early in the development of PSA.

Number of Items: 36 Time Limit: about 6 minutes

Speededness: self-paced Apparatus: Computerized

Psychometrics:

Scoring: The test yields two scores: proportion correct and decision time. Decision time has better variability and is more reliable than proportion correct. Decision time is reflected such that higher scores are "better."

Correlations with other constructs: Correlates with Target Identification Test scores and tends to load with Target Identification in factor solutions that include a wide range of tests (i.e., ASVAB, MAP, Assembling Objects, Psychomotor).

Subgroup Differences: On average, males receive higher decision time scores than females (.09 to .19 of an sd), but females are more accurate than males (females score about 1/3 of an sd higher than males on proportion correct). Whites score slightly higher than blacks on both decision time and proportion correct with effect size of .02 to .04 on decision time and .12 to .24 for accuracy.

Reliability: In three samples, each having more than N=6,000, decision time split half reliability estimates ranged from .94 to .96. Proportion correct has less variability and is less reliable (split half estimates range from .61 to .65). Test retest estimates (N=473) were .63 for decision time and .51 for proportion correct.

Practice and Coaching Effects: Gains due to practice are small to moderate ranging from .08 SD gain (1 month interval between testing) .35 (with 2wk break) for Decision time and .05 (2 wk. break) to .11 (1 month break) for proportion correct.

Validity Evidence: In Project A, PSA was combined with Target Identification to form a composite and the validity of a set of computer test composites was compared with the validity of the ASVAB. While the computer composites yielded high validities in predicting technical proficiency, those validities were typically no higher than the level of validity achieved by the ASVAB alone.

22. Experimental: Project A Target Identification Test (TID)

Construct Measured:

Perceptual speed involving matching stimuli rapidly.

Short Description of Test:

A target (e.g., helicopter, tank) is presented near the top of the screen, and 3 stimuli appear in a row near the bottom. The respondent must identify which of the 3 stimuli represents the same object as the target as quickly as possible. The target may need to be rotated relative to its current position to better match the stimulus object in terms of position. The target objects are military vehicles and aircraft used by various nations. The position, orientation, angle and size of the object is manipulated.

Number of Items: 36 Time Limit: about 4 minutes
Speededness: self-paced Apparatus: Computerized

Psychometrics:

Scoring: The test yields two scores: proportion correct and decision time. Decision time has better variability and is more reliable than proportion correct. Decision time is reflected such that higher scores are "better."

Correlations with other constructs: Correlates with Perceptual Speed and Accuracy scores and tends to load with Perceptual Speed and Accuracy in factor solutions that include a wide range of tests (i.e., ASVAB, MAP, Assembling Objects, Psychomotor).

Subgroup Differences: On average, males receive higher decision time scores than females (about 1/2 of an sd), but females are more accurate than males (females score about 1/10 of an sd higher than males on proportion correct. Whites slightly score higher than blacks on both decision time and proportion correct with effect size of .65 to .71 on decision time and .13 to .23 for accuracy.

Reliability: In three samples, each having more than N=6,000, decision time split half reliability was .97. Proportion correct has less variability and is less reliable (split half estimates range from .62 to .69). Test retest estimates (N=473) were .78 for decision time and .40 for proportion correct.

Practice and Coaching Effects: Gains due to practice are moderate to large ranging from .32 SD gain (1 month interval between testing) to .47 SD (2 wk interval between testing) Decision time.

Validity Evidence: In Project A, TID was combined with Perceptual Speed and Accuracy to form a composite, and the validity of a set of computer test composites was compared with the validity of the ASVAB. While the computer composites yielded high validities in predicting technical proficiency, those validities were typically no higher than the level of validity achieved by the ASVAB alone.

23. Experimental: Project A Number Memory (NM)

Construct Measured:

Basic math ability and working memory capacity.

Short Description of Test:

A number is presented on the computer screen. After studying the number, the subject pushes a button to receive the next part of the problem. At the press of the button, the first part disappears. Another number and an operation term (+, -, *, or /) appear. After completing the operation (e.g., 39 + 18), the subject pushes the button again to receive the third part of the problem, another number along with an operation term. This proceeds until a solution to the problem is presented. Then the subject must indicate whether the solution is right or wrong. The number of operations to be performed varies from 4, 6, or 8; interstimulus delay time also measures short term memory. This test is similar to a working memory test developed by the Air Force.

Number of Items: 28 items Time Limit: about 10 minutes
Speededness: self-paced Apparatus: Computerized

Psychometrics:

Scoring: The test yields two scores: proportion correct and decision time. They are combined to form a composite.

Correlations with other constructs: Number Memory loads with ASVAB AR and MK in factor solutions including a wide range of cognitive test scores. It also yields moderate correlations with ASVAB NO and to a lesser extent ASVAB CS (speeded tests).

Subgroup Differences: Males score higher than females, but the differences are relatively small, .13 to .18 SD. Whites score higher than blacks by about one-half of an SD.

Reliability: In three samples, each having more than N=6,000, decision time split half reliability ranged from .93 to .95. Proportion correct has less variability and is less reliable (split half estimates range from .53 to .59). Test retest estimates (N=473) were .73 for decision time and .53 for proportion correct. The internal consistency of the composite score (decision time and proportion correct) was .83.

Validity Evidence: The validity of a set of computer test composites was compared with the validity of the ASVAB. While the computer composites yielded high validities in predicting technical proficiency, those validities were typically no higher than the level of validity achieved by the ASVAB alone.

24. Experimental: Project A Short Term Memory (STM)

Construct Measured:

Ability to store and recall information in short term memory.

Short Description of Test:

A box appears containing 1, 3, or 5 objects. After a delay period of .5 to 1.0 seconds the box disappears. After another delay a probe item appears. The subject must decide if the probe item was included in the original stimulus set and press a white key if it was or blue key if it was not.

Number of Items: 36 items Time Limit: about 7 minutes
Speededness: self-paced Apparatus: Computerized

Psychometrics:

Scoring: The test yields two scores: proportion correct and decision time. They are combined to form a composite.

Correlations with other constructs: Short Term Memory forms a factor of its own in factor analyses including ASVAB, spatial, psychomotor, and computer tests. The STM composite score is not correlated very highly (i.e., always less than .32) with other variables.

Subgroup Differences: Females tend to perform better than males on this task with effect sizes ranging from -.05 and -.11. Whites perform better than blacks, but the effect sizes are relatively small .19 and .21.

Reliability: In three samples, each having more than N=6,000, decision time split half reliability ranged from .96 to .97. Proportion correct has less variability and is less reliable (split half estimates range from .48 to .60). Test retest estimates (N=473) were .66 for decision time and .41 for proportion correct. The internal consistency of the composite score (decision time and proportion correct) was .80.

Practice and Coaching Effects: There is an increase in performance with practice up to .15 SD over a 1 month interval (N=473).

Validity Evidence: The validity of a set of computer test composites was compared with the validity of the ASVAB. While the computer composites yielded high validities in predicting technical proficiency, those validities were typically no higher than the level of validity achieved by the ASVAB alone.

25. Operational: Defense Language Aptitude Battery (DLAB)

Construct Measured:

Aptitude for Language Acquisition

Short Description of Test:

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.

Number of Items: 119 Time Limit: 90 minutes
Apparatus: paper and pencil test, audio equipment

Psychometrics:

Scoring: All items are pooled to form one composite but there is some factor analytic research supporting three factors.

Correlations with other constructs: 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. Silva et al. (1991) computed 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.

Reliability: Peterson and Al-Haik (1976) report KR-21 reliabilities for the 3 factors or subtests of the DLAB. The KR-21 estimates ranged from .78 to .82; with an estimate of .89 for the total test.

Validity Evidence: 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 and White 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.

Experimental: Cognitive and Meta-Cognitive Predictors of Leadership Potential Introduction

These measures tap the skills crucial to leader performance, basic cognitive capacities and social skills. These skills facilitate development of the knowledge structures and problem solving skills that leaders need to apply in ill-defined problem solving situations.

Short Description of Test:

These are a series of computer administered tests in development under an Army contract. They tap various cognitive capacities. Descriptions of individual tests developed and tested on a college sample are defined. Further description is provided of a series of computerized tests that tap actual problem solving behaviors. These tests are oriented to an Army audience.

There are 11 tests in the battery. Five of these tests have been tested on 161 undergraduates and the results are summarized on the following pages:

- Problem Construction (PC)
- Information Encoding (IE)
- Category Search and Specification (CS)
- Category Combination (CC)
- Wisdom. (W)

Correlations among four of the measures based on 161 students' scores were:

	CS	PC	IE	CC
Category Search (CS)	1.00			
Problem Construction (PC)	.10	1.00		
Information Encoding (IE)	.12	.27	1.00	
Category Combination (CC)	.21	.23	.20	1.00
Total SAT Score	.13	.20	.17	.24

A written format of the next six tests was administered to a sample of Army officers. Due to the labor requirements for coding the results a computerized format is in the development phase with final tests ready in January, 1995:

- Problem Solving Skills,
- Solution Characteristics,
- Problem Evaluation,
- Planning and Implementation,
- Leadership Knowledge, and
- an alternative Wisdom measure.

26. Experimental: Problem Construction

Construct Measured:

Problem construction or problem finding skills. Individuals must identify and structure the problem to be solved rather than working with "givens."

Short Description of Test:

The respondent is presented with 4 problem scenarios developed by Baer (1988). They are a series of complex, ill-defined situations which can be defined a number of ways. The respondent must restate the problem scenario and is scored on the number of times they choose to structure the problem in terms of goals, procedures, key information or restrictions. They are also judged on the quality and originality of these restatements. 16 response alternatives are presented covering the range of content preferences and are rated for quality and originality.

Example: "You are selected to represent your country in the Olympic track and field. You are one of the top "hopefuls," but your doctor has advised you to have surgery immediately or risk a debilitating injury. However, to have the surgery would mean missing the games."

Potential Responses:

- How can I use my fame so as to help others avoid this condition? (Goal information)
- How can I get a bionic replacement part so I can participate? (Procedures information)
- How can I find out if other athletes dealt with this same condition successfully? (Key information)
- How can I make this decision on the basis of what is best for the team? (Restrictions information)"

Psychometrics:

Correlations with Other Measures: Was correlated .20 with Total SAT and .00 with GPA in a college student sample (N=161).

Scoring: The score is the total number of times that respondents chose high originality and high quality responses for each information content type.

Validity Evidence: Problem construction had a mean validity of r=.28 (N=161) with performance on ill-defined complex problems.

27. Experimental: Information Encoding

Construct Measured:

Information encoding skills

Short Description of Test:

The respondent is presented with 4 problems (2 business, and 2 political). Each problem is presented on 6 "index cards" displayed on a computer screen. Respondents may only view one card at a time, and may "page back" to any card after initially viewing all six. Respondents are asked to type a one paragraph solution to the problem.

Psychometrics:

Scoring: A score is obtained for the relative time spent viewing different kinds of information, i.e., discrepant facts, key diagnostic facts, abstract principles, etc.
Quality and originality of the solution is also rated by 4 judges.

Correlations with Other Measures: Was correlated .17 with Total SAT, .15 with GPA, and .20 with verbal reasoning ability in a college student sample (N=161).

Validity Evidence: Information Encoding had a mean validity of r=.36 (N=161) with performance on ill-defined complex problems.

28. Experimental: Category Search and Specification

Construct Measured:

Ability to link information to existing concepts or schema.

Short Description of Test:

Respondents are presented 4 2-3 paragraph complex, ill-defined organizational scenarios from Shorris (1981). Next, they are asked to review eight concepts that might be useful in generating a solution to the problem. These concept statements reflect four dimensions: general principles, long-term goals, evaluation of others, and discrete action plans. The subjects must answer a series of questions regarding the scenario.

Why did the situation occur?
What were the major mistakes in handling the situation?
What would you do in this situation?

Example: "The amounts charged to the expense account were exorbitant. This was not the occasional three martini lunch or theater tickets--the sales rep was spending over one hundred and fifty thousand dollars a year--which was more than the whole regional office travel and entertainment budget. The receipts were all there, but the legitimacy of the expenses was questionable. However, the sales rep had been an assistant to the Undersecretary of the Navy during a previous administration and he really knew his way around Washington....."

What would you do in this situation?

- The regional manager has to decide whether to take the fall or expose the situation (Relatedness)
- Fiscal irresponsibility can set a bad precedent for other reps. (Long-Term Goals)
- The regional manager should take into account ethics, customary Washington lobbying practices, and personal and career considerations in deciding what to do. (Integration)

Psychometrics:

Scoring: Each of the four content dimensions were scored by summing the total number of statements selected for each dimension across all of the problems.

Correlations with Other Measures: Was correlated .15 with Total SAT, -.05 with GPA, and .22 with verbal reasoning ability in a college student sample (N=161).

Validity Evidence: Category Search had a mean validity of r=.25 (N=161) with performance on ill-defined complex problems.

29. Experimental: Category Combination

Construct Measured:

The ability to synthesize various concepts and to construct a coherent model of the phenomenon.

This measure taps combination - reorganization skills, resulting in synthesis of ideas. It requires respondents to a) search for key features of each category; b) identify the shared/nonshared features to be used in linking concepts; c) provide elaboration of implications of new concept through identification of new features. Category combination is the basis for synthesis and generation of new models for understanding a problem situation.

Short Description of Test:

The respondent is presented with 4 problems. For each problem, subjects are presented 3 lists of four words each. Each word list contains the names of related items that comprise a specific concept category, e.g., "birds," etc. The lists, however, are relatively unrelated.

The respondent is instructed to consider the three short word lists and think of how they might be combined to obtain a list of related items. Next, respondents are asked to generate:

1) a descriptive label the category
2) additional features or attributes of the category
3) additional exemplars or members of the new category

Example: "Your task is to look at these three categories and combine them into one category. Approach this task as though the 12 words are a single list of words, and you had to invent a name for the list.

seat	glove	bicycling
tire	baseball	running
brakes	baseball bat	swimming
wheel	football	lifting weights"

In a new version of this test, data from several hundred people who have completed this task will be used to generate labels, attributes, and exemplars that will be presented on the computer screen. Respondents will then choose from among a number of these alternatives. Scores will be computed for the quality and originality of their selections based on normative data.

Psychometrics:

Scoring: Scoring is by an expert scoring system. The categories are rated for quality and originality. They are contrasted to an existing pool of 1) labels, 2) features and 3) exemplars (which were previously rated). An average of all the scores constitutes the principle score.

Correlations with Other Measures: Was correlated .23 with Total SAT, .03 with GPA, and .21 with verbal reasoning ability in a college student sample (N=161).

Validity Evidence: Category Combination had a mean validity of r=.28 (N=161) with performance on ill-defined complex problems.

30. Experimental: Wisdom I

Construct Measured:

Measures several wisdom dimensions:

- self-objectivity - awareness of personal strengths and weaknesses
- self-reflection - willingness to learn from mistakes
- judgement under uncertainty - capacity to work with conflicting demands
- system perception - awareness of others' needs and concerns
- sensitivity to fit - awareness whether solution is consistent with ongoing patterns of social interaction
- social commitment - willingness to resolve conflict for betterment of others

Short Description of Test:

The respondent is presented with 10 examples of an Aesop's fables. They are asked to read the fable and identify the moral of the story. The fable asks the individual to resolve a complex social conflict. In this case students were offered 5 alternatives and picked the best response. The alternatives are rated with regard to approximation of the actual moral as well as the other wisdom dimensions (mentioned above).

Example: "A Fox had by some means got into the store-room of a theater. Suddenly he observed a face glaring down on him and began to be very frightened; but looking more closely he found it was only a Mask such as actors use to put over their face. "Ah," said the Fox, "you look very fine; it is a pity you have not got any brains."

What is the moral of the story?

- Only unintelligent people hide behind masks.
- Outside show is a poor substitute for inner worth.
- What is inside matters more than what is outside.
- Confronting your fears may show you there is not substance behind them.
- People use their appearances to deceive others."

Psychometrics:

Scoring: Score is the sum of the scores across the 10 fables. Alternatives are previously rated according to approximation of fable moral and other wisdom dimensions.

Correlations with Other Measures: Was correlated .06 with Total SAT, .03 with GPA, and .14 with verbal reasoning ability in a college student sample (N=161).

Validity Evidence: Wisdom had a mean validity of r=.15 (N=161) with performance on ill-defined complex problems.

31. Experimental: Problem Solving Skills (in development)

Construct Measured:

The ability to discern the usefulness and applicability of existing knowledge to a given problem or situation.

Short Description of Test:

Subjects are told that they will have to solve a military problem scenario. Their first task is to pick a group of 6 individuals to act as their staff to provide recommendations of solutions. The must read through resume information and make choices from 18 candidates. Then they are presented with the problem scenario, a series of 12 recommendations, 2 from each staff member, are presented on the screen (these recommendations will be developed from a pool of open-ended responses gathered from an initial Army sample). Subjects are required to select the 4 recommendations they believe are the most viable solutions for the given problem.

Psychometrics:

Scoring: The recommendations provided will be rated on quality, final score will be a summed score regarding the quality of the alternative chosen.

32. Experimental: Solution Characteristics (in development)

Construct Measured:

Problem Solving: This test is expected to measure the way that leaders/officers approach and structure complex ill-defined problems.

It will be a measure of the characteristics such as time frame, attention to restrictions, and goal preferences (and other characteristics important in decision making as identified in the literature) that leaders attend to in defining problems.

Short Description of Test:

This measure taps how officers structure and approach leadership problems. Given two problem scenarios (one military and one organizational) respondents are required to answer three questions:

(1) If you were in this situation what would be the one most important problem for you to address?
(2) What 6 key pieces of information would you need to solve the problem?
(3) What 6 other problems do you have to consider?

Respondents will be given 18 information statements (drawn from an earlier Army study's open-ended format answers) to choose from in answering those questions.

Psychometrics:

Scoring: Scores will be calculated for each dimension identified as important in characterizing a problem. The score will be a sum of the total number of statements chosen from each dimension across the three questions and two problems. A score will be calculated based on the rating of the statement as representing high, low or neither high nor low examples of behavior.

33. Experimental: Problem Evaluation (in development)

Construct Measured: Problem Evaluation

Short Description of Test:

In this computer task subjects are presented problem scenarios and they are to select from a series of questions, those that best help to thoroughly evaluate the stated problem. The question alternatives (drawn from responses in an Army sample) can be categorized as searching for objective evaluation criteria or social evaluation questions.

Psychometrics:

Scoring: A score can be calculated, assessing the comprehensiveness of the questions selected in terms of adequately addressing each both objective and social components. The quality of the questions chosen will also be evaluated.

34. Experimental: Planning and Implementation (in development)

Construct Measured:

Planning and Implementation

Short Description of Test:

Subjects are asked to develop a viable solution to a problem. A problem is presented, and subjects are given a number of tasks to choose from in creating their solution. They are instructed to create a good solution to the problem using the fewest number of tasks possible.

Responses collected from an Army sample will be categorized representing high and low level performance from high, mid-, and low level leaders. These will be used to develop prototypical response patterns for making ratings. These response patterns will include content and structure information.

Psychometrics:

Scoring: Subjects' responses will be scored by completing a profile analysis against the developed response categories.

35. Experimental: Leadership Knowledge (in development)

Construct Measured:

Leader expertise in organizing problems and plans.

Short Description of Test:

Subjects read through a list of tasks, and then groups together job tasks that are similar. Responses should reflect grouping-by-principle and similarity-to-taxonomy.

Psychometrics:

Scoring: Three scores will be obtained: (1) the first score is an objective index of integration - or a count of the tasks assigned to any groups, (2) similarity of generated categories to those as proposed by Fleishman et al. (1991), and (3) extent the grouping tends to reflect superficial categories.

36. Experimental: Wisdom II (in development)

Construct Measured:

Measures several wisdom dimensions:

- self-objectivity - awareness of personal strengths and weaknesses
- self-reflection - willingness to learn from mistakes
- judgement under uncertainty - capacity to work with conflicting demands
- system perception - awareness of others' needs and concerns
- sensitivity to fit - awareness whether solution is consistent with ongoing patterns of social interaction
- social commitment - willingness to resolve conflict for betterment of others

Short Description of Test:

In this test, respondents will read a problem regarding negative organizational outcomes, caused by a failure on the part of leadership to attend to complex social cues. Subjects will be asked to select responses to answer the following questions:

(1) Why did this situation occur?
(2) What was the central mistake made by the manager?
(3) What would you do if you were the manager?

The response alternatives provided will vary in terms of level of wisdom in the following dimensions (listed above) and will be developed from the responses provided from a sample of Army officers.

Psychometrics:

Scoring: Responses will be scored on wisdom dimensions (above), and a total wisdom score will be the sum of all high wisdom statements chosen across dimensions.

37. Experimental: Leadership Problems Inventory (in development in ECQUIP)

Construct Measured:

Ability to set priorities for problems encountered by supervisors.

Short Description of Test:

This test is akin to a paper-and-pencil version of an in-basket. Each item presents five problem scenarios. Subjects indicate which problem scenario they would attend to first, second, ... to last. In this way they indicate which problems are most critical and must be given priority over the rest. The problem scenarios were developed from critical incidents collected from NCOs. There are two parallel forms of the test, each containing 24 items.

Number of Items: 24

Apparatus: Paper and pencil

Psychometrics:

Scoring: The key will be based on the responses of high ranking NCOs.

38. Published: Army Radio Code Test

Construct Measured:

This test measures the speed that an individual can learn Morse code characters.

Short Description of Test:

The individual practices with the Morse code characters for 25 minutes using a tape recording device. The test consists of 150 items, the individual is tested on ability to learn the Morse characters for letters "I," "N," and "T." The test takes about 30 minutes to complete, and consists of two trials of 75 items. During the first trial 11 words are presented per minute, and 15 words per minute in the second trial. During the trial the individual marks on the score sheet under the "I", "N", or "T" as the stimuli signals are presented.

Psychometrics:

Scoring: The number of items that are identified correctly.

Correlations with other constructs: Fleishman (1955) found correlations between the ARC and the Signal Corps Code Aptitude Test (SCCAT) and the Radio Operator Aptitude Index (ROAI) of .45 and .50 respectively, with a sample size of 400. The latter tests had been previously used as a first screen for selection into Radio Operator Training. Due to this range restriction correlations are corrected for double restriction.

Reliability: Fleishman (1955) reported a split-half reliability (N=400), corrected to the full length of the test of .98.

Validity Evidence: Original work by Fleishman (1955) found the ARC had a corrected validity coefficient of .44 with the criterion, success in the Radio Operator Course (N=400). Fleishman et. al., (1958) found that the ARC had a corrected correlation (N=310) of $\underline{r}$=.27 with the training criterion, time to proficiency for receiving code. However, more recent studies suggest that the ASVAB composites are better predictors of Morse training performance and attrition than the ARC (Russell, Reynolds, & Campbell, 1994).

Prediction of attrition from Morse training was calculated at just .08 with the ARC (which is also called the Auditory Perception Test - APT) (Silva, personal communication, 1994). This unexpectedly low relationship may be due to testing conditions which possibly cause ceiling effects on the test scores, decreasing test variance, and due to database problems with these variables.

39. Experimental: Superdit - Sound Memory

Construct Measured:

This test measures memory of auditory stimuli.

Short Description of Test:

The subject is presented with a stimulus sound of Morse "dots and dashes." After a short delay, of 1, 2, or 3 seconds, the subject is presented another Morse signal and must indicate whether the two sounds were the same or different. The length of the stimulus Morse sounds is varied from two to four elements. The test is comprised of 10 practice trials and 24 test trials. This is a useful test for those tasks that have high information processing demands, but will be less useful for prediction of general language skills (Silva, personal communication, 1994).

Psychometrics:

Scoring: Scored simply as the number of sounds identified correctly (accuracy) and a measure of how long it takes the subject to respond (reaction time).

Validity Evidence: Initial data analyses indicate that with a sample size of 93, prediction of attrition from Morse training was $r=.29$ ($p<.05$) (Silva, personal communication, 1994).

40. Experimental: Superdit - Sound memory with interference

Construct Measured:

Measures reaction time to auditory and visual stimuli.

Short Description of Test:

This test is similar to the Sound Memory Test. The subject is presented with a stimulus sound of Morse "dots and dashes." After a short delay, of 1, 2, or 3 seconds, the subject is presented another Morse signal and must indicate whether the two sounds were the same or different. During the delay, a monotone sound (white noise) is presented increasing the difficulty of recalling the first stimulus sound. The length of the stimulus Morse sounds is varied from two to four elements. The test is comprised of 10 practice trials and 24 test trials. This is a useful test for those tasks that have high information processing demands, but will be less useful for prediction of general language skills (Silva, personal communication, 1994).

Psychometrics:

Scoring: Scored simply as the number of sounds identified correctly (accuracy) and a measure of how long it takes the subject to respond (reaction time).

Validity Evidence: Initial data analyses indicate that with a sample size of 93, prediction of attrition from Morse training was $r=.27$ ($p<.05$) (Silva, personal communication, 1994).

41. Experimental: Superdit - Motor Programming Test

Construct Measured:

Measure of the time required for an individual to translate a Morse character (e.g., "e") into a physical/motor Morse Code response.

Short Description of Test:

This test presents the subject with stimulus Morse characters of "dots and dashes." The subject is given time to organize their response, then they are required to replicate the stimulus Morse characters at a specific identified time. The subject replicates the stimulus sound by pressing two keys for the "dots" and "dashes." The stimulus sound is varied from two to four elements, across the 10 practice trials and 24 test trials.

Psychometrics:

Scoring: Score is based on the accuracy of the response, and a measure of how long it takes the subject to respond (reaction time).

Validity Evidence: Initial data analyses indicate that with a sample size of 93, prediction of attrition from Morse training was $r=.27$ ($p<.05$) (Silva, personal communication, 1994).

Appendix C
Biographical, Interest, and Temperament Measures

Experimental: Army Biodata Inventory--Introduction

Biodata inventories work on the axiom that prior events predict future events. There are links between prior situations and behaviors, and current capabilities. Various choices and adaptive processes allow an individual to build a repertoire of skills, abilities and knowledge that can be applied to various contexts.

General Biodata examples:

What is your height?
What is your birth-order in your family?
The place in which you spent most of your high school years was a...
How old were you when you had your first steady paid job outside you home?
There are generally 4 or 5 alternatives to choose an answer from that best applies to the individual.

Short Description of Test:

The Army Biodata Inventory (ABI) was designed for standardized use with a wide array of populations (Grey & Mael, in preparation). Based on a large body of biodata literature, the authors drew biodata items that appeared to be stable across multiple samples. They administered the ABI to Captains and Majors (N=600), empirically scored it against a leadership criterion measure, and factor analyzed biodata items. The intent was to identify stable factors such that ABI users would be able to select appropriate biodata factors for the population of interest. The factors were revalidated against data gathered from the West Point study with 2500 cadets (Mael and Hirsch, 1993) and found to be predictive consistent with a priori hypotheses. Six biodata factors based on West Point results are suggested for inclusion in the SF ABI:

- Academic Performance
- Formal Leadership
- Ruggedness
- Mechanical Activities
- Work Experience
- Home Economics

Two additional biodata factors from an attrition study (Mael & Ashforth, in preparation) are suggested for inclusion:

- Nondelinquency
- Team Sports/Group Orientation

Three additional biodata scales are proposed for development for SF based on SF job analysis results (Russell, Crafts, Tagliareni, McCloy, & Barkley, 1994):

- Work Skills
- Family/Community
- Cross-Cultural Sensitivity.

Examples of biodata research are provided in two studies described on the following pages.

Mael & Ashforth

A biodata form was developed to predict attrition of a total of 2500 U.S. Army recruits, based on their level of organizational identification with the Army. The measure had 115 items.

Scoring: Items were keyed to an Organizational Identification measure.
Four factors emerged from the data set:

1. Rugged/outdoors (11 items): enjoys outdoor activities and hands-on work
2. Solid citizen (10 items): nondelinquent, dependable pattern of work
3. Team Sports/Group Orientation (6 items): interest and involvement in team-oriented sports; preference for working in a group
4. Intellectual/Achievement Oriented (7 items): diligent involvement in intellectual pastimes.

Correlations with other constructs: Of the four biodata factors (based on N=1021) 'solid citizen' correlated at the $p<.01$ level with educational level ($r=.15$) and AFQT score ($r=.10$). The 'team sports/group orientation' factor and the 'intellectual/achievement orientation' also correlated with the AFQT score ($r=-.07$, $p<.05$) and ($r=.14$, $p<.01$) respectively.

Reliability: The internal consistency (alpha) ranged from (based on N=2535) .37 for Intell/Achievement Oriented, to .85 for Rugged/outdoors. It should be noted that high internal consistency is neither expected nor particularly desired with this type of measure.

Validity Evidence: All the biodata factors significantly predicted attrition (N=1021) at 6 months (validities ranged from .07 to .30). Solid citizen; Team Sport/Group Orientation and Intellectual/Achievement Orientation predicted attrition across 24 months, with prediction strongest at the earlier time periods.

Mael and Hirsch (1993)

A biodata form was developed to tap relevant temperament constructs and to minimize socially desirable responding that is found on the ABLE. It was administered to U.S. Military Academy cadets at West Point to predict leadership ratings. The measure had 73 items.

Scoring: These biodata items were empirically keyed to both sets of criteria. There is the basic training and field training leadership scores, along with fall and spring semester leadership ratings during the academic year.

Correlations with other constructs: As expected from the keying, the biodata scales were strongly related to the ABLE scales. Correlations ranged from .37 to .53 for the Work Orientation scale. The biodata keyed to the criterion formed two separate dimensions. The fall and spring ratings of leadership where highly related, and the basic and field ratings of leadership were also related, indicating separate dimensions of leadership that are fairly different from one another.

Gender/Race Differences In this data, two of the criterion measures, the Fall & Spring leadership ratings, showed race differences of about 1 SD (Mael, personal communication, 1994). These criterion measures were collected during the academic year and are significantly related to high school rank. Biodata keyed to these measures also show similar race differences. Biodata keyed to the Basic and Field leadership ratings do not show these race differences (blacks tend to score higher), these measures are more related to physical fitness attributes than to academic achievements.

Fakability:
The biodata items that were keyed based to the ABLE constructs showed significantly lower correlations with social desirability than did the ABLE scales. But the biodata items empirically keyed to the criteria, showed even lower levels of socially desirable responding.

Validity Evidence: Biodata items empirically keyed to the criterion measures predicted the leadership ratings (Basic=.30; Fall=.39; Spring=.40; Field=.34; N=1325). 1994 keys were cross-validated on the 1995 class sample, which did not show excessive shrinkage. The biodata showed incremental validity to the Whole Candidate Score (primary measure used for selection into West Point) in predicting leadership ratings R^2=.02 to .05 in 1994 (N=1325) and in 1995 (N=1240).

1. Experimental: Army Biodata Inventory - Academic Performance

Construct Measured:

Academic Performance

Short Description of Test:

The 5 items which tap this factor have to do with High School grade, and class rank.

Psychometrics:

Reliability: Alpha obtained from the Mael and Hirsch (1993) data set was .84.

Validity Evidence: This factor was found to be most predictive of the fall and spring leadership grade (from Mael & Hirsch data).

2. Experimental: Army Biodata Inventory - Formal Leadership

Construct Measured:

Formal Leadership

Short Description of Test:

The 4 items which tap this factor have to do with having leadership positions in the student body during high school, being a class officer.

Psychometrics:

Reliability: Alpha obtained from the Mael and Hirsch (1993) data set was .82.

Validity Evidence: This factor was found to be most predictive of the fall and spring leadership grade (from Mael & Hirsch data).

3. Experimental: Army Biodata Inventory - Ruggedness

Construct Measured:

Ruggedness

Short Description of Test:

The 4 items which tap this factor have to do with interests in mountain climbing, and camping.

Psychometrics:

Reliability: Alpha obtained from the Mael and Hirsch (1993) data set was .73.

Validity Evidence: This factor was found to be most predictive of the basic field training scores (from Mael & Hirsch data).

4. Experimental: Army Biodata Inventory - Mechanical Activities

Construct Measured:

Mechanical Activities

Short Description of Test:

The 3 items which tap this factor have to do with interests and experience with car repairs and operating machinery.

Psychometrics:

Reliability: Alpha obtained from the Mael and Hirsch (1993) data set was .71.

Validity Evidence: This factor was found to be most predictive of the basic field training scores (from Mael & Hirsch data).

5. Experimental: Army Biodata Inventory - Work Experience

Construct Measured:

Work Experience

Short Description of Test:

The 3 items which tap this factor have to do with how much work experience individuals had during High School and asks the age at which individuals started working, and how much they worked on average during high school.

Psychometrics:

Reliability: Alpha obtained from the Mael and Hirsch (1993) data set was .63.

Validity Evidence: This factor was found to be most predictive of the basic field training scores (from Mael & Hirsch data).

6. Experimental: Army Biodata Inventory - Home Economics

Construct Measured:

Home Economics

Short Description of Test:

The 3 items which tap this factor have to do with experiences with cooking, sewing and babysitting. This factor taps into a level of self-sufficiency.

Psychometrics:

Reliability: Alpha obtained from the Mael and Hirsch (1993) data set was .69.

Validity Evidence: This factor was found to be most predictive of the basic field training scores (from Mael & Hirsch data).

7. Experimental: Army Biodata Inventory - Nondelinquency
Construct Measured: Nondelinquency Short Description of Test: The 10 items making up the nondelinquency or Solid citizen factor tap into nondelinquent, dependable patterns of work behavior and work. Psychometrics: Correlations with other constructs: This nondelinquency factor correlated with educational level and with AFQT score (r=.15 and r=.10; p<.01; respectively). Reliability: Alpha for the entire sample (n=2500) from the Mael and Ashforth study was .60. Validity Evidence: Nondelinquency significantly predicted attrition from the Army across 24 months (N=1021). Prediction at 6 months was .17 (p<.01) and .10 (p<.01) at 24 months.

8. Experimental: Army Biodata Inventory - Team Sports/Group Orientation
Construct Measured: Team Sport/Group Orientation Short Description of Test: The 6 items comprising the Team Sports/Group Orientation factor tap into an interest and involvement in team-oriented sports and a preference for working in a group. Psychometrics: Reliability: The coefficient alpha for the entire sample (N=2500) in the Mael and Ashforth study was .45, however with a biodata instrument high internal consistency is not a goal, and is not expected. Validity Evidence: The team sports/group orientation factor predicted attrition from the Army for 24 months (N=1021). Prediction at 6 months was .30 (p<.01).

9. Proposed Experimental: Army Biodata Inventory - Work Skills
Construct Measured: Work experiences defined by workshop participants in the SF job analysis. Short Description of Test: Items which will tap this factor have to do with what type or the content of previous work experience. This includes work experience individuals had in the conventional Army, i.e., specific MOS. In addition, experiences work experiences outside the Army, what types of work was actually performed, skilled trades or farming jobs.

10. Proposed Experimental: Army Biodata Inventory - Family/Community
Construct Measured: Family/Community biographical items suggested by SF job analysis workshop participants. Short Description of Test: These items will tap the individual's early family experiences in terms of having to move frequently, being brought up in a military family, being exposed to hardship as a child and the strength of family ties.

11. Proposed Experimental: Army Biodata Inventory - Cross-Cultural Sensitivity

Construct Measured:

Awareness and sensitivity to cross-cultural differences and similarities.

Short Description of Test:

This factor will tap an individual's awareness of cross-cultural differences and sensitivity in dealing with indigenous populations. Items will tap past experiences, and curiosity about other cultures and people. Potential items currently exist that have been tested on peacekeeping troops, and soldiers who deal with indigenous peoples.

Experimental: Ranger Biodata Inventory--Introduction

Biodata inventories work on the axiom that prior events predict future events. There are links between prior situations and behaviors, and current capabilities. Various choices and adaptive processes allow an individual to build a repertoire of skills, abilities and knowledge that can be applied to various contexts.

Short Description of Test:

The Ranger Biodata Inventory (RBI) was designed to predict advancement and performance in Ranger Battalions. It contains 138 items that query respondents about their past behavior and reactions to specific life events. Individual items consist of multiple choice questions with five response options. Administration time is 30-40 minutes.

Items are scored on nine scales:

- Cognition Under Stress
- Mature Team Commitment
- Self-Esteem
- Combat Motivation
- Need for Achievement
- Outdoor Orientation
- Physical Endurance
- Physical Strength
- Object Belief

Psychometrics:

Scoring: The RBI is rationally-keyed. Items are scored a priori to reflect the degree to which the response measures the intended construct. Item scores are then summed to obtain construct scale scores.

Correlations among constructs: RBI scales correlate -.22 to .61 with each other.

Correlations with Other Measures: Correlations with ABLE scales ranged from .00 to .53. The highest correlations were obtained between biodata and ABLE scales measuring similar constructs (e.g., Work Orientation and Need for Achievement, r=.54).

Reliability: The internal consistency estimates range from .55 to .81 for the scales.

Fakability: Correlations of the biodata scales with a validity scale to measure deliberate faking ranged from .01 to .25. The magnitude of the correlations is lower than that obtained with previously developed temperament scales. Preliminary analyses show little faking in operational use where the temptation to fake may be high.

Validity Evidence: Analysis of the data is ongoing, and some scales may be substantially revised. However, preliminary results suggest that the biodata scales are strong predictors of advancement and various administrative performance criteria among Rangers. Shrunken multiple-Rs range from .25 to .50 (N=300).

The RBI will be administered experimentally in SFAS in the fall of 1994.

12. Experimental: Ranger Biodata Inventory - Cognition Under Stress
Cognition under stress contains items measuring the ability to think rationally under pressure.

13. Experimental: Ranger Biodata Inventory - Mature Team Commitment
Mature team commitment contains items measuring the willingness to make sacrifices and to assume informal leadership roles to benefit the team.

14. Experimental: Ranger Biodata Inventory - Self Esteem
Self Esteem contains items measuring confidence in one's own abilities.

15. Experimental: Ranger Biodata Inventory - Combat Motivation
Combat Motivation contains items measuring one's willingness to be aggressive and confront adversaries when called upon.

16. Experimental: Ranger Biodata Inventory - Need for Achievement
Need for achievement contains items measuring the desire to set and attain difficult work-related objectives.

17. Experimental: Ranger Biodata Inventory - Outdoor Orientation
Outdoor Orientation contains items measuring one's preference for engaging in outdoor activities.

18. Experimental: Ranger Biodata Inventory - Physical Endurance
Physical Endurance contains items tapping the ability to perform demanding physical work without becoming fatigued.

19. Experimental: Ranger Biodata Inventory - Physical Strength
Physical Strength contains items tapping the ability to lift and carry heavy objects.

20. Experimental: Ranger Biodata Inventory - Object Belief
Object Belief contains items designed to tap the tendency to treat others merely as tools for personal gain. It is reverse scored.

Experimental: Forced Choice Assessment of Background and Life Experiences (FCABLE)--Introduction

The Assessment of Background and Life Experiences (ABLE) developed during the Army's Project A was a highly useful temperament instrument. It added substantial incremental validity over that afforded by the Armed Services Vocational Aptitude Battery (ASVAB) in the prediction of Effort and Leadership, Maintaining Personal Discipline, and Physical Fitness and Bearing (Campbell & Zook, 1993; McHenry, Hough, Toquam, Hanson, & Ashworth, 1990). The ABLE is, however, susceptible to response distortion in samples of applicants who are highly motivated (DeMatteo, White, Teplitzky, & Sachs, 1991).

The Forced Choice ABLE (FCABLE) was developed to overcome the problem of socially desirable responding while retaining the construct validity of the original ABLE. In trying to find a way to reduce social desirability, ARI researchers chose a partially ipsative format. (There are several problems with fully ipsative formats.) In this format, the respondent is given four statements--two positive statements and two negative statements. S/he is asked to select (out of the four statements) which one is "most like me" and the one that is "least like me." Respondents have not reacted negatively to the format in field tests of the FCABLE.

FCABLE statements were written to reflect five ABLE construct scales:

- Work Orientation
- Dominance
- Dependability
- Agreeableness
- Emotional Stability

FCABLE also contains a Social Desirability scale. It has 30 items and takes about 30 minutes with instructions.

Psychometrics:

Subgroup Differences: There are gender differences on the ABLE scales. Females have higher scores on Nondelinquency, Internal Control, and Self-Knowledge; males have higher scores on the physical condition scale. White et al., (1993) report that females score higher on Dependability, Work Orientation and Cooperation; while males score higher on Emotional Stability, Dominance and Physical Condition yet notes that these differences are rather small. Race differences on the ABLE are small but blacks have slightly higher scores than whites on 7 of the 11 scales. White et al. (1993) report a .23 STD difference with blacks scoring higher than whites on 10 of 11 scales. Hispanics tend to be more conscientious, less delinquent, and respond in less socially desirable manner than whites.

ABLE-FCABLE Correlations: FCABLE scales correlate between .60 an .70 with their ABLE counterparts. These intercorrelations are slightly lower than, but in the same ballpark as the ABLE's test-retest reliabilities which range from .64 to .84.

ABLE Validity: Concurrent validity obtained as part of the Project A research: a sample of 18 MOS with 9430 personnel had the following results: Dependability correlated highly with Personal Discipline, those with low scores on Nondelinquency scales had the most Articles 15s and other disciplinary actions. As expected Physical Condition scale predicted physical fitness. Project A longitudinal sample had Emotional Stability and Nondelinquency the best predictors of 1 year and 36 month attrition respectively.

Correlations between ABLE scales and second tour NCO performance: Dominance and Work Orientation predicted Leadership Achievement (r=.30-.34, $p<.05$); Self-Esteem, Dominance, Internal Control and Conscientiousness were related to training and counseling subordinates (r=.15-.20, $p<.05$).

FCABLE Validity: Initial results from administrations of the FCABLE to Rangers suggest that it predicts advancement, the number of badges, commendations, and other awards received, and entry-level attrition.

The FCABLE will be administered experimentally in SFAS in the fall of 1994.

21. Experimental: FC-ABLE Work Orientation
Work orientation assesses the tendency to strive for competence in one's work. The work-oriented person works hard, sets high standards, tries to do a good job, endorses the work ethic, and concentrates on and persists in the completion of the task at hand. The less achievement-oriented person has little ego involvement in his or her work, does not expend much effort, and does not feel that hard work is desirable.

22. Experimental: FC-ABLE Dominance
Dominance is defined as the tendency to seek and enjoy positions of leadership and influence over others. The highly dominant person is forceful and persuasive when adopting such appropriate behavior. The relatively non-dominant person is less likely to seek leadership positions and is timid about offering opinions, advice, or direction.

23. Experimental: FC-ABLE Dependability
Dependability assesses a person's tendency to be reliable. The person who scores high on this scale is well organized, planful, prefers order, thinks before acting, and holds him- or herself accountable. The person who scores low tends to be careless and disorganized, and acts on the spur of the moment.

24. Experimental: FC-ABLE Agreeableness
Agreeableness assesses the degree of pleasantness versus unpleasantness a person exhibits in interpersonal relations. The agreeable and likeable person is pleasant, tolerant, tactful, helpful, not defensive and is generally easy to get along with. His or her participation in a group adds cohesiveness rather than friction. A disagreeable and unlikable person is critical, faultfinding, touchy, defensive, alienated, and generally contrary.

25. Experimental: FC-ABLE Emotional Stability
Emotional stability assesses the amount of emotional stability and tolerance for stress a person possesses. The well-adjusted person is generally calm, displays an even mood, and is not overly distraught by stressful situations. He or she thinks clearly and maintains composure and rationality in situations of actual or perceived stress. The poorly adjusted person is nervous, moody, and easily irritated, tends to worry a lot, and "goes to pieces" in times of stress.

Experimental: Army Vocational Interest Career Examination (AVOICE) Introduction

The Army Vocational Interest Career Examination (AVOICE) is an occupational interest instrument that was developed to measure vocational interests relevant to jobs in the Army. The Air Force's Vocational Interest Career Examination was used as a model early in its development. It has 166 items and takes about 20 minutes to administer.

The AVOICE has 22 specific scales that are organized into eight composite scores:

- Rugged/Outdoors
- Audiovisual Arts
- Interpersonal
- Skilled Technical
- Administrative
- Food Service
- Protective Service
- Structural/Machines

Psychometrics:

Correlations with other constructs: Correlations between interest in an occupational field (measured with the AVOICE) and job satisfaction averages less than .20. However this is suggested to be due to range restriction in the measure of job satisfaction (Carter, 1991). A study which looked at vocational interests and actual job performance found reasonable correlations between interest and the "can do" aspects of the job (Technical Proficiency r=.44; General Soldiering Proficiency r=.44) and slightly lower correlations with the "will do" aspects of the job (Effort and Leadership r=.38; Personal Discipline r.35; Physical Fitness and Military Bearing r=.38). Interests however, do not typically increment prediction of job performance above that predicted by cognitive and personality variables.

Subgroup Differences: Females tend to score higher than males in Audiovisual Arts, Interpersonal, Administrative, and Food Service. Males score higher on Rugged/Outdoors, Structural/Machines, and Protective Services. Blacks score higher than whites on all but two of the eight composites--Rugged/Outdoors, and Protective Services.

Experimental: Army Vocational Interest Career Examination (AVOICE) Introduction

Reliability:	alpha	test-retest
AVOICE scale (items)	N=8224-8493	N=389-409
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	.70
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

Jobs used for in the past: A number of studies suggest that the VOICE or the AVOICE is able to differentiate between occupations within the military (Personnel Selection and Classification: New Directions; cf..pg. 22)

Validity Evidence: Campbell & Zook (1991) report validities for the AVOICE in predicting first-tour job performance. These correlations, corrected for range restriction, are:

Criterion	Multiple Corr.
Core Tech. Proficiency	.38
General Soldiering Proficiency	.37
Effort and Leadership	.17
Maintaining Personal Discipline	.05
Physical Fitness and Military Bearing	.05

The AVOICE fails to add incremental variance over the ASVAB for any of the criteria.

26. Experimental: AVOICE-Rugged/Outdoors
The Rugged/Outdoors composite is composed of three subscales: Combat, Rugged Individualism, and Firearms Enthusiast.

27. Experimental: AVOICE-Audiovisual Arts
The Audiovisual Arts composite is composed of three subscales: Drafting, Audiographics, and Aesthetics.

28. Experimental: AVOICE-Interpersonal
The Interpersonal composite is composed of two subscales: Medical Services and Leadership/Guidance.

29. Experimental: AVOICE Skilled/Technical
The Skilled/Technical composite is composed of four subscales: Science/Chemical, Computers, Mathematics, and Electronic Communications.

30. Experimental: AVOICE Administrative
The Administrative Composite is composed of two subscales: Clerical/Administrative and Warehousing/Shipping.

31. Experimental: AVOICE-Food Service
The Food Service composite is composed of two scales: Food Service Professional and Food Service Employee.

32. Experimental: AVOICE Protective Service
The Protective Service composite is composed of two scales: Fire Protection and Law Enforcement.

33. Experimental: AVOICE Structural/Machines
The Structural/Machines composite is composed of four scales: Mechanics, Heavy Construction, Electronics, Vehicle Operator.

Experimental: Job Orientation Blank (JOB) Introduction

The Job Orientation Blank (JOB) taps individual's preferences for work environments. It asks the respondent to rate the strength of his/her preferences for certain job outcomes (e.g., rewards) on a 7-point scale. Theoretically, it is based on studies of job satisfaction and work motivation. It has 31 items that are grouped into six scales. The six scales are combined to form three composites:

	Composite	Subscales
●	High Expectations	Job Pride, Job Security/Comfort, Serving Others, and Ambition
●	Routine	Job Routine
●	Autonomy	Job Autonomy

Correlations among constructs: High expectations is correlated .09 with Routine and .31 with Autonomy; Autonomy and Routine are correlated .08 (N=6116).

Subgroup Differences: The Career Force data show that females score higher than males on High Expectations (by 1/3 of an SD) and slightly higher on Routine (by less than 1/10th of an SD). Blacks score higher than whites on Routine (.43 SD) and on High Expectations (.32 SD).

Reliability:

Items	**Subtest**	**alpha**	**N**
10	Job Pride	.79	6309
6	Job Security/Comfort	.76	6322
3	Serving Others	.80	6290
4	Job Autonomy	.59	6228
4	Job Routine	.63	6234
4	Ambition	.67	6239

In the Career Force study the data loaded on 3 factors High Expectations, Routine, and Autonomy the Spearman Brown reliabilities were: .84, .65 and .47 respectively.

Validity Evidence: Campbell & Zook (1991) reported validities of JOB factors for predicting first tour job performance criteria. These correlations, corrected for range restriction are:

	Multiple Corr.
Core Tech. Proficiency	.29
General Soldiering Proficiency	.29
Effort and Leadership	.18
Maintaining Personal Discipline	.06
Physical Fitness and Military Bearing	.06

The JOB added incremental variance over the ASVAB for predicting ratings of Physical Fitness and Military Bearing (.01).

34. Experimental: JOB - High Expectations

High Job Expectations is comprised of the following:

Job Pride includes preferences for work environments that are characterized by such positive characteristics as friendly coworkers, fair treatment, and comparable pay. Persons who score high on this scale like the work environment to allow them to feel a sense of accomplishment and to receive recognition for accomplishment.

Job Security/Comfort includes preferences for work environments that provide secure and steady employment, environments where persons receive good training and can utilize their abilities.

Serving Others includes preferences for work environments where persons are reinforced for doing things for other people and for serving others through the work preformed.

Ambition measures preferences for work environments that have prestige and status. Persons who score high on this scale prefer work environments that have opportunities for promotion and for supervising or directing others' activities.

35. Experimental: JOB- Routine

Job Routine includes preferences for work environments that lack variety, where people do the same or similar things every day, have about the same level of responsibility for quite a while, and follow others' directions.

36. Experimental: JOB- Autonomy

Job Autonomy includes preferences for work environments that reinforce independence and responsibility. Persons who score high on this construct prefer to work alone, try out their own ideas, and decide for themselves how to get the work done.

Proposed: Job Compatibility Questionnaire (JCQ) Introduction

A Job Compatibility Questionnaire (JCQ) could be developed for Special Forces. A JCQ contains statements describing job activities and characteristics (e.g., listen to angry people vent their problems, convince customers to buy a product). Job applicants use a forced-choice format to indicate the characteristics they prefer. The key is composed of statements that job incumbents and supervisors rate as highly job descriptive.

The JCQ focuses on the person-job fit (Bernardin, 1989)--the degree to which characteristics of a job satisfy individuals' preferences--and is intended to be consistent with the Theory of Work Adjustment. JCQs have proven useful for customer service representative and telephone interviewer jobs (Bernardin, 1987; Villanova & Bernardin, 1990), and are intended for use in placement and classification decisions. JCQs are based on job content and have a good deal of face validity. The forced-choice format makes the JCQ less prone to response distortion. JCQs have been shown to predict job performance and turnover criteria (Villanova, Bernardin, Johnson, & Dahmus, 1994).

Like the FCABLE, the JCQ uses a partially ipsative format. Each item presents four job characteristics or situations. On some items, four undesirable characteristics are presented and the respondent must indicate the two most undesirable choices. On other items, four desirable characteristics are presented and the respondent must select the two most desirable choices.

A JCQ with these five scales could be developed for SF jobs:

- Special Forces
- Weapons
- Engineering
- Communications
- Medic

Psychometrics:

Scoring: A priori scoring keys are developed based on job incumbents' ratings of the degree to which characteristics are relevant to their jobs. Individuals receive high scores if their preferences match job characteristics that are relevant to the job.

Correlations with other constructs: JCQs have yielded correlations with a measure of cognitive ability ($r=.30$, $p<.05$) and numerical ability ($r=.44$, $p<.05$) (Villanova et al., 1994).

Reliability: Villanova et al. (1994) report an internal consistency estimate of .65.

Validity Evidence: The JCQ has been found to correlate with voluntary termination of customer service personnel (Bernardin, 1987), intentions to quit for fast food personnel (Bernardin, 1989) and criteria for telephone interviewers (Villanova and Bernardin, 1990).

37. Proposed: JCQ SF Scale
The SF scale would include job activities and characteristics that are common to all SF jobs such as: teaching, interacting with indigenous people, getting along with others, contributing to the effectiveness of the team, navigating and surviving in the field, exhibiting effort and motivation.

38. Proposed: JCQ Weapons Scale
The Weapons scale would contain job tasks and duties that are central to the Weapons sergeant job, such as tasks that involve loading, firing, assembling, and disassembling direct and indirect fire weapons.

39. Proposed: JCQ Engineering Scale
The Engineering scale would contain job tasks and duties that are central to the SF Engineering sergeant job, such as tasks that involve emplacing or detonating mines or explosives or building structures or bridges.

40. Proposed: JCQ Communications Scale
The Communications scale would contain job tasks and duties that are central to the SF Communications sergeant job, such as tasks that involve using Morse coding, constructing antennas, and operating communication equipment.

41. Proposed: JCQ Medic Scale
The Medic scale would contain job tasks and duties that are central to the SF Medic job, such as evaluating and treating medial conditions and injuries, determining and administering medications and dosages, and maintaining health standards in facilities.

42. Experimental: Organizational Identity

Construct Measured:

Level of individual identification with a psychological group - the organization

Short Description of Test:

This instrument is comprised of 5 questions which ask the respondent to rate the degree they agree or disagree a the statement. Questions ask the respondent to rate the degree they identify with the specific organization from several different perspectives.

Example: "This organization's successes are my successes." (Mael & Tetrick, 1992)

Number of Items: 5 items Time Limit: approx. 5 min.

Speededness: N/A Apparatus: Paper and pencil

Psychometrics:

Scoring: The score is the sum of the ratings for the 5 items.

Correlations with other constructs: Correlation with Organizational Commitment r=.77; Org. Satisfaction r=.55; Job Satisfaction r=.55 Job Involvement r=.55, Ns=263 (Mael & Tetrick, 1992). Org. Distinctiveness r=.17; Org. Prestige r=.32 N=297 (Mael & Ashforth, 1992).

Reliability: alpha coefficients: .81 N=263 (Mael & Tetrick, 1992); .87 N=297 (Mael & Ashforth, 1992); .88 N=1012 (Mael & Alderks)

Fakability: The items are fairly transparent in intent - identification with the organization. Thus, the individual is able to "fake" higher levels of organizational identification.

Jobs used for in the past: Used to look at alumni identification with their alma mater (Mael & Ashforth, 1992); comparison of perceived team cohesion across Army platoons (Mael & Alderks).

Validity Evidence: High levels of organization identification of alumni with their alma maters were associated with their subsequent contributions to the alm mater .38, willingness to advise (a) a son and (b) others to attend the school (r=.43 r=.39, respectively); and reports of attending special lectures through the alma mater r=.40 (Mael & Ashforth, 1992).

Mael and Ashforth report that organizational identification with the Army significantly predicted attrition across 24 months (N=1021). These Zero-order correlations range from .30 at 6 months to .12 at 24 months (both significant at the p<.01 level).

43. Experimental: Occupational Stress Assessment Inventory (OSAI)

Construct Measured:

The ability to cope in stressful or changing environments.

Short Description of Test:

The Occupational Stress Assessment Inventory (OSAI) measures individual's coping abilities (Heslegrave & Colvin, 1994). It is a written measure which includes 26 stressful situations. The respondent makes four judgments about each situation, and those judgments yield four scores:

- how stressful s/he finds the situation (stress);
- whether s/he would cope with or change the situation (change);
- how effectively s/he would cope relative to others (others); and
- what coping strategies s/he would use in the situation (coping).

Psychometrics:

Scoring: The score is the sum of the ratings for each subscale across situations.

Correlations with other constructs: For all of the following correlations, $p<.05$ and $N=41$. The 'others' subscale is correlated with several problem-focused coping strategies from the Ways of Coping Questionnaire (planning $r=.39$; suppressing competing activities $r=.31$; restraint coping $r=.34$; reinterpretation and growth $r=.47$). The 'others' subscale is also correlated with Self-Efficacy $r=.45$; and Eysenck's Lie and Neuroticism scales ($r=.25$, $-.44$). The 'coping' subscale correlates with Active Coping Strategy $r=.33$); the 'change' subscale correlated with Self-Efficacy ($r=.45$), the Eysenck Lie Scale ($r=.25$) and neuroticism ($r=-.36$).

Subgroup Differences: The only gender differences were on the 'coping' scale with females using more active coping styles than males. Results on the 'stress' subscale indicate that Asian-Americans and Latinos tend to view situations as significantly less stressful than European-Americans and African-Americans.

Reliability: 'Stress' subscale .81; 'Change' .86; 'Others' .88; 'Coping' .50; N=41; Heslegrave & Colvin (1994)

Fakability: The 'stress' subscale taps an individual's perceived ability to cope with stress, in a selection situation this subscale may be fakable.

Validity Evidence: In predicting actual performance on criteria under stressful conditions, the OSAI and other personality variables proved less consistent than psychophysiological measures (such as heart rate). The 'stress' subscale did have limited success in predicting performance under stress. Perceptions of one's own performance under stress seems to have some relationship to subsequent performance.

44. Social Intelligence (Biodata Measure)

Construct Measured:

This instrument is a biodata measure designed to tap an individuals' effectiveness in social functioning. It is designed to assess four social intelligence constructs (Zazanis, Zaccaro, Diana, Teplitzky, & Gilbert, 1994):

(1) interpersonal perceptiveness-ability to understand or perceive persons,
(2) systems level perceptiveness-ability to be aware and sensitive to the needs, goals, demands, and problems at multiple system levels,
(3) behavioral flexibility- ability to act appropriately in situations and achieve one's socially-oriented goals,
(4) social competence- demonstrates successful social accomplishments.

The instrument is in a paper/pencil format and has 41 biodata items.

Psychometrics:

Scoring: Items are rationally keyed to meaningful constructs associated with social intelligence.

Construct Validity: There is some evidence that this measure correlates with other measures of social intelligence. Zazanis et al. (1994) administered the SI measure and SI marker tests to two samples of SFAS candidates (Ns=189, 528). The scales were moderately correlated with the Lennox and Wolfe self-monitoring scales and Guilford's Test of Social Intelligence.

Reliability:	Study 1 (N=189) alpha	Study 2 (N=528) alpha
1) **interpersonal perceptiveness**	.82	.85
2) **systems level perceptiveness**	.72	.66
3) **behavioral flexibility**	.76	.79
4) **social competence**	.72	.63

Fakability: The instrument is not expected to be highly fakable since it elicits prior life history events (Zazanis et al, 1994).

Validity Evidence: In **Study 1**, (N=189) peer rankings was correlated with social perceptiveness (r=.17, p<.05); interpersonal perception (r=.15, p<.04), system perception (r=.16, p<.05), and with social competence (r=.22, p<.05).
In **Study 2** peer ranking correlated again with social competence (r=.21, p<.01) and with behavioral flexibility (r=.10, p<.05) but did not correlate with the perceptiveness factors as in study 1. These shifts in the correlational framework suggest the need of additional investigation regarding the structure of this measure.

Appendix D
Psychomotor and Physical Measures

1. Experimental: Project A Target Tracking I

Construct Measured:

Control precision-- precision and steadiness of muscular movements.

Short Description of Test:

The task is a pursuit tracking task. Subjects are shown a path consisting of vertical and horizontal line segments. The target box appears at the beginning of the path and moves at a constant rate along the path. At the start of the task a pair of crosshairs are centered in the target box. The task is to keep crosshairs centered on the target at all times through the use of a joystick controlled by one hand. The speed of the crosshairs and the target are manipulated, as well as the length of the path, and average time the target moves along the segment. Tracking accuracy and improvement in tracking performance measured by accuracy measures, time on target, and distance from center of crosshairs to center of target (several times each second) and an average is taken to derive overall accuracy score for the trial. Early development of Target Tracking was based on the AAF Rotary Pursuit Test.

Number of Items: 18 items Time Limit: about 8 minutes
Apparatus: Computer administered with joysticks and a response pedestal

Psychometrics:

Scoring: Distance from the center of the crosshairs to center of the target (i.e., log(distance +1)). Distance scores are reflected so that higher scores are "better."

Correlations with other constructs: The psychomotor tests from Project A (Tracking I, Tracking II, Target Shoot, and Cannon Shoot) were highly correlated with each other and consistently loaded together in factor solutions. In particular, Target Tracking I and Target Tracking II typically correlate with each other in the .90s. Tracking I, Tracking II, and Cannon Shoot also yield moderate correlations with spatial test scores.

Subgroup Differences: In three large samples (N > 6000), males scored higher than females by 1.25 to 1.28 SD; whites scored higher than blacks by .66 to .78 SD; and whites scored higher than Hispanics by .15 to .33 SD.

Reliability: Split-half reliability was .98 in two samples (N=9251; N=6754). Test-Retest reliability were .74 (N=460 with a two-week interval between testing) and .84 (N=313 with a four week interval between testing).

Practice and Coaching Effects: Improvement with practice on psychomotor measures is a common finding (McHenry & Rose, 1988). Two studies have examined practice effects on the Tracking I:

McHenry et al. (1987) administered Tracking I twice, and one group practiced between testing sessions; 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. The control group (N=113) improved slightly on Target Tracking 1 (.07 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.

Validity Evidence: There is evidence that psychomotor tests predict proficiency in military enlisted jobs. McHenry et al. (1990) formed six composites of Project A psychomotor and perceptual test scores (including Tracking 1 and Tracking 2). Mean validity coefficients for the combination of six composites were .53 for the core technical proficiency criterion and .57 for the general soldiering proficiency criterion, which subsume job knowledge and hands-on task proficiency measures. Similar results were obtained when the measures were administered to a longitudinal sample (Oppler, Peterson, & Russell, 1993).

Tracking I and Tracking II have been used by ARI to predict training performance in combat jobs, particularly Infantryman, Cannon Crewmember, and Tube-launched, Optically-tracked, Wire-guided Gunner performance (Busciglio, 1990; Busciglio, Silva, & Walker, 1990; Silva, 1989).

2. Experimental: Project A Target Tracking II

Construct Measured:

Multilimb coordination--two-handed tracking.

Short Description of Test:

The test is similar to Target Tracking I. Subjects are presented a path with horizontal and vertical lines. At the beginning of the path is a target box with crosshairs. The target moves along the path at a constant rate.

The major difference between Tracking I and II is that in Tracking II, the subject moves two sliding resistors to control the crosshairs. [In Tracking I the subject tracks the target using a joystick in one hand.] One controls vertical movement and one controls horizontal movement. The task is to keep crosshairs centered on the target box.

Number of Items: 18 items Time Limit: about 7 minutes
Apparatus: Computer administered with a special response pedestal.

Psychometrics:

Scoring: Distance from the center of the crosshairs to center of the target (i.e., log(distance +1). Distance scores are reflected so that higher scores are "better."

Correlations with other constructs: The psychomotor tests from Project A (Tracking I, Tracking II, Target Shoot, and Cannon Shoot) were highly correlated with each other and consistently loaded together in factor solutions. In particular, Target Tracking I and Target Tracking II typically correlate with each other in the .90s. Tracking I, Tracking II, and Cannon Shoot also yield moderate correlations with spatial test scores.

Subgroup Differences: In three large samples (N > 6000), males scored higher than females by .92 to 1.28 SD; whites scored higher than blacks by .83 to .90 SD; and whites scored higher than Hispanics by .23 to .40 SD.

Reliability: Split-half reliability was .98 in two samples (N=9251; N=6754). Test-Retest reliabilities were .85 (N=460 with a two-week interval between testing) and .91 (N=313 with a four week interval between testing).

Practice and Coaching Effects: Improvement with practice on psychomotor measures is a common finding (McHenry & Rose, 1988). Three studies have examined practice effects on Tracking II:

> McHenry et al. (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 .21 SD for Target Tracking 2. The control group (N=113) 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 .24 SD on Target Tracking 2.

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.

Validity Evidence: McHenry & Rose (1989) conducted a meta-analysis of psychomotor predictors. They found that measures of Multilimb Coordination have been effective predictors of criteria for pilot, aircrew, infantry and combat jobs.

There is evidence that psychomotor tests predict proficiency in military enlisted jobs. McHenry et al. (1990) formed six composites of Project A psychomotor and perceptual test scores (including Tracking 1 and Tracking 2). Mean validity coefficients for the combination of six composites were .53 for the core technical proficiency criterion and .57 for the general soldiering proficiency criterion, which subsume job knowledge and hands-on task proficiency measures. Similar results were obtained when the measures were administered to a longitudinal sample (Oppler, Peterson, & Russell, 1993).

Tracking I and Tracking II have been used by ARI to predict training performance in combat jobs, particularly Infantryman, Cannon Crewmember, and Tube-launched, Optically-tracked, Wire-guided Gunner performance (Busciglio, 1990; Busciglio, Silva, & Walker, 1990; Silva, 1989).

3. Experimental: Project A Target Shoot Test

Construct Measured:

Psychomotor precision and steadiness.

Short Description of Test:

At the start of the trial, a target box and crosshairs appear at different locations on the computer screen. The target moves about the screen in an unpredictable manner, changing speed and direction. The subject moves crosshairs with a joystick with the goal of keeping it centered in the target box, and then to fire on the target. This must be accomplished before the end of the time limit for the trial. Parameters vary trial to trial - maximum speed of crosshairs, average speed of target, difference between the 2 speeds, number of changes in target speed, number of line segments comprising path of target, time required for target to travel segments. This test was modified after tests used by AAF in Aviation Psychology Program.

Number of Items: 30 items Time Limit: about 5 minutes

Apparatus: Computer administered with a special response pedestal

Psychometrics:

Scoring: Scores include: (a) distance from the center of the crosshairs to center of the target (i.e., log(distance +1) and (b) the time elapsed from trial onset until subject fires at the target. Distance and time scores are reflected so that higher scores are "better."

Correlations with other constructs: The psychomotor tests from Project A (Tracking I, Tracking II, Target Shoot, and Cannon Shoot) were highly correlated with each other and consistently loaded together in factor solutions. In particular, Target Tracking I and Target Tracking II typically correlate with each other in the .90s. Tracking I, Tracking II, and Cannon Shoot also yield moderate correlations with spatial test scores.

Subgroup Differences: In three large samples (N > 6000), males scored higher than females by .63 to .90 SD; whites scored higher than blacks by .23 to .25 SD; and white-hispanic effect sizes ranged from -.04 to .13 SD.

Reliability: Split-half reliabilities were .85 and .84 in two samples (N=9251; N=6754). Test-Retest reliability was .58 (N=460 with a two-week interval between testing).

Practice and Coaching Effects: Improvement with practice on psychomotor measures is a common finding (McHenry & Rose, 1988). Gains similar to those for the tracking tests can probably be expected.

Validity Evidence: McHenry & Rose (1989) conducted a meta-analysis of psychomotor predictors. They found that measures of psychomotor precision and steadiness have not been included in many validity studies. There is evidence that psychomotor tests predict proficiency in military enlisted jobs. McHenry et al. (1990) formed six composites of Project A psychomotor and perceptual test scores. Mean validity coefficients for the combination of six composites were .53 for the core technical proficiency criterion and .57 for the general soldiering proficiency criterion, which subsume job knowledge and hands-on task proficiency measures. Similar results were obtained when the measures were administered to a longitudinal sample (Oppler, Peterson, & Russell, 1993).

4. Experimental: Cannon Shoot Test

Construct Measured:

This test taps movement judgment, the ability to judge the relative speed and direction of one or more moving objects to determine where those objects will be at a given point in time, or when objects will intersect.

Short Description of Test:

Subjects fire a cannon at a moving target. At the start of the trial, a stationary cannon appears on the computer screen. The cannon is able to fire a shell that travels at a constant speed on each trial. After the cannon appears a circular target appears and moves in a constant direction and rate of speed (though the speed varies trial to trial). The subject must fire a shell so that the shell intersects with the target as the target crosses the shell's line of fire. The angle of target movement relative to the cannon, the distance from cannon to point of contact, distance from impact point to fire point are all varied.

Number of Items: 36 items Time Limit: about 7 minutes
Apparatus: Computer administered with a special response pedestal.

Psychometrics:

Scoring: The primary score is the deviation score, or difference between time of fire and optimal fire time (direct hit has a deviation score of 0). The deviation score is reflected so that higher scores are "better."

Correlations with other constructs: The psychomotor tests from Project A (Tracking I, Tracking II, Target Shoot, and Cannon Shoot) were highly correlated with each other and consistently loaded together in factor solutions. In particular, Target Tracking I and Target Tracking II typically correlate with each other in the .90s. Tracking I, Tracking II, and Cannon Shoot also yield moderate correlations with spatial test scores.

Subgroup Differences: In three large samples (N > 6000), males scored higher than females by .84 to .99 SD; whites scored higher than blacks by .45 to .55 SD; and whites scored higher than Hispanics by .07 to .12 SD.

Reliability: Split-half reliabilities were .65 and .64 in two samples (N=9251; N=6754). Test-Retest reliability was .52 (N=460 with a two-week interval between testing).

Practice and Coaching Effects: Improvement with practice on psychomotor measures is a common finding (McHenry & Rose, 1988). Gains similar to those for the tracking tests can probably be expected.

Validity Evidence: McHenry & Rose (1989) conducted a meta-analysis of psychomotor predictors. They found that measures of movement judgment have not been included in many validity studies. There is evidence that psychomotor tests predict proficiency in military enlisted jobs. McHenry et al. (1990) formed six composites of Project A psychomotor and perceptual test scores. Mean validity coefficients for the combination of six composites were .53 for the core technical proficiency criterion and .57 for the general soldiering proficiency criterion, which subsume job knowledge and hands-on task proficiency measures. Similar results were obtained when the measures were administered to a longitudinal sample (Oppler, Peterson, & Russell, 1993).

5. Operational: Army Physical Fitness Test (APFT)

Construct Measured:

upper body strength and physical fitness

Short Description of Test:

The Army Physical Fitness Test (APFT) is administered during in-processing for the Special Forces Assessment and Selection (SFAS) program. It has three components: sit-ups, push-ups, and a two-mile run. For sit-ups and push-ups individual are told to do as many as possible in a one minute period; the score is a count of the number of repetitions. The two-mile run is timed.

Psychometrics:

Scoring: Counts of push-ups and sit-ups and the run time are compared against standards established for 17 to 21 year olds to derive scores. The total APFT score is the sum of the three component scores.

Correlations with Other Measures: Teplitzky (1990) reported an average correlation of .34 between APFT and Ruckmarch scores across SFAS classes in FY 89, 90, and 91, each with N > 2000.

Validity Evidence: For Project A/Career Force Physical Fitness and Bearing measured during soldiers' first tour was correlated .46 with Physical Fitness and Bearing measured as in the second tour (NCO).

Burke & Dyer (1984) found that APFT events predicted success in Ranger training and were correlated with the occurrence of nonserious injuries during training.

6. Operational: SFAS Physical Endurance Composite

Construct Measured:

Physical strength, physical and mental endurance, perseverance

Short Description of Test:

Physical endurance is a composite of scores on three exercises: a ruckmarch, a battle march, and an obstacle course. Although the specific standards and conditions of the exercises are considered sensitive, the exercises resemble infantry exercises used by conventional forces. The ruckmarch involves carrying a heavy load (at least 45 lbs--based on information given to SFAS applicants) for a long distance. The obstacle course involves climbing obstacles 20-30 ft high as well as maneuvering through other types of obstacles. All events are timed.

Psychometrics:

Scoring: Scores on individual events are time scores. They are standardized and then added together to form a composite.

Correlations with Other Measures: Teplitzky (1990) reported an average correlation of .34 between APFT and Ruckmarch scores across SFAS classes in FY 89, 90, and 91, each with N > 2000.

Initial analyses of data from 1989 SFAS classes suggest that the endurance composite is very highly correlated (.80-.90) with the fitness composite (described on the next page).

7. Operational: SFAS Physical Fitness Composite

Construct Measured:

Aerobic fitness and ability to run

Short Description of Test:

A composite score of performance on three timed runs: 2.8 mile, 3.8 mile, and 4.8 mile runs.

Psychometrics:

Scoring: Time scores are converted to T-Scores and then added to form a composite.

Correlations with other measures: Initial analyses of data from 1989 SFAS classes suggest that the endurance composite is very highly correlated (.80-.90) with the fitness composite (described on the next page).

8. Operational: SFAS Swim Test
Construct Measured: Ability to swim. Short Description of Test: SFAS candidates must swim 50 meters unassisted wearing combat boots. Psychometrics: Scoring: Pass/fail Subgroup Differences: Blacks tend to fail the swim test at a much greater proportion than whites. During FY 1991 for example, 15.4% of blacks failed the SFAS swim test as opposed to 2.8% of whites.

Appendix E
Simulations, Administrative Indices
Ratings and Other Performance Measures

1. Operational: Self-Development Test (SDT)

Construct Measured:

This test taps the skills and competencies needed for NCO leadership development and is a measure of personal motivation for self-development.

Short Description of Test:

The SDT is a formally administered test with 100 items. The test is used to evaluate self-development progress and provides a guide of development need areas. The content areas of the SDT are:

1. Army Leadership - 20 questions
2. Training Management Principles - 20 questions
3. MOS Knowledge - about 60 questions on MOS specific knowledge

Psychometrics:

Scoring: 0 - 100 range in scores which vary based on MOS. The average score is 78%.

Correlations with other constructs: SDT is likely related to scores on the SQT. The tests have similar content areas (MOS knowledge).

Subgroup Differences: Results indicate that there is a .12 SD difference in 9 Conventional Army MOS (from Project A; 11B, 13B, 19E, 31C, 63B, 64C, 71L, 91A, 95B) with males outscoring females. Three MOS that showed greatest gender differences (71M, 88M, 91C) show a .39 SD difference in favor of males (Silva; 1994) . Results from the 9 Project A MOS indicate a .45 SD difference with whites outscoring blacks. The MOS with the greatest race differences (12C, 63B, 63H) have a .94 SD difference in favor of whites.

Reliability: Reliability information should become available in Fall 1994.

Practice and Coaching Effects: A different test is developed each year which should minimize the scoring advantage of those who retake the test.

Jobs used for in the past: To this point the SDT has been used as a tool to aid NCOs in their own self-development. The SDT will be used in school selection and promotion decisions in FY94 for those in active component, and will also be used for reserves in FY95.

Validity Evidence: Validities are unavailable at this time, but should become available in Fall, 1994. The test does have considerable face and content validity given the content development is based on technical manuals and tests are developed by 1) Leadership component by Center for Army Leadership; 2) Training Management by US Army Sergeants Major Academy; 3) MOS knowledge by MOS proponent.

2. Operational: SFAS Military Orienteering (MO)

Construct Measured:

Basic navigation skills, ability to function under stress, and motivation to succeed are required to complete these exercises (Busciglio & Teplitzky, 1990).

Short Description of Test:

There are six military orienteering (MO) exercises, four take place during the day and two at night. They occur at the end of Phase I events, between day 7 and 10 of the SFAS. The first two day time events are followed with night time events. The last two events occur during the day and are considerably longer (4-6 hours; and 7-10 hours respectively). In this exercise the soldier is equipped with a heavy rucksack, is taken to an undefined location and must navigate his way from one point to the next.

Psychometrics:

Scoring: Soldiers receive a time score and a rating score which is based on their time (3=outstanding; 2=satisfactory; 1=unsatisfactory). If a soldier does not finish the exercise they receive no time score and are rated as unsatisfactory (3). Cadre also observe the performance of the soldier to evaluate his motivation, physical endurance, and ability and comfort navigating at night (Busciglio & Teplitzky, 1994). Factor analyses of rating and time scores suggest that it is reasonable to develop two composites, a ratings composite and a time score composite. The ratings and time score composites are highly correlated with each other.

Correlations with other constructs: The ratings and time composite scores have virtually identical relationships with spatial scores. Busciglio and Teplitzky (1990) found correlations (N=398) between the MO exercise **time** composite and the Map test -.32, the Orientation test -.26; and the Maze -.24. Similarly, the **ratings** composite correlates .33 with the Map test, .30 with the Orientation test, and .24 with the Maze test.

Validity Evidence: Busciglio and Teplitzky (1994) found that **time** for one (Task IV) of the MO exercises (N=167) and **ratings** for two events (Day II, and Task IV) are related to passing land navigation in the SFQC on the first try.

3. Operational: SFAS Peer Rankings

Construct Measured:

Leadership potential

Short Description of Test:

After four days in SFAS peers rank each individual in their group (aside from themself) regarding leader potential. Rankings are also collected at the end of each of the events (Situation Reaction). These peer rankings are used to rate each team member according to their contribution to the team's overall performance. This includes their effectiveness as a team member and their effectiveness during their rotation in the team leader role.

Psychometrics:

Scoring: Ranks for each soldier are averaged across all raters to produce an overall rank.

Jobs used for in the past: The SFAS board uses these rankings in making pass/fail graduation decisions.

Validity Evidence: There is some evidence that the peer evaluations in SFAS may be more psychometrically sound than the cadre ratings. The peer evaluations during SFAS are a good predictor of success in Q-course (personal communication, Zazanis, 1994).

4. Operational: SFAS Situation Reaction Exercises (SR)

Construct Measured:

The Situation Reaction (SR) exercises purport to assess the individuals' level of Trustworthiness, Responsibility, Motivation, Stability, Intelligence, Communication, Physical Fitness, Teamwork, Decisiveness, Judgment and Leadership.

Short Description of Test:

The Situation Reaction Exercises are among the Phase II exercises in SFAS. They are ten realistic job simulation exercises that usually involve 12-man team participation. The SRs are commonly considered the most difficult aspect of the SFAS training. Team leaders are assigned to each exercise, and each soldier acts as the leader in at least one exercise. The leader must develop and implement the plan for the team, organize the team and direct them in the plan. Ratings are made along the above mentioned dimensions.

Psychometrics:

Scoring: Performance is judged along a 3 point scale, 3=outstanding; 2=satisfactory; 1=unsatisfactory. Ratings are only required during those events when the individual is leader of the exercise. A tally is made of the number of outstanding and unsatisfactory ratings the individual receives.

Data collection practices result in missing data (e.g., a rating of 2 is "assumed" and therefore is omitted). To analyze SR ratings data, one must make assumptions about missing data which may or may not be true.

Even so, factor analyses of the SFAS data between the years of 1989 and 1993 show some evidence for 2 or 3 underlying factors: (1) **Effort and Dependability** consisting of Responsibility, Motivation, Teamwork, Stability, Communication, and Intelligence; (2) **Judgment** consisting of Judgment and Decisiveness; (3) **Physical Fitness** consists of the Physical Fitness score. Trustworthiness and Influence loadings on factors flucuate drastically across years and should probably be dropped from across-year analyses.

Reliability: No investigations of reliability of the ratings has been completed.

5. Proposed: Personnel File Form - Number of Awards and Certificates

Construct Measured:

This measure taps items that reflect recognition of exceptional job performance. Awards are weighted based on their importance to the Army. Respondents also indicate memoranda they have received that recognize situations of outstanding performance.

Short Description of Test:

The Personnel File Form was used during Project A to collect information regarding the number of awards and certificates that individuals had received. Respondents reported the number of certificates of appreciation, commendation, and achievement they had received. The awards section contained a checklist of awards and a section to write in additional awards that had not been included.

In the study of first-tour job proficiency (Project A/Career Force data), Awards and Certificates were included in the **Effort/Leadership** (ELS) variable combined with Army-wide Ratings of Effort/Leadership and Overall Effectiveness Ratings; MOS Specific Rating Scales. In the study of second-tour NCO job proficiency, Number of Awards was included in the **Effort and Achievement** (EA) variable combined with Army-wide Ratings of Technical Skill/Effort and Overall Effectiveness; and Average MOS Ratings.

Psychometrics:

Scoring: A tally of awards and memoranda that were received in grades E-4 and above were used in the Project A study.

Correlations with other constructs: The Project A Administrative Index for Awards (for Second Tour Performance) correlated with the other Administrative Indices (Ns=817 to 1,035, correlations are significant at $p<.01$): Article 15/Flag Actions $r=-.08$ (as is expected); Physical Readiness $r=.13$; M16 Qualification $r=.14$; Military Training $r=.31$; Promotion Rate $r=.31$.

Fakability: Project A included a pilot test which indicated that self-report of the administrative indices appeared to be the most up-to-date information available (Riegelhaupt et. al., 1987). The fact that answers can be verified against existing records, increases the veracity of the self-report items.

Validity Evidence: Number of Awards received is dependent in part on assigned military post and the MOS of the respondent. Certain Posts have more advanced training and also administer more letters and certificates to foster a greater level of esprit de corps.
The ELS variable, for first-tour performance was correlated with several second-tour performance variables (Ns range 333-413): Core Technical Proficiency $r=.25$; General Soldiering Proficiency $r=.22$; Effort and Achievement $r=.45$; Leadership $r=.38$; Maintain Personal Discipline $r=.12$; Physical Fitness and Bearing $r=.22$; Rating of Overall Effectiveness $r=.35$.

6. Proposed: Personnel File Form - Number of Article 15's and Flag Actions

Construct Measured:

Both measures are considered indicators of poor soldier performance.

Short Description of Test:

The Personnel File Form was used during Project A as a self-report measure to collect information regarding the number of Articles 15s and Flag Actions that individuals had received. The Project A/Career Force data of first and second-tour job proficiency, the Administrative Index - Number of Article 15's and Flag Actions was included in the **Maintaining Personal Discipline** (MPD) variable combined with Army-wide Ratings of Personal Discipline. In the Promotion Grade Deviation Score for the first-tour study, another Administrative Index was also included.

Psychometrics:

Scoring: A tally was computed of the total number of Article 15's and Flag Actions the soldier had received in paygrades E-4 and above (Campbell & Zook, 1990).

Correlations with other constructs: The Project A Administrative Index for Number of Article 15's and Flag Actions (LVII study) correlated with the other Administrative Indices (Ns=817 to 1,035; all correlations are significant at $p<.01$ except as identified): Awards $r=-.08$; Physical Readiness $r=-.11$; M16 Qualification $r=-.03$ ns; Military Training $r=-.16$; Promotion Rate $r=-.19$.

Subgroup Differences: There were no differences in the receipt of Article 15/Flag by demographic group.

Reliability: Riegelhaupt et al, (1987) report that there are low base rate problems with this variable. 11% of the entire sample have a score on this variable. It may be best for reliability purposes not to use this variable alone, but rather as one component of a composite.

Fakability: Project A included a pilot test which indicated that self-report of the administrative indices appeared to be the most up-to-date information available (Riegelhaupt, et. al., 1987). The fact that answers provided can be verified against existing records, increases the veracity of the self-report items.

Validity Evidence: The first-tour job proficiency MPD variable was correlated with subsequent second-tour performance variables (Ns range 333-413): Core Technical Proficiency $r=.08$; General Soldiering Proficiency $r=.09$; Effort and Achievement $r=.28$; Leadership $r=.27$; Maintain Personal Discipline $r=.26$; Physical Fitness and Bearing $r=.14$; Rating of Overall Effectiveness $r=.25$.

7. Proposed: Personnel File Form - Promotion Rate

Construct Measured:

Promotion rate is a measure of the soldiers' progress and movement through the enlisted ranks. This measure should have a direct relationship with the soldiers' performance and success in their position.

Short Description of Test:

Using the Personnel File Form respondents report recommendations for promotion which occurred prior to having put in the requisite time in grade. In addition, grade deviation score was calculated from the Enlisted Master File. These two measures were used to form "Promotion Rate."

For the Project A/Career Force data for first-tour job proficiency, Promotion Grade Deviation Score was included in the **Maintaining Personal Discipline (MPD)** variable combined with Army-wide Ratings of Personal Discipline and an Administrative Index for Number of Article 15's and Flag Actions.

In the Second-tour NCO data, Promotion Rate was included in the **Leadership (LEAD)** variable combined with Army-wide Ratings of Leading/Supervisory skills, the Situation Judgment Test, and Role Play exercises.

Psychometrics:

Scoring: The grade deviation score was calculated from the Enlisted Master File. This measure adjusts the soldier's paygrade to the mean of those who have the same time in service. This was combined with an indicator of whether soldier had been recommended for promotion in the secondary zone. These two measures were used to form "Promotion Rate."

Correlations with other constructs: The Project A Administrative Index for Promotion Rate (LVII study) correlated with the other Administrative Indices (Ns=817 to 1,035; all correlations are significant at $p<.01$ except as identified): Awards $r=.31$; Article 15's/Flag Action $r=-.19$; Physical Readiness $r=.14$; M16 Qualification $r=.14$; Military Training $r=.39$.

Validity Evidence: The MPD variable for first-tour performance was correlated with subsequent Second Tour Performance variables (ns range 333-413): Core Technical Proficiency $r=.08$; General Soldiering Proficiency $r=.09$; Effort and Achievement $r=.28$; Leadership $r=.27$; Maintain Personal Discipline $r=.26$; Physical Fitness and Bearing $r=.14$; Rating of Overall Effectiveness $r=.25$.

8. Proposed: Work and Training Portfolio

Construct Measured:

Conventional Army Experience

Short Description of Test:

Project A/Career Force data suggests that first-tour job proficiency is a good predictor of second-tour job proficiency (NCO). This instrument will extend that principle to SF; it will be designed to measure conventional Army experiences (likely to be relevant to SF performance). It will be similar to the Personnel File Form used by the Army during Project A, but it will incorporate ideas from accomplishment record research. The instrument will have three parts:

Training History will ask respondents to document their Army training experiences, focusing on MOS-specific technical skills training, leadership training, and other education.

Work History will ask the respondent to rate himself on a variety of Army-wide tasks and conventional Army tasks that are similar/identical to Special Forces tasks. Such tasks are likely to include: land navigation, communications, first aid, weapons, and other general soldiering/combat-related tasks.

References will ask the respondent to provide names, addresses, and phone numbers of individuals who can verify the information.

Psychometrics:

Scoring: We expect to develop a priori scoring keys in workshops with SF personnel. The participants will be asked to rate the relevance of various conventional Army tasks and experiences to the 26 SF job performance categories that emerged from the analysis of SF jobs.

Fakability: Project A data suggests that self-report of administrative data are reasonably accurate. Self-report information was compared with data from the Enlisted Master File and the Military Personnel Records Jacket. There were relatively few discrepancies, and in some cases the discrepancies were due to out-of-date data files.

One common premise of accomplishment record and biodata instrumentation is that verifiable information is less likely to be faked. Accomplishment records and application blanks often ask for references in order to enhance the veracity of the information.

Validity Evidence: Project A/Career Force data suggest that performance in training predicts performance on the job and, in turn, performance first-tour Army jobs predicts performance in an non-commissioned officer. Taking that finding one step further, measures of conventional army experience may predict performance in Special Forces.

Accomplishment records, scored descriptions of work accomplishments, have yielded useful validities with performance criteria for professional jobs (Hough, Keyes, & Dunnette, 1983).

9. Proposed: Language Training Record

Construct Measured:

Language proficiency

Short Description of Test:

This instrument will be similar to the Personnel File Form used by the Army during Project A, but it will focus on language training and proficiency. Respondents will be asked to document all of their foreign language training experiences (e.g., date attended, course taken, etc.) and to provide levels of proficiency in any languages they have been trained in. They will also be asked to indicate any languages for which they are "native speakers."

Psychometrics:

Scoring: We expect to develop a priori scoring keys in workshops with SF personnel.

Fakability: Language proficiency is easily verifiable, and respondents are, therefore, not expected to exaggerate their experiences.

10. Experimental: Army wide--Performance Rating Scales

Construct Measured:

Dimensions of performance were tapped during Project A are relevant across MOS for all NCO levels. The 12 performance dimensions include: Technical Knowledge/Skill; Effort; Supervising; Following Regulations and Orders; Integrity; Training/Developing; Maintaining Assigned Equipment; Physical Fitness; Self-Development; Consideration for Subordinates; Military Appearance/Bearing; and Self-Control.

Short Description of Test:

We propose to collect information on the above rating scales to tap NCO conventional Army performance (which is likely to be relevant to SF performance). Factor analyses of the rating scales suggest that four factors were consistently identified for Army-wide ratings of performance. The scales are:

Factor	Rating Scale	
1)Leading/Supervising	Supervising Training/Development Consideration for Subordinates	
2)Personal Discipline	Following Regs/Orders Integrity	Self-Control
3)Technical Skill/Effort	Technical Knowledge/Skill Effort	Maintain Equipment
4)Physical Fitness/ Military Bearing	Military Bearing Physical Fitness	

Psychometrics:

Scoring: Individual scores from ratings scales can be used, an average for each dimension, or an average across all dimensions.

Internal Consistency Reliability: Interrater reliabilities were calculated in the Project A data sets. For individual scales intraclass correlation coefficients range from .30-.45. When the Army-wide rating scales and the MOS specific ratings were combined the reliabilities increased to .65 and .55 for supervisors and .58 and .42 for peer ratings. The Career Force longitudinal study showed very similar levels of interrater reliabilities as Project A. As noted here supervisor ratings were consistently more reliable, than peers.

Jobs used for in the past: Supervisory performance rating scales were developed for ECQUIP as a criterion measure of NCO typical performance. At this point data is unavailable on these scales.

Validity Evidence: Project A data gave consistent evidence that training performance ratings predict first and second tour of duty performance; and that first tour performance (LVI) ratings also predicted second tour performance (LVII) adding incremental variance above prediction added by aptitude tests (ASVAB).

11. Proposed: Training Role Play

Construct Measured:

The proposed training role play will be designed to tap attributes gathered from the Special Forces job analysis. The role play will be designed to elicit some general attributes such as planning and creativity; communication attributes including communication, and non-verbal communication; and several interpersonal skills including motivating others.

Short Description of Test:

Role plays are developed based on job analysis information identifying those attributes which are readily assessed by an interactive method. A role play to tap teaching skills of SF candidates is useful because the focus of many special forces missions revolves around these skills.

The role play will be structured such that the SF candidate will be provided a list of about 20 alternate tasks and asked to select one to teach (e.g., dress a wound, assesmbel/disassemble a weapon, knot tying). The SF candidate will be given a kit to use in his "course" and two days to prepare a one hour training session. The soldier will be required to conduct training to a group of about three to five peers. The soldier is responsible for developing training content as well as structuring the training session in terms of hands-on training versus lecture style, pulling together training materials and teaching aids, and planning and outlining training and final presentation. Manuals and instructional materials will be provided in the kit.

Psychometrics:

Scoring: Rating scales or a behavior checklist will be developed to rate role play performance. Peers and cadre who participate or watch the training session will rate training performance. In addition, participants will be given a skills test to determine if they learned the topic matter that was to be covered in training.

12. Proposed: Cultural Adaptability Role Play

Construct Measured:

The proposed role play, Interacting with Different Cultures will be designed to tap most communication attributes particularly language ability for dealing with a language barrier; general communication ability and non-verbal communication will be important; interpersonal skills particularly related to adaptability and diplomacy will be important and a level of perceptiveness and interest in the other culture are critical to rapport building.

Short Description of Test:

A role play to tap into the attributes necessary to interact with other cultures is useful because many Special Forces missions revolve around these skills. The role play will be developed with parallel "forms" so that new troops can not be debriefed regarding the role play content and demands prior to testing.

The role play will be designed as a scenario of a traditional cultural ceremony (perhaps a formal feast/meal or a wedding) in which the soldier must participate in order to maintain and build good relations and rapport with the host nation. This role play will necessitate the cooperation of native-born Spanish speaking soldiers, or those soldiers thoroughly familiar and knowledgeable of "Latino" culture (as one example). The subject will be required to interact and behave in such a way as is customary with the host culture. Some of the cultural traditions will be evident and "spoken" traditions; other traditions will be "unspoken." To interact successfully the subject will need to pick up on these unspoken behaviors and follow suit so as not to offend the host. Some of the customs may involve eating food typically found repulsive in our country (such as dog meat); while other customs might involve how food is eaten (with the appropriate hand), the way one sits at the table, and customs such as the correct way to touch one another (that Americans might feel uncomfortable or unfamiliar with).

Psychometrics:

Scoring: Rating scales or a behavior checklist will be developed to rate role play performance. Peers and cadre who participate in the role play will rate performance according to the constructs of interest.

13. Proposed: Structured Interview

Construct Measured:

Structured Interview questions can be developed to tap a range of cognitive and non-cognitive skills and abilities. This structured interview will be designed to elicit behaviors relevant to the attributes required for the SF job; some of the attributes that will be tapped include communication skills, both verbal and non-verbal; interpersonal skills; previous Army accomplishments and past performance, including experiences relevant to teaching others, interacting with others, and personal discipline.

Short Description of Test:

Structured questions are designed to get at individual's behaviors in situations similar to those likely to be encountered on a job (Motowidlo, Russell, Carter, & Dunnette, 1988; Pulakos, Schmitt, & Keenan, in preparation). Such interviews elicit short vignettes explaining how the applicant handled a certain kind of situations. Interviewers use behaviorally anchored rating forms to assess applicants. Questions will be developed based on the job analysis which identified the critical attributes to the position. The interview will consist of about ten questions and will last about one-half hour.

Psychometrics:

Scoring: Behaviorally anchored rating scales are often developed to help raters evaluate responses to the questions. There will be at least two trained interviewers making these ratings.

Correlations with other constructs: In research by Campion and colleagues interviews and cognitive measures have been found to have corrected correlations of about $r=.75$. However, Pulakos et al., 1994 found that cognitive ability and the experience-based interviews were not significantly correlated ($r=.09$).

Subgroup Differences: Pulakos et al., (1994) report fairly small differences between subgroups for performance on the interview or performance ratings. The male/female difference on the interview was -.05 with females performing better. The white/black difference on the interview was .12 and white/Hispanic difference was .22 with whites having higher scores.

Reliability: Pulakos et al., (1994) found intra-class correlations for reliability ranged from .74-.86 ($n=108$) for experience based interviews.

Validity Evidence: McDaniel et al., (1994) did a meta-analysis looking at the validities of over 14 studies, the validity of the situational interviews was .27 (Corrected for unreliability in the criterion and range restriction validity was .50).

Pulakos et al. (1994) found that experience-based interviews had validities of .32 ($p<.05$) ($Ns=108$) and .38 ($p<.05$; $N=464$; .48 when corrected for unreliability on the criterion) using performance ratings as the criterion. The experience-based interview significantly incremented the amount of performance predicted beyond that explained by the cognitive test.

14. Proposed: Low Fidelity Situational Simulation - Situational Judgements (SJ)

Construct Measured:

SJ measures tap an individuals' situational effectiveness/ineffectiveness in social functioning in interpersonal and communication areas. This includes conflict resolution, negotiation skills, interpersonal problem solving, supervisor/subordinate interaction, directing/leading teams, communication with peers, subordinates and supervisors, training subordinates, counseling, acting as a model, reasoning with soldiers, rewarding and disciplining, facilitating teamwork development and unit cohesion, motivating others, working with culture and or gender differences. In addition, situational judgment simulations are useful in assessing managerial and leadership abilities (Motowidlo, Dunnette, & Carter, 1990; Pulakos, et al., in preparation; Sternberg, Wagner, & Okagaki, in press).

Short Description of Test:

Low fidelity formats provide a verbal description of the scenario and a list of potential plans of action. There are currently two relevant situational judgment measures available: 1) The NCO Situational Judgment test developed for Project A and used as a criterion measure of NCO performance; 2) The Army Leadership Questionnaire developed in the Army's ECQUIP project. This measure has two alternative forms, but has not yet been tested.

Example of an SJ Exercise:
The directions for this test require the respondent to read the situation and mark the response that they believe is the most effective response. Next, they are to indicate which response they believe is the least effective response.

You are a squad leader. Over the past several months you have noticed that one of the other squad leaders in your platoon hasn't been conducting his CTT training correctly. Although this hasn't seemed to affect the platoon yet, it looks like the platoon's marks for CTT will go down if he continues to conduct CTT training incorrectly. What should you do?

a. Do nothing since performance hasn't yet been affected.
b. Have a squad leader meeting and tell the squad leader who has been conducting training improperly that you have noticed some problems with the way he is training his troops.
c. Tell your platoon sergeant about the problem.
d. Privately pull the squad leader aside, inform him of the problem, and offer to work with him if he doesn't know the proper CTT training procedure.

Psychometrics:

Scoring: Test questions are generally developed based on critical incidents. Alternatives are generated by incumbents and supervisors. Scores are based on subject matter experts' ratings of the best and worst alternatives.

Correlations with other constructs: In the Motowidlo study ($N=120$), aptitude test measures did not correlate with the SJ, except for GPA in major ($r=.30$, $p<.05$). However, SJ ratings did correlate significantly with interview ratings of interpersonal skills ($r=.21$), communication skills ($r=.16$) and negotiation ratings ($r=.50$).

Subgroup Differences: With the exception of one small sample (Motowidlo study) where women performed better than men, no gender or race differences have been found.

Reliability: Motowidlo, et al. (1990) reported an internal consistency estimate of .56 although they suggest that test-retest statistics might be more appropriate.

Jobs used for in the past: SJ tests have been used for law enforcement jobs, military leadership assessment, and managerial positions.

Validity Evidence: Motowidlo, et al (1990) reported validity estimates of .30 ($p<.01$) for overall effectiveness ratings for externally hired managers (N=120-140).

15. Proposed: High Fidelity Situational Simulation (HFSS)

Construct Measured:

HFSS taps an individuals' situational effectiveness/ineffectiveness in social functioning in interpersonal and communication areas. This includes conflict resolution, negotiation skills, interpersonal problem solving, supervisor/subordinate interaction, directing/leading teams, communication with peers, subordinates and supervisors, training subordinates, counseling, acting as a model, reasoning with soldiers, rewarding and disciplining, facilitating teamwork development and unit cohesion, motivating others, working with culture and or gender differences. In addition, situational judgment simulations are useful in assessing managerial and leadership abilities (Motowidlo, Dunnette, & Carter, 1990; Pulakos, et al., in preparation; Sternberg, Wagner, & Okagaki, in press).

Short Description of Test:

This proposed high fidelity situational simulation will present a social problem scenario in an audiovisual medium. This provides visual and auditory information and cues, requiring subjects to interpret subtle cues in these mediums. Benefits over paper-and-pencil SJs include a minimization of method variance in tests of reading/writing skills highly correlated with g; and the ability to capitalize on 'richness' and 'subtleness' of sociobehavioral information.

A measure that is currently being developed by ARI (Busciglio) presents the subject with a 2 to 3 minute problem scenario and is asked 2 written questions. 4 alternative responses appear in video format, about 20/30 seconds in length. The total vignette averages 4-6 minutes. The vignettes are developed based on critical incident information regarding difficult interpersonal and social problems.

Psychometrics:

Scoring: Test questions are generally developed based on critical incidents. Alternatives are generated by incumbents and supervisors. Scores are based on subject matter experts' ratings of the best and worst alternatives.

Correlations with other constructs: There has been considerable work in the social intelligence (SI) domain. In the past measures of SI have had been difficult to separate from a general "g" factor. Thus, the present recommendation for audio-visual medium for testing is proposed to increment the g factor getting at a unique component of SI.

Fakability: All alternatives to the questions are developed to be reasonable answers reducing faking problems.

Validity Evidence: Video Based Assessments developed by the ESS Corporation (Frank, 1993) report validities ranging from .38 to mid .50s with criteria such as turnover and performance data (no studies are referenced, and N sizes are not reported).
Pine (1994) developed a video-based Situational Response Test for correctional officers. The situations are presented in video and the alternatives are presented in a written format. Correlations with overall effectiveness ratings ($r=.26$ $p<.01$).

16. Experimental: Project A NCO Role Plays

Construct Measured:

Role plays can be used to predict performance through exercises that simulate the job environment. The basic skills tapped by role plays are interpersonal skills and oral communication. Beyond these two skills, role plays can be developed to tap many other constructs including counseling and training subordinates (Pulakos, Schmitt & Keenan, 1994, Tech. Report).

Short Description of Test:

Role plays developed in Project A to measure NCO performance could be administered to SF candidates. In Project A the following contexts provided the role play content:

(1) Counseling a subordinate with personal problems; in this scenario a PFC is having financial difficulties caused by his wife's excessive spending, he is also having a difficult time being separated from his new wife. He is not happy in Korea and further, is having trouble getting reaching his wife in the States.

(2) Counseling a subordinate with performance problems; in this scenario a PFC has missed work and upon checking on his reason he is caught in a lie. He has lied in the past and has been late to work several times although his work is generally up to standard.

(3) Remedial training with a subordinate; a PVT is having difficulty with drill and ceremony activities. He is an enthusiastic soldier yet needs some extra training in these areas. The PVT must go through additional training to perform the Hand Salute and an About Face.

Psychometrics:

Scoring: Typically there will be a confederate who is assigned to play a role in the role play exercise. The confederate also has full or part responsibility for making ratings of the examinee's performance. The role play score is based on a behaviorally based rating scales (or checklists). In Project A a 5-pt rating scale was used to rate specific behaviors.

Correlation with other constructs: In Project A analyses, the role play variables all loaded on a general Leadership factor (LEAD) along with the Administrative Index for promotion rate, Army-wide BARS for Leading/Supervisory rating score, and the Situational Judgement Test. The LEAD variable was related to other measures of performance taken at the end of training and the first tour.

Internal Consistency Reliability: Project A data showed the following reliabilities:

	Type of scenario (role play)		
	Personal Counseling	Disciplinary Counsel	Training
Median 1-rater rel.	.68	.78	.68
Overall effectiveness (1-rater)	.84	.76	.89

Validity Evidence: Often, role play simulations are defended based on content validity bases. Schmitt, Gooding, Noe, and Kirsch (1984) reported average validities for a wide variety of work sample tests in the range of .38 (18 validity studies).

Appendix F
Predictor Description Bibliography

Baer, J. M. (1988). Long-term effects of creativity training with middle school students. Journal of Early Adolescence, 8, 183-193.

Bernardin, H. J. (1987). Development and validation of a forced choice scale to measure job-related discomfort among customer service representatives. Academy of Management Journal, 30, 162-173.

Bernardin, H. J. (1989). Innovative approaches to personnel selection and performance appraisal. Journal of Management Systems, 1, 25-36.

Burke, W. P., & Dyer, F. N. (1984). Physical fitness predictors of success and injury in Ranger training (ARI Research Report No. 1366). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Busciglio, H. H. (1990). The incremental validity of spatial and perceptual-psychomotor tests relative to the Armed Services Vocational Aptitude Test Battery (ARI Technical Report No. 883). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A220 903)

Busciglio, H. H. (1994). Defining and measuring social intelligence: Preliminary steps (RACO-Draft Report). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

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., Silva, J. M., & Walker (1990). The potential of new Army test to improve performance. Paper presented at the 1990 Army Science Conference, Alexandria, VA.

Busciglio, H. H., & Teplitzky, M. L. (1990). Project A spatial tests and military orienteering performance in the Special Forces assessment and selection program (ARI Technical Report). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A233 432)

Busciglio, H. H., & Teplitzky, M. L. (1994). Predicting land navigation performance in the special forces qualification course (ARI Technical Report). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Campbell, J. P., & Zook, L. M. (1990). Building and retaining the career force: New procedures for accessing and assigning Army enlisted personnel, FY 1990 Annual Report. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Campbell, J. P., & Zook. L. M. (1991). Building and retaining the career force: New procedures for accessing and assigning Army enlisted personnel, FY 1991 Annual Report. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Campbell, J. P., & Zook. L. M. (1992). Building and retaining the career force: New procedures for accessing and assigning Army enlisted personnel, FY 1992 Annual Report. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Campbell, J. P., & Zook. L. M. (1993). Building and retaining the career force: New procedures for accessing and assigning Army enlisted personnel, FY 1993 Annual Report. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

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.

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.

CTB/McGraw-Hill. (1987). Tests of Adult Basic Education: Forms 5 and 6 Technical Report. Monterey, CA: CTB/McGraw-Hill.

CTB/McGraw-Hill. (1987). Tests of Adult Basic Education: Forms 5 and 6 Norms Book. Monterey, CA: CTB/McGraw-Hill.

DeMatteo, J. S., White, L. A., Teplitzky, M. L., & Sachs, S. A. (1991). Relationship between temperament constructs and selection for Special Forces training. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Department of Defense. (1984). Armed Services Vocation Aptitude Battery: Test Manual. Chicago, IL: United States Military Entrance Processing Command.

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 Technical Report No. 739). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

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., Roberts, M. M., & Friedman, M. P. (1958). A factor analysis of aptitude and proficiency measures in radiotelegraphy. Journal of Applied Psychology, 42(2), 129-135.

Fleishman, E. A., Mumford, M. D., Zaccaro, S. J., Levin, K. Y., & Hein, M. B. (1991). Taxonomic efforts in the description of leader behavior: A synthesis and functional interpretation. Leadership Quarterly, 2, 245-287.

Frank. (1993). Computer and multimedia testing for new skills and abilities: Practical issues. In Video based Assessment. Symposium conducted at the Society of Industrial and Organizational Psychologist Convention, San Francisco, CA.

Heslegrave, R. J., & Colvin, C. (1993). Selection of personnel for stressful occupations: The potential utility of psychophysiological measures as selection tools (ARI Technical Report No. 975). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Heslegrave, R. J., & Colvin, C. (1994). An exploration of psychological and psychophysiological measures as predictors of successful performance under stress (ARI Report). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Hough, L. M., Keyes, M. A., & Dunnette, M. D. (1983). An evaluation of three "alternative" selection procedures. Personnel Psychology, 36, 261-276.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternate predictors of job performance. Psychological Bulletin, 96, 72-98.

Ilgen, D. R. (1994). Jobs and roles: Accepting and coping with the changing structure of organizations. In M. G. Rumsey, C. B. Walker, & J. H. Harris (Eds.), Personnel selection and classification (pp. 13-32). Hillsdale, NJ: Erlbaum.

Karlsen, B., & Gardner, E. F. (1986). Adult Basic Learning Examination, Norms Booklet Level 3 (2nd ed.). USA: Harcourt Brace Jovanovich.

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.

Lohman, D. R. (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: Erlbaum.

Mael, F. A., & Alderks, C. E. (1993). Leadership team cohesion and subordinate work unit morale and performance. Military Psychology, 5(3), 141-158.

Mael, F. A., & Ashforth, B. E. (1992). Alumni and their alma mater: A partial test of the reformulated model or organizational identification. Journal of Organizational Behavior, 13, 103-123.

Mael, F. A., & Ashforth, B. E. (1994). Loyal from day one: Biodata, organizational identification, and turnover among newcomers. Paper presented at the annual meeting of the Academy of Management, Dallas, TX.

Mael, F. A., & Hirsch, A. C. (1993). Rainforest empiricism and quasi-rationality: Two approaches to objective biodata. Personnel Psychology, 46, 719-738.

Mael, F. A., & Tetrick, L. E. (1992). Identifying organizational identification. Educational and Psychological Measurement, 52, 813-824.

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

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.

McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. Journal of Applied Psychology, 79(4), 599-616.

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., & 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.

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 Technical Report No. 739). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

McKelvie, S. J. (1989). The Wonderlic Personnel Test: Reliability and validity in an academic setting. Psychological Reports, 65, 161-162.

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.

Motowidlo, S. J., Russell, T. L., Carter, G. W., & Dunnette, M. D. (1988). Revision of the Management Selection Interview (Institute Report #156). Minneapolis, MN: Personnel Decisions Research Institute.

Mumford, M. D., Baughman, W. A., Supinski, E. P., Costanza, D. P., & Threlfall, D. V. (1993). Cognitive and metacognitive skill development: Alternative measures for predicting leadership potential (Report No. SBIR A92-154). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

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

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., 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.

Pine, D. E. (1994). The development and validation of a video-based situational response test. Paper presented at the Society of Industrial Organizational Psychologist Convention, Nashville, TN.

Pleban, R. J., Allentoff, H. L., & Thompson, T. J. (1989). Preliminary assessment of selected predictors of special forces qualification course success (ARI Research Report 1539). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Pulakos, E., Schmitt, N., & Keenana, P. (in preparation). Validity and implications of the FBI special agent entry level selection system (ARI Technical Report 94-20). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Ree, M. J., & Earles, J. A. (1991). Predicting training success: Not much more than *g*. Personnel Psychology, 44, 321-332.

Ree, M. J., & Earles, J. A. (1992). 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., 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: Human Resources Directorate, Armstrong Laboratory.

Riegelhaupt, B. J., Harris, C. D., & Sadacca, R. (1987). The development of administrative measures as indicators of soldier effectiveness (ARI Technical Report No. 754). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Russell, T. L., Crafts, J. L., Tagliareni, F. A., McCloy, R. A., & Barkley, P. (1994). Job analysis and Special Forces jobs (ARI Contract No. MDA903-92-D-0206, Del. Order 1). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Russell, T. L., Humphreys, L., Rosse, R., & Peterson, N. G. (1992). The factor structure of a spatial test battery. Paper presented at the meeting of the Military Testing Association, San Diego, CA.

Russell, R. L., Reynolds, D. H., & Campbell, J. P. (1994). Building a joint-service classification research roadmap: Individual differences measurement (AL/HR-TP-1994-0009). Brooks Air Force Base, TX: Armstrong Laboratory.

Schmitt, N., Gooding, R. Z., Noe, R. A., & Kirsch, M. P. (1994). Meta-analyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. Personnel Psychology, 37, 407-422.

Shorris, E. (1981). The oppressed middle: Politics of middle-management/scenes from corporate life. New York: Anchor Press/Doubleday.

Silva, J. (1989). Usefulness of spatial and psychomotor testing for predicting TOW and UCOFT gunnery performance (ARI Working Paper WP-RS-89-21). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Silva, J. M. (1994). Examining the self-development test for race and gender fairness (ARI Project No. 2O465803D730, Technical Report). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Silva, J. M., & White, L. A. (1993). Relation of cognitive aptitudes to success in foreign language training. Military Psychology, 5(2), 79-93.

Silva, J. M., White, L. A., & Rumsey, M. G. (1991). Relationship of cognitive aptitudes to success in foreign language training. Paper presented at the meeting of the American Psychological Association, San Francisco, CA.

Sternberg, R. J., Wagner, R. K., & Okagaki, L. (in press). Practical intelligence: The nature and role of tacit knowledge in work and at school. In H. Reese & J. Puckett (Eds.), Advances in lifespan development. Hillsdale, NJ: Erlbaum.

Teplitzky, M. L. (1990). Spatial abilities and land navigation performance (ARI Research Report). 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.

Villanova, P., & Bernardin, H. J. (1990). Work behavior correlates of interviewer job compatibility. Journal of Business and Psychology, 5, 179-195.

Villanova, P., Bernardin, H. J., Johnson, D. L., & Dahmus, S. A. (1994). The validity of a measure of job compatibility in the prediction of job performance and turnover of motion picture theater personnel. Personnel Psychology, 47, 73-90.

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

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.

White, L. A., Nord, R. D., Mael, F. A., & Young, M. C. (1993). The assessment of background and life experiences. In T. Trent & J. H. Laurence (Eds.), Adaptability screening for the Armed Services (pp. 101-162). Washington, DC: Office of Assistant Secretary of Defense (Force Management and Personnel).

Wonderlic Personnel Test, Inc. (1992). Wonderlic Personnel Test & Scholastic Level Exam. Libertyville, IL: Wonderlic Personnel Test, Inc.

Zazanis, M. M., Zaccaro, S. J., Diana, M., Teplitzky, M. L., & Gilbert, J. A. (1994). The construct validity of a measure of social intelligence. Paper presented at the meeting of the American Psychological Society, Washington, DC/

Appendix G
Predictor Expert Judgment Exercise Instructions

MEMORANDUM

To:
From: Teresa Russell, Rita Nee, Michelle Rohrback, and Norm Peterson
Re: Expert Judgment Exercise

Thank you for agreeing to participate in the expert judgment exercise. The ratings you make, along with those made by the other judges, will be used to identify measures for selecting Army personnel into Special Forces (SF) jobs. Applicants for SF jobs are typically conventional Army NCOs (usually E-5s). Previous job analysis research has identified the attributes (e.g., creativity, adaptability) necessary for successful performance in SF; this portion of the project seeks to lay the foundation for a comprehensive validation of new and existing predictors of performance for Special Forces.

The purpose of this expert judgement exercise is to evaluate the usefulness of a number of potential predictors for measuring the attributes defined in the SF job analysis. Here predictors have been broadly defined to include a host of different types of measures such as role play exercises, interviews, peer and supervisory ratings of performance, administrative indices, personality and interest inventories, computerized measures, and of course traditional cognitive paper-and-pencil tests.

Based on the ratings you made of your familiarity with different domains and on our needs for expertise, we have enclosed the following exercises for you to complete:

[insert here: name of exercise (e.g., cognitive) and an estimate of the amount of the exercise will take]

Please give your ratings to Jennifer Crafts by FRIDAY, SEPTEMBER 30th, 1994. Please consult with Jennifer on billing information.

Thanks again for your time.

MEMORANDUM

To:
From: Teresa Russell, Rita Nee, Michelle Rohrback, and Norm Peterson
Re: Expert Judgment Exercise

Thank you for agreeing to participate in the expert judgment exercise. The ratings you make, along with those made by the other judges, will be used to identify measures for selecting Army personnel into Special Forces (SF) jobs. Applicants for SF jobs are typically conventional Army NCOs (usually E-5s). Previous job analysis research has identified the attributes (e.g., creativity, adaptability) necessary for successful performance in SF; this portion of the project seeks to lay the foundation for a comprehensive validation of new and existing predictors of performance for Special Forces.

The purpose of this expert judgement exercise is to evaluate the usefulness of a number of potential predictors for measuring the attributes defined in the SF job analysis. Here predictors have been broadly defined to include a host of different types of measures such as role play exercises, interviews, peer and supervisory ratings of performance, administrative indices, personality and interest inventories, computerized measures, and of course traditional cognitive paper-and-pencil tests.

Based on the ratings you made of your familiarity with different domains and on our needs for expertise, we have enclosed the following exercises for you to complete:

[insert here: name of exercise (e.g., cognitive) and an estimate of the amount of the exercise will take]

Please give your ratings to Rita Nee by FRIDAY, SEPTEMBER 30th, 1994. Your time should be billed the HumRRO SFROAD.

Thanks again for your time.

INSTRUCTIONS

Background Information Form

Please begin by completing the background information form enclosed in the packet. It will help us document the expertise of participants.

Then read the executive summary of the job analysis before proceeding (it's only two pages long).

Materials

Enclosed you will find four important materials:

- *Predictor Descriptions*-- You will find one set of predictor descriptions for each exercise (e.g., cognitive, non-cognitive) you are completing.

  There are four major kinds of predictor measures:

  (1) Proposed-- instruments that we are considering for development for Special Forces.
  (2) Experimental-- instruments that have been developed and field tested but are not currently in use.
  (3) Operational-- instruments that are currently in use.
  (4) Published-- instruments developed and controlled by a test publisher.

  The predictor description includes a synopsis of research on each predictor. The amount of information available for the predictors varies greatly depending upon the type of predictor, extensiveness of its research history, and the availability of documentation about the predictor.

- *Attribute Definitions*-- You will find one set of attribute definitions that are generally applicable to all the exercises.

  The attribute definitions were developed through a job analysis of SF jobs, coupled with a literature review. There are 47 attributes tapping a wide range of individual differences characteristics. Many of the characteristics are traditional cognitive or non-cognitive constructs (e.g., mechanical ability, dependability). Because the SF applicant pool is composed of non-commissioned officers (NCO) who already have a track record in the conventional Army, some of the attributes are "performance" constructs that are targeted at conventional Army experience.

- *Predictor x Attribute Rating Form--* This is the form on which you will record your judgments about the extent to which each of the predictors measure each of the attributes. Predictors are listed in the rows of the form. Attributes are listed in the columns. Your numeric ratings will be written in the cells. You will find one predictor x attribute rating form for each expert jugdment task you are completing.

- *Best Bet Predictor Form--* After you complete the predictor x attribute judgments, we would like you to indicate which predictor is your "Best Bet" for measuring the attribute. Record your Best Bets on this form.

Please ensure that all of these materials are present. If you are missing any forms, call Rita Nee at (703) 706-5663 to obtain them.

Specific Instructions

We are interested in your estimate of how well the predictors measure each attribute. Please follow these steps to make your ratings:

1. Scan the entire set of *Predictor Descriptions* to get a feel for the type and level of information available about each predictor.

2. Read the entire list of *Attribute Definitions* carefully.

3. Now carefully read the information about the first predictor. Consider the first attribute. To what extent does the first predictor measure the first attribute? Use this rating scale to quantify your judgment:

0----------1----------2----------3----------4----------5----------6----------7----------8

This attribute is not at all measurable by the predictor (it is almost useless)	This attribute is measured partly by the predictor (it is of some use)	This attribute is entirely measured by the predictor (it is highly useful)

Factors to Consider in Making Your Extent-of-Measurement Judgments

What does the predictor measure? The description of the test and information about the construct validity (i.e., correlations with other constructs) are intended

to help you better understand what the test measures.

How well does the predictor measure the construct? Internal consistency reliability and test retest reliability are intended to help you understand how much measurement error is associated with each predictor. Large practice effects can also attenuate the extent of measurement and should result in lower test-retest reliabilities. On self-report measures, such as personality, faking (or responding in a socially desireable way) may also reduce the predictor's usefulness in measuring a construct.

What is the predictor's track record? Where available, validity evidence is provided to help you better understand what the test measures.

What if there isn't much information about a predictor? The amount of information available for the predictors varies greatly depending upon the type of predictor, extensiveness of its research history, and the availability of documentation about the predictor. Your job as an expert rater will be to make the best judgment you can about each predictor, given the amount of information available for the predictor and your expertise with different measurement methods.

4. Record your judgment on the *Predictor x Attribute Rating Form* by writing in the number from the scale that best represents your judgment (e.g., "6" or "7").

5. Go to the next attribute and judge the extent to which the first predictor measures it. This measn that you will work your way across the first row on the *Predictor x Attribute Rating Form*. Continue until you have rated the first predictor for all of 47 attributes. Please note the 47 attributes are continued on multiple pages.

6. Move to the second predictor measure and repeat steps 1-5. Continue this process until you have rated all of the predictors for all of the attributes.

7. After you have completed all of your ratings, please indicate which predictor (if you could only use one) you would use to measure each attribute. Write the number and name of the your "Best Bet" predictor next to the name of the attribute on the *Best Bet Predictor Form.* If none of the predictors were adequate measures of the attribute, write "none" in the blank next to the name of the attribute.

   To make this judgment, you should consider your "extent of measurement" ratings and take into account any other factors you consider relevant (e.g., the extent of subgroup differences on the predictor).

8. When you have completed your ratings, please return them, by September 30, to

Rita Nee at HumRRO. Those of you who work at ARI should mail your packets. AIR personnel should give packets to Jennifer Crafts.

Job Analysis of Special Forces Jobs Executive Summary

Research Requirement:

The overall goal of the project was to gather information that will aid in the development of new Special Forces (SF) performance measures. This goal required two types of information: (a) the individual attributes requisite to SF performance and (b) the behavioral dimensions of field performance of SF jobs. The research involved five major steps:

(1) Development of workshop materials and logistics,
(2) Administration of workshops to collect critical incidents and task and attribute ratings,
(3) Analysis of critical incident, task, and attribute data,
(4) Development of performance categories and behavior based rating scales, and
(5) Analysis of linkages between attributes and performance categories.

Procedure:

Step 1, Development of workshop materials and logistics, involved: (1) collecting and reviewing documents to form initial lists of job tasks and personal attributes relevant to SF Military Occupational Specialties (MOS), (2) conducting interviews with SF officers and NCOs to obtain critical incidents and feedback on the initial lists of tasks and attributes, and (3) preparing and pilot testing job analysis data collection procedures.

Steps 2 and 3 were accomplished over the course of May through July of 1993. In total, 175 NCOs, officers, and warrant officers participated. They represented various SF MOS and the five major SF groups (i.e., Special Forces Group Airborne [SFGA]). On average, the participants had 13 years of Army experience and 8 years of SF experience. Seventy-seven percent of participants were currently assigned to A Detachments (B Detachment = 17%, C Detachment = 6%).

The participants in Step 2 provided three major types of information:

(1) judgments about individual attributes (such as judgment and decision making ability, non-verbal communication ability, endurance, motivation)
(2) judgments about task areas relevant to SF MOS, and
(3) descriptions of critical incidents (scenarios that describe a situation, an SF individual's behavior in that situation, and the outcome of the individual's actions).

Step 3, data analysis, involved: (1) editing and categorizing critical incidents, (2) computing means, standard deviations, and reliability coefficients for the task ratings, and (3) computing means, standard deviations, and reliability ratings for the attribute ratings.

In total, the participants provided 1,767 critical incidents, and in turn, the research team organized the incidents into job performance categories. Step 4 involved collecting and analyzing additional information on the performance categories and critical incidents. It had two goals: (1) to get input from SF NCOs, officers, and warrant officers on the performance categories and (2) to obtain judgments about the effectiveness of different behaviors that are represented in the critical incidents. One hundred and thirteen SF NCOs, officers, and warrant officers representing the five SFG[A] made the judgments. In turn, we used the effectiveness data to develop behavior-based performance evaluation scales relevant to each of the performance categories.

Step 5, Analysis of linkages between attributes and performance categories, involved collecting judgments from NCOs, officers, and researchers familiar with SF jobs about the importance of each attribute for effective performance in each of the job performance categories.

Findings:

The critical incident technique yielded 26 performance dimensions that describe SF jobs. These behavioral dimensions are diverse such as "Building Effective Relationships with Indigenous Populations," "Decision-Making," and "Navigating in the Field."

A wide variety of attributes (e.g., physical endurance, reasoning ability, language adeptness) are needed for effective performance in the 26 performance areas. Forty seven relevant attributes were defined.

Utilization of Findings:

The information developed in this project formed the foundation for the identification and validation of tests or other tools likely to predict performance in SF jobs. The behavior-based rating scales may be used to gather criterion data. Task ratings will guide development of hands-on or job knowledge performance criteria. Definitions of job-relevant individual attributes will guide identification of appropriate predictors for SF job performance.

ACRONYMS AND TERMS OF INTEREST

Effect size- the difference in standard deviation unit between two means.

With regard to subgroup differences the formal is: ($mean_{subgroup\ 1}$ - $mean_{subgroup\ 2}$)/ total group standard deviation. An effect size of 1.00 would signify that the mean for subgroup 1 was one standard deviation higher than the mean for subgroup 2.

With regard to practice effects the formula is: ($mean_{Time\ 2}$ - $mean_{Time\ 1}$)/ average standard deviation across times 1 and 2. An effect size of 1.00 would signify that the mean for the Time 2 administration of the measure was one standard deviation higher than the mean for the Time 1 administration.

Project A- A concurrent validity study sponsored by the Army Research Institute to identify predictors that could supplement the Armed Services Vocational Aptitude Battery (ASVAB). Its sister project, Building the Career Forces, involved longitudinal validation of the Project A measures.

Q-Course- Special Forces Qualification Course, also called SFQC. Includes MOS specific training and training in small group tactics. Individuals attend SFQC if they complete SFAS successfully.

SF- Special Forces

SFAS- Special Forces Assessment and Selection - a 3 week assessment program used to select soldiers for the Special Forces Qualification Course. SFAS is composed of military orienteering exercises, physical strength and endurance exercises, and peer and observer assessments on field exercises, as well as paper-and-pencil testing.

SFQC- Special Forces Qualification Course, also called the Q-Course. Includes MOS specific training and training in small unit tactics. Individuals attend SFQC if they complete SFAS successfully.

BACKGROUND INFORMATION FORM

Name: _____

Please circle "yes" or "no" to describe your background in each area.

Development/ Design of **Cognitive** Tests	-Heard about this task in undergraduate courses or general sources	Yes	No
	-Studied this task in graduate courses (or in depth on your own)	Yes	No
	-Performed parts of this task under supervision	Yes	No
	-Supervised others performing this task	Yes	No
	-Taught others this task	Yes	No
	-Wrote a scholarly article/book about this task	Yes	No
Development/ Design of **Physical/ Psychomotor** tests	-Heard about this task in undergraduate courses or general sources	Yes	No
	-Studied this task in graduate courses (or in depth on your own)	Yes	No
	-Performed parts of this task under supervision	Yes	No
	-Supervised others performing this task	Yes	No
	-Taught others this task	Yes	No
	-Wrote a scholarly article/book about this task	Yes	No

Development/ Design of **Noncognitive** Measures (e.g., personality, interest)	-Heard about this task in undergraduate courses or general sources	Yes	No
	-Studied this task in graduate courses (or in depth on your own)	Yes	No
	-Performed parts of this task under supervision	Yes	No
	-Supervised others performing this task	Yes	No
	-Taught others this task	Yes	No
	-Wrote a scholarly article/book about this task	Yes	No
Development/ Design of **Job Performance** Measures	-Heard about this task in undergraduate courses or general sources	Yes	No
	-Studied this task in graduate courses (or in depth on your own)	Yes	No
	-Performed parts of this task under supervision	Yes	No
	-Supervised others performing this task	Yes	No
	-Taught others this task	Yes	No
	-Wrote a scholarly article/book about this task	Yes	No

Final Attribute Definitions

General Attributes

1. **Judgment and Reasoning** - to make sound decisions; using common sense; improvising; extracting general principles and applying them in new situations.

2. **Planning** - to plan and organize activities and resources such that mission objectives are met.

3. **Adaptability** - to switch gears; modifying plans to fit the situation.

4. **Creativity** - to find novel ways to use the resources at hand in solving problems.

5. **Auditory Ability** - to detect, memorize, retain, and distinguish tonal patterns or sounds.

6. **Mechanical Ability** - to understand electrical and mechanical principles; to understand how equipment works.

7. **Spatial Ability** - to readily orient oneself in an unfamiliar environment; reading maps or diagrams; forming mental pictures of things (e.g., equipment, terrain).

8. **Perceptual Ability** - to notice details of the physical environment; to be attentive to and observant of surroundings.

9. **Basic Math** - to add, subtract, multiply, divide, and use formulas.

10. **Advanced Math** - to use advanced math such as geometry or algebra.

Communication Attributes

11. **Reading Ability** - to read and comprehend written materials.

12. **Writing Ability**- to write materials that are easily understood; using appropriate grammar, punctuation, and level (for the audience).

13. **Language Ability** - to be multi-lingual; learning new languages.

14. **Communication Ability** - to present information clearly; using voice inflection and eye contact for emphasis; tailoring presentations to the audience.

15. **Non-Verbal Communication** - to use and read non-verbal behaviors (e.g., posture, gestures) accurately.

Interpersonal Skills, Motivation, and Character

16. **Persuasiveness/Diplomacy** - to be tactful, pleasant, and diplomatic toward others; to be persuasive.

17. **Cultural/Interpersonal Adaptability** - to modify own style and behavior to fit the situation and culture; being tolerant of other cultures and value systems.

18. **Maturity** - to be level-headed and emotionally stable; to remain calm under stress.

19. **Autonomy** - to be self-confident, self-sufficient, and comfortable when working alone.

20. **Team Playership** - to be cooperative--to support the team effort, making contributions to the team.

21. **Dependability** - to be responsible and loyal; following through on duties.

22. **Initiative** - to be self-motivated, self-starting, and achievement-oriented.

23. **Perseverance** - to sustain a high level of effort over long periods of time, in spite of hardships.

24. **Moral Courage** - to act on own convictions, despite consequences; choosing the more difficult "right" over the easier "wrong."

25. **Motivating Others** - to encourage team work and maintain esprit d'corps; setting an example for others.

26. **Supervising** - to organize and monitor the work of others.

Physical and Psychomotor Attributes

27. **Swimming** - to swim capably; using water survival skills; avoiding water hazards.

28. **Physical Flexibility and Balance** - to kneel, stoop, reach, or get into awkward physical positions, maintaining balance.

29. **Physical Strength** - to push, pull, lift, or carry heavy objects.

30. **Physical Endurance** - to do cardiovascular activities, such as running, skiing, climbing;

achieving and maintaining a high level of physical readiness.

31. **Psychomotor Ability** - to have good eye-hand coordination and quick reaction time.

Interests

32. **Interest in Adventure and Outdoor Activities** - to like adventurous activities such as riding motorcycles or parachuting; to like hunting, fishing, and camping.

33. **Interest in Skilled Trades** - to like auto mechanics, carpentry, or other skilled types of work.

34. **Interest in Other Cultures** - to like learning about other cultures.

35. **Interest in People** - to like people, enjoying being around people.

36. **Enterprising Interests** - to like activities that involve leading others or being persuasive or assertive.

Conventional Army Experiences

37. **Leadership** - to use good judgment in dealing with subordinates (e.g., counseling, disciplining); acting as a role model, communicating, and supervising effectively.

38. **Achievement and Effort** - to produce high quality work, exhibiting effort and initiative; to achieve notable accomplishments.

39. **Personal Discipline** - to follow regulations/orders; to exhibit integrity and self-control.

40. **Physical Fitness and Military Bearing** - to maintain physical fitness, strength, and stamina; to maintain proper military appearance and bearing.

41. **General Soldiering Proficiency** - to perform basic soldiering tasks (e.g., first aid, land navigation, NBC activities, field techniques, weapons, communications, mines) effectively.

42. **Infantry (11 CMF) Core Technical Proficiency** - to perform infantryman tasks proficiently.

43. **Combat Engineer (12 CMF) Technical Proficiency** - to perform combat engineering tasks proficiently.

44. **Other Combat MOS Technical Proficiency** - to be proficient in combat MOS other than 11 or 12 CMF (e.g., 13B, 16S, 19E).

45. **Radio Teletype Operator (31 CMF) Technical Proficiency** - to perform radio teletype operator tasks proficiently.

46. **Medical Care Specialist (91 CMF) Technical Proficiency** - to perform medical care specialist tasks proficiently.

47. **Other Non-Combat MOS Technical Proficiency** - to be proficient in non-combat MOS other than 31 or 91 CMF (e.g., 63B, 64C, 71L, 95B).

Appendix H
Predictor Measures Rated Most Highly for Each Attribute

Tests and Scales that are Likely to be Good Measures of SF Attributes

SF Attributes	Measures	Mean Extent of Measurement (0 to 8)
1. **Judgment and Reasoning** - to make sound decisions; using common sense; improvising; extracting general principles and applying them in new situations.	C11. Wonderlic	5.92
	C31. Problem Solving Skills	5.33
	C4. ASVAB PC	5.25
	C32. Solution Characteristics	5.17
	C28. Category Search and Specification	5.08
	C36. Wisdom 2	5.00
	C34. Planning and Implementation	4.92
	C27. Information Encoding	4.83
	C29. Category Combination	4.83
	C37. Leadership Problems Inventory	4.83
	C33. Problem Evaluation	4.83
	C2. ASVAB AR	4.83
2. **Planning** - to plan and organize activities and resources such that mission objectives are met.	C34. Planning and Implementation	5.50
	C37. Leadership Problems Inventory	4.92
	C32. Solution Characteristics	4.83
	C31. Problem Solving Skills	4.42
	P4. SFAS Situation Reaction Exercises	4.17
	C26. Problem Construction	4.08
	P11. Teaching Role Play	3.92
	C11. Wonderlic	3.83
	C33. Problem Evaluation	3.83
	C27. Information Encoding	3.83
3. **Adaptability** - to switch gears; modifying plans to fit the situation.	C26. Problem Construction	4.17
	C32. Solution Characteristics	3.75
	C11. Wonderlic	3.58
	NC10. ABI Family/Community	3.55
	C31. Problem Solving Skills	3.42
	NC12. RBI Cognition Under Stress	3.36
	C34. Planning and Implementation	3.33
	C28. Category Search and Specification	3.33
	C29. Category Combination	3.25
	C27. Information Encoding	3.25
4. **Creativity** - to find novel ways to use the resources at hand in solving problems.	C29. Category Combination	4.75
	C26. Problem Construction	4.67
	C32. Solution Characteristics	4.33
	C27. Information Encoding	4.17
	C34. Planning and Implementation	3.83
	C30. Wisdom 1	3.75
	C31. Problem Solving Skills	3.75
	C11. Wonderlic	3.75
	C36. Wisdom 2	3.50
	C28. Category Search and Specification	3.50
5. **Auditory Ability** - to detect, memorize, retain, and distinguish tonal patterns or sounds.	C39. Superdit-Sound Memory	6.92
	C40. Superdit-Sound Memory with Interference	6.50
	C38. Army Radio Code Test	5.50
	C41. Superdit-Motor Programming	5.17
		3.50

Tests and Scales that are Likely to be Good Measures of SF Attributes

SF Attributes	Measures	
6. **Mechanical Ability** - to understand electrical and mechanical principles; to understand how equipment works.	C9. ASVAB MC	7.33
	C7. ASVAB AS	6.42
	NC4. ABI Mechanical Activities	5.55
	C10. ASVAB EI	5.25
	C13. Project A Assembling Objects	4.00
	C2. ASVAB AR	3.33
	C1. ASVAB GS	3.17
	NC33. AVOICE-Structural/Machines	3.09
7. **Spatial Ability** - to readily orient oneself in an unfamiliar environment; reading maps or diagrams; forming mental pictures of things (e.g., equipment, terrain).	C13. Project A Assembling Objects	7.17
	C12. Project A Map Test	6.92
	C9. ASVAB MC	5.08
	P2. SFAS Military Orienteering	4.50
	C22. Project A TID	4.25
	C7. ASVAB AS	3.75
	PS4. Cannon Shoot Test	3.60
	C10. ASVAB EI	3.50
	C8. ASVAB MK	3.42
	PS2. Project A Target Tracking 2	3.40
8. **Perceptual Ability** - to notice details of the physical environment; to be attentive to and observant of surroundings.	C21. Project A PSA	6.25
	C22. Project A TID	6.00
	C12. Project A Map Test	4.58
	C13. Project A Assembling Objects	4.50
	C9. ASVAB MC	3.58
	C6. ASVAB CS	3.50
	C24. Project A STM	3.42
	C23. Project A NM	3.42
	P2. SFAS Military Orienteering	3.33
	C7. ASVAB AS	3.33
9. **Basic Math** - to add, subtract, multiply, divide, and use formulas.	C2. ASVAB AR	7.42
	C16. TABE Mathematics Composite	7.17
	C20. ABLE Mathematics Composite	6.83
	C8. ASVAB MK	6.58
	C5. ASVAB NO	5.92
	C23. Project A NM	5.67
	C11. Wonderlic	4.83
	C9. ASVAB MC	3.83
	C10. ASVAB EI	3.75
	NC1. ABI Academic Performance	3.27
10. **Advanced Math** - to use advanced math such as geometry or algebra.	C8. ASVAB MK	7.50
	C16. TABE Mathematics Composite	6.25
	C20. ABLE Mathematics Composite	6.17
	C2. ASVAB AR	6.17
	C5. ASVAB NO	4.75
	C11. Wonderlic	4.58
	C9. ASVAB MC	3.75
	C10. ASVAB EI	3.50
	NC1. ABI Academic Performance	3.27

Tests and Scales that are Likely to be Good Measures of SF Attributes

SF Attributes	Measures	
11. **Reading Ability** - to read and comprehend written materials.	C4. ASVAB PC	7.33
	C15. TABE Reading Composite	7.25
	C18. ABLE Reading Composite	7.25
	C3. ASVAB WK	6.25
	C11. Wonderlic	5.33
	C17. ABLE Vocabulary Composite	5.08
	C14. TABE Language Composite	5.00
	C19. ABLE Language Composite	4.75
	C1. ASVAB GS	4.08
	C2. ASVAB AR	4.00
12. **Writing Ability**- to write materials that are easily understood; using appropriate grammar, punctuation, and level (for the audience).	C14. TABE Language Composite	6.33
	C19. ABLE Language Composite	6.00
	C4. ASVAB PC	4.92
	C3. ASVAB WK	4.83
	C15. TABE Reading Composite	4.58
	C18. ABLE Reading Composite	4.33
	C17. ABLE Vocabulary Composite	4.08
	C11. Wonderlic	3.67
	NC1. ABI Academic Performance	3.64
13. **Language Ability** - to be multi-lingual; learning new languages.	C25. DLAB	6.67
	P9. Language Training Record	5.58
	C3. ASVAB WK	4.25
	C4. ASVAB PC	4.17
	C19. ABLE Language Composite	3.58
	C14. TABE Language Composite	3.42
	C15. TABE Reading Composite	3.25
	C18. ABLE Reading Composite	3.17
	C11. Wonderlic	3.08
	C17. ABLE Vocabulary Composite	3.00
14. **Communication Ability** - to present information clearly; using voice inflection and eye contact for emphasis; tailoring presentations to the audience.	P11. Teaching Role Play	5.25
	P13. Structural Interview	5.08
	P16. NCO Role Plays	4.83
	P12. Cultural Adaptability Role Play	4.25
	C4. ASVAB PC	4.00
	C3. ASVAB WK	3.75
	C17. ABLE Vocabulary Composite	3.33
	C14. TABE Language Composite	3.25
	C11. Wonderlic	3.25
	P4. SFAS Situation Reaction Exercises	3.25
15. **Non-Verbal Communication** - to use and read non-verbal behaviors (e.g., posture, gestures) accurately.	P12. Cultural Adaptability Role Play	4.75
	P13. Structural Interview	4.50
	P16. NCO Role Plays	4.33
	P11. Teaching Role Play	4.08

Tests and Scales that are Likely to be Good Measures of SF Attributes

SF Attributes	Measures	
16. **Persuasiveness/Diplomacy** - to be tactful, pleasant, and diplomatic toward others; to be persuasive.	P12. Cultural Adaptability Role Play	5.00
	NC22. FCABLE-Dominance	4.55
	NC2. ABI Formal Leadership	3.64
	P16. NCO Role Plays	3.58
	NC24. FCABLE-Agreeableness	3.55
	P13. Structural Interview	3.33
	NC44. SI Biodata	3.27
17. **Cultural/Interpersonal Adaptability** - to modify own style and behavior to fit the situation and culture; being tolerant of other cultures and value systems.	NC11. ABI Cross Cultural Sensitivity	6.09
	P12. Cultural Adaptability Role Play	5.58
	NC44. SI Biodata	3.64
	NC10. ABI Family/Community	3.27
	NC24. FCABLE-Agreeableness	3.27
18. **Maturity** - to be level-headed and emotionally stable; to remain calm under stress.	NC25. FCABLE-Emotional Stability	5.45
	NC7. ABI Nondelinquency	3.91
	NC23. FCABLE-Dependability	3.73
	NC14. RBI Self Esteem	3.36
	NC12. RBI Cognition Under Stress	3.09
	NC10. ABI Family/Community	3.00
19. **Autonomy** - to be self-confident, self-sufficient, and comfortable when working alone.	NC36. JOB-Autonomy	6.00
	NC14. RBI Self Esteem	4.09
	NC6. ABI Home Economics	3.64
20. **Team Playership** - to be cooperative--to support the team effort, making contributions to the team.	NC13. RBI Mature Team Commitment	5.55
	P3. SFAS Peer Rankings	5.25
	NC8. ABI Team Sports/Group Orientation	5.09
	P4. SFAS Situation Reaction Exercises	
	NC24. FCABLE-Agreeableness	4.50
	NC20. RBI Object Belief	4.45
	NC2. ABI Formal Leadership	3.91
	NC42. Organizational Identity	3.64
		3.00

Tests and Scales that are Likely to be Good Measures of SF Attributes

SF Attributes	Measures	
21. **Dependability** - to be responsible and loyal; following through on duties.	NC23. FCABLE-Dependability	6.73
	NC7. ABI Nondelinquency	4.45
	P4. SFAS Situation Reaction Exercises	3.83
	P3. SFAS Peer Rankings	3.58
	P7. Promotion Rate	3.42
	NC13. RBI Mature Team Commitment	3.36
	P6. Number of Article 15s and Flag Actions	3.25
	C21. FCABLE-Work Orientation	3.00
22. **Initiative** - to be self-motivated, self-starting, and achievement-oriented.	NC16. RBI Need for Achievement	5.64
	P4. SFAS Situation Reaction Exercises	3.83
	C21. FCABLE-Work Orientation	3.73
	P5. Number of Awards and Certificates	3.42
	P7. Promotion Rate	3.42
	NC5. ABI Work Experience	3.18
	P13. Structural Interview	3.08
	NC2. ABI Formal Leadership	3.00
23. **Perseverance** - to sustain a high level of effort over long periods of time, in spite of hardships.	C21. FCABLE-Work Orientation	5.09
	NC16. RBI Need for Achievement	5.00
	PS6. SFAS Physical Endurance Composite	4.90
	P2. SFAS Military Orienteering	
	PS7. SFAS Physical Fitness Composite	4.17
	NC18. RBI Physical Endurance	4.00
	PS8. SFAS Swim Test	3.36
	P4. SFAS Situation Reaction Exercises	3.20
		3.17
24. **Moral Courage** - to act on own convictions, despite consequences; choosing the more difficult "right" over the easier "wrong."	NC7. ABI Nondelinquency	3.36
25. **Motivating Others** - to encourage team work and maintain esprit d'corps; setting an example for others.	P3. SFAS Peer Rankings	
	NC2. ABI Formal Leadership	4.50
	P11. Teaching Role Play	4.00
	P4. SFAS Situation Reaction Exercises	3.92
	NC22. FCABLE-Dominance	3.92
	P16. NCO Role Plays	3.64
	NC13. RBI Mature Team Commitment	3.58
		3.18

Tests and Scales that are Likely to be Good Measures of SF Attributes

SF Attributes	Measures	
26. **Supervising** - to organize and monitor the work of others.	P10. Army Wide-Performance Rating Scales	4.42
	P16. NCO Role Plays	4.33
	C37. Leadership Problems Inventory	3.75
	P3. SFAS Peer Rankings	3.58
	P4. SFAS Situation Reaction Exercises	3.58
	NC2. ABI Formal Leadership	3.45
	NC22. FCABLE-Dominance	3.27
	P15. High Fidelity Situational Simulation	3.25
	P14. Low Fidelity Situational Simulation	
	C32. Solution Characteristics	3.08
		3.00
	PS8. SFAS Swim Test	
27. **Swimming** - to swim capably; using water survival skills; avoiding water hazards.		6.80
	PS6. SFAS Physical Endurance Composite	
28. **Physical Flexibility and Balance** - to kneel, stoop, reach, or get into awkward physical positions, maintaining balance.	PS5. Army Physical Fitness Test	3.80
		3.30
	NC19. RBI Physical Strength	
29. **Physical Strength** - to push, pull, lift, or carry heavy objects.	PS6. SFAS Physical Endurance Composite	
	PS5. Army Physical Fitness Test	6.36
	PS8. SFAS Swim Test	6.20
	NC18. RBI Physical Endurance	
	PS7. SFAS Physical Fitness Composite	6.00
	NC3. ABI Ruggedness	4.40
	P2. SFAS Military Orienteering	4.36
		4.10
	PS6. SFAS Physical Endurance Composite	3.09
	NC18. RBI Physical Endurance	3.08
30. **Physical Endurance** - to do cardiovascular activities, such as running, skiing, climbing; achieving and maintaining a high level of physical readiness.	PS7. SFAS Physical Fitness Composite	
	PS8. SFAS Swim Test	7.30
	PS5. Army Physical Fitness Test	
	P2. SFAS Military Orienteering	6.55
	NC3. ABI Ruggedness	6.20
	NC19. RBI Physical Strength	5.10
		5.10
		4.58
		3.64
		3.36

Tests and Scales that are Likely to be Good Measures of SF Attributes

SF Attributes	Measures	
31. **Psychomotor Ability** - to have good eye-hand coordination and quick reaction time.	PS2. Project A Target Tracking 2	6.90
	PS1. Project A Target Tracking 1	6.70
	PS3. Project A Target Shoot Test	6.60
	PS4. Cannon Shoot Test	6.00
32. **Interest in Adventure and Outdoor Activities** - to like adventurous activities such as riding motorcycles or parachuting; to like hunting, fishing, and camping.	NC17. RBI Outdoor Orientation	6.82
	NC26. AVOICE-Rugged/Outdoors	6.64
	NC3. ABI Ruggedness	6.55
33. **Interest in Skilled Trades** - to like auto mechanics, carpentry, or other skilled types of work.	C7. ASVAB AS	5.50
	NC4. ABI Mechanical Activities	5.36
	NC29. AVOICE-Skilled Technical	4.73
	NC33. AVOICE-Structural/Machines	4.73
	C9. ASVAB MC	4.25
	C10. ASVAB EI	4.17
	NC9. ABI Work Skills	3.18
34. **Interest in Other Cultures** - to like learning about other cultures.	NC11. ABI Cross Cultural Sensitivity	5.91
	P12. Cultural Adaptability Role Play	3.92
	NC37. JCQ-SF Scale	3.36
35. **Interest in People** - to like people, enjoying being around people.	NC20. RBI Object Belief	4.00
	NC11. ABI Cross Cultural Sensitivity	4.00
	NC28. AVOICE-Interpersonal	3.91
	NC24. FCABLE-Agreeableness	3.55
	NC44. SI Biodata	3.18
	P12. Cultural Adaptability Role Play	3.17

Tests and Scales that are Likely to be Good Measures of SF Attributes

SF Attributes	Measures	
36. **Enterprising Interests** - to like activities that involve leading others or being persuasive or assertive.	NC22. FCABLE-Dominance	5.00
	NC2. ABI Formal Leadership	4.09
	NC16. RBI Need for Achievement	3.00
37. **Leadership** - to use good judgment in dealing with subordinates (e.g., counseling, disciplining); acting as a role model, communicating, and supervising effectively.	P3. SFAS Peer Rankings	5.67
	NC2. ABI Formal Leadership	5.36
	NC22. FCABLE-Dominance	5.36
	P4. SFAS Situation Reaction Exercises	4.92
	P16. NCO Role Plays	4.75
	P10. Army Wide-Performance Rating	4.67
	C36. Wisdom 2	4.58
	C37. Leadership Problems Inventory	4.50
	NC13. RBI Mature Team Commitment	4.36
	P1. Self Development Test	4.33
38. **Achievement and Effort** - to produce high quality work, exhibiting effort and initiative; to achieve notable accomplishments.	NC16. RBI Need for Achievement	5.82
	C21. FCABLE-Work Orientation	4.64
	P5. Number of Awards and Certificates	4.58
	P10. Army Wide-Performance Rating	4.33
	P7. Promotion Rate	4.00
	NC2. ABI Formal Leadership	4.00
	NC1. ABI Academic Performance	3.64
	P4. SFAS Situation Reaction Exercises	3.42
	P1. Self Development Test	3.17
	P8. Work and Training Portfolio	3.17
39. **Personal Discipline** - to follow regulations/orders; to exhibit integrity and self-control.	P10. Army Wide-Performance Rating	5.33
	NC7. ABI Nondelinquency	5.27
	P6. Number of Article 15s, Flag Actions	5.17
	P7. Promotion Rate	4.25
	NC23. FCABLE-Dependability	3.82
	C21. FCABLE-Work Orientation	3.27
	P13. Structural Interview	3.17
	NC25. FCABLE-Emotional Stability	3.09
	NC1. ABI Academic Performance	3.00
	NC5. ABI Work Experience	3.00
40. **Physical Fitness and Military Bearing** - to maintain physical fitness, strength, and stamina; to maintain proper military appearance and bearing.	PS6. SFAS Physical Endurance Composite	6.10
	PS5. Army Physical Fitness Test	
	PS7. SFAS Physical Fitness Composite	5.90
	P10. Army Wide-Performance Rating	5.70
	PS8. SFAS Swim Test	5.58
	NC3. ABI Ruggedness	4.20
	NC18. RBI Physical Endurance	3.91
	P2. SFAS Military Orienteering	3.73
	P4. SFAS Situation Reaction Exercises	3.50
	P7. Promotion Rate	3.17
		3.00

Tests and Scales that are Likely to be Good Measures of SF Attributes

SF Attributes	Measures	
41. **General Soldiering Proficiency** - to perform basic soldiering tasks (e.g., first aid, land navigation, NBC activities, field techniques, weapons, communications, mines) effectively.	P10. Army Wide-Performance Rating	4.50
	P2. SFAS Military Orienteering	4.17
	C11. Wonderlic	3.75
	P1. Self Development Test	3.75
	P8. Work and Training Portfolio	3.58
	C12. Project A Map Test	3.58
	P4. SFAS Situation Reaction Exercises	3.58
	PS1. Project A Target Tracking 1	3.40
	C16. TABE Mathematics Composite	3.33
	PS2. Project A Target Tracking 2	3.30
42. **Infantry (11 CMF) Core Technical Proficiency** - to perform infantryman tasks proficiently.	P1. Self Development Test	4.33
	C9. ASVAB MC	4.08
	NC38. JCQ-Weapons	4.00
	C22. Project A TID	4.00
	C7. ASVAB AS	3.92
	C11. Wonderlic	3.83
	P2. SFAS Military Orienteering	3.83
	C4. ASVAB PC	3.75
	C12. Project A Map Test	3.75
	C2. ASVAB AR	3.67
43. **Combat Engineer (12 CMF) Technical Proficiency** - to perform combat engineering tasks proficiently.	NC39. JCQ-Engineer	5.09
	C9. ASVAB MC	5.00
	C2. ASVAB AR	4.33
	C7. ASVAB AS	4.25
	C8. ASVAB MK	4.25
	C10. ASVAB EI	4.17
	C11. Wonderlic	4.17
	P1. Self Development Test	4.08
	C20. ABLE Mathematics Composite	4.00
	C16. TABE Mathematics Composite	3.92
44. **Other Combat MOS Technical Proficiency** - to be proficient in combat MOS other than 11 or 12 CMF (e.g., 13B, 16S, 19E).	P1. Self Development Test	4.17
	C11. Wonderlic	4.00
	C9. ASVAB MC	3.92
	C7. ASVAB AS	3.67
	C22. Project A TID	3.58
	C4. ASVAB PC	3.42
	C2. ASVAB AR	3.42
	C10. ASVAB EI	3.42
	C12. Project A Map Test	3.33
	C3. ASVAB WK	3.25

Tests and Scales that are Likely to be Good Measures of SF Attributes

SF Attributes	Measures	
45. **Radio Teletype Operator (31 CMF) Technical Proficiency** - to perform radio teletype operator tasks proficiently.	C10. ASVAB EI	5.33
	C38. Army Radio Code Test	4.83
	NC40. JCQ-Commo	4.64
	C41. Superdit-Motor Programming	4.58
	C11. Wonderlic	4.50
	C40. Superdit-Sound Memory with Interference	4.42
	C39. Superdit-Sound Memory	4.25
	P1. Self Development Test	4.00
	C21. Project A PSA	3.75
	C4. ASVAB PC	3.75
46. **Medical Care Specialist (91 CMF) Technical Proficiency** - to perform medical care specialist tasks proficiently.	NC41. JCQ-Medic	5.27
	C1. ASVAB GS	4.75
	C11. Wonderlic	4.50
	C4. ASVAB PC	4.00
	C3. ASVAB WK	3.75
	C2. ASVAB AR	3.67
	NC28. AVOICE-Interpersonal	3.64
	P1. Self Development Test	3.58
	C18. ABLE Reading Composite	3.42
	NC1. ABI Academic Performance	3.36
47. **Other Non-Combat MOS Technical Proficiency** - to be proficient in non-combat MOS other than 31 or 91 CMF (e.g., 63B, 64C, 71L, 95B).	C11. Wonderlic	4.17
	C4. ASVAB PC	3.58
	P1. Self Development Test	3.58
	C3. ASVAB WK	3.50
	C2. ASVAB AR	3.33
	C1. ASVAB GS	3.33
	C20. ABLE Mathematics Composite	3.17
	NC1. ABI Academic Performance	3.09
	C18. ABLE Reading Composite	3.08
	C16. TABE Mathematics Composite	3.00

Appendix I
Descriptions of Criterion Measures

Operational Measure 1

Measure: SFAS Graduation

Short Description of Measure: SFAS is a three week assessment program in which SF candidates participate in a series of rigorous physical exercisessuch as ruckmarches, obstacle course, and runs. Participants are deprived of sleep and put under extreme physical stress in a series of team events that require planning, teamwork, and physical endurance.

SFAS Graduation is thus a measure of completion of the SF assessment and selection program, the first major step in becoming a member of SF. This variable is recorded in the SFAS database.

Psychometrics:

Scoring: This variable has values of 0 for "NO" (did not graduate) and 1 for "YES."

Note that additional categorical variables exist that have more detail than just "Yes/No," e.g., SFAS Final Status (HISTORY) and Reason Dropped from SFAS (RESULT).

Relevance:: Relevant to the assessment/selection domain; this measure of performance in the assessment/selection program is collected before any job experience is gained

Comprehensiveness: This is a summary level measure of performance, summarizing behavior on many individual-level physical activities/tests and team-level exercises (e.g., situation-reaction and military orienteering events).

Discriminability: The dichotomous scoring does not provide as much information as a continuous distribution would; it also provides less information than the HISTORY and RESULT variables do about the final outcome of the SFAS performance.

Practicality/feasibility: No extra time is required to develop this measure; it is available in archival records.

Susceptibility to contamination: There are reasons beyond the individual's performance/control (e.g., medical and involuntary drop reasons) that can account for not graduating.

Correlations with other Variables: Analyses to correlate SFAS Graduation with other variables have not been run yet because the process of building the SFdata base is not complete.

Variables that predict it best: SFAS decision-makers were found to rely heavily on ruck march scores in making the graduation decision (Teplitzky, 1991). Validation analyses have not been completed; these will be run when the database is completed.

Operational Measure	2

Measures: Peer Rankings of Leadership Potential (collected during SFAS)

Short Description of Measure: This ranking is done by all members of every team during SFAS. Each team member ranks every other member (but not themselves) from Most to Least in terms of their contribution to the team's overall performance. They also write a short justification for the decisions they made about the highest and lowest ranking individuals.

Psychometrics:

Scoring: The peer rank is computed from the variables: 1) size of the peer-rated group and 2) raw score ranking.

Relevance: Relevant to the assessment/selection domain and the interpersonal domain.

Comprehensiveness: Rating scales (if constructed to do so) can measure performance more comprehensively than most other measurement methods; performance ratings have been shown to be determined by declarative knowledge, procedural knowledge and skill, and motivation (McCloy, 1990).

Discriminability: The ranking procedure forces team members to make discriminations (more so than a rating procedure might).

Practicality/feasibility: This type of measure does not require a lot of resources to develop and does not require a lot of time to collect.

Susceptibility to contamination: These rankings of contribution to team performance can be subject to personal views about the individuals on the rater's team. The raters are not given training in how to avoid bias.

Correlations with other Variables: Peer ratings of leadership potential collected during first tour predict performance as an NCO in second tour (Campbell, Peterson, & Johnson, 1994).

Variables that predict it best: This variable is potentially useful as a predictor (e.g., of Q-course performance, SFAS pass/fail).

Operational Measure	3

Measure: Honors Received in SFQC (HONORS)

Short Description of Measure: This is a measure of the level of performance in the Q-course, available in the post-1992 SFQC longitudinal database for each time that the subject went through the course (1 - 9 times).

Psychometrics:

Scoring: Categorical variable: No Honors, Commandant's List (CL), Distinguished Graduate (DG), or Honor Graduate (HG). HG is highest.

Relevance: Relevant to the training domain; this is a measure of performance over the course of the training, which occurs before any job experience is gained.

Comprehensiveness: This measure captures the "grades" dimension of training performance.

Discriminability: This is a potentially useful measure for distinguishing between good and poor performers. However, it is not available for 4055 of 4192 subjects.

Practicality/feasibility: No extra time would be required to develop or administer this measure. However, it is available in archival records for only a limited subset -- post 1992 -- of Q-course students.

Susceptibility to contamination: There is a concern that others' honors status may influence the designation of an individual's honors status, if decide on the basis of a curve (e.g., if honors are given to a set number of Q-course students each year/session regardless of quality, then honors status is not comparable across years/sessions).

Correlations with other Criteria: Number of awards and certificates (a similar variable) was combined with ratings on effort and MOS proficiency to form a criterion composite for the convential Army -- Effort and Leadership (ELS). ELS during first tour was a good predictor of NCO performance.

Variables that predict it best: Validation analyses have not been completed; these will be run when the database is completed.

Operational Measure	4

Measure: Final Training Status (FINISH)

Short Description of Measure: The Q-course is an intensive training course where students are taught both the skills necesssary for all SF team members to have (such as navigation and small unit tactics) and the skills that each individual needs to have to perform as a specialist in his own assigned MOS (e.g., medical, combat engineer, weapons, communications).

The FINISH variable is contained in the SFQC database and describes the final outcome assigned to each student in the database -- whether they completed the course, are still eligible for repeating the course, or are no longer eligible (as of the specified end date FY92).

Psychometrics:

Scoring: Scoring categories are:
- Academic Relief (out due to problems with grades)
- Graduate - multiple tries (took course more than once)
- Graduate - first try (passed on first try)
- No entry since FY90 (not re-entered into course after FY90)
- Future entries possible (student would be allowed back)
- Relief - non-academic (out due to reason other than grades)

Relevance: Relevant to the training domain, this is an indicator of the final outcome recorded in the database as of the end of FY92.

Comprehensiveness: This is a summary-level variable, and contains more information than the simple graduated/not graduated outcome measure.

Discriminability: Variations in reasons for non-performance can be captured by this variable.

Practicality/feasibility: This measure is available from archival records; it does not require any development or administrative time.

Susceptibility to contamination: There are potential sources of variance beyond the individual's performance/control, such as personal/family crisis that causes a student to leave, or school changes in policy for setting passing scores, etc.

Correlations with other Criteria: These analyses have not yet been completed.

Variables that predict it best: Validation analyses have not been completed; these will be run when the database is completed.

Operational Measure	5

Measure: Retrained into a Different MOS (RETRAIN)

Short Description of Measure: This variable is contained in the SFQC longitudinal database. It identifies the soldiers who began training in one MOS, failed to graduate from that MOS, then retrained into another MOS in the Q-course. (These cases were identified by comparing trainees' first and last MOS.)
Psychometrics:

Scoring: "N" for No, were not retrained and "Y" for Yes, retrained into a different MOS.

Relevance: Relevant to training domain as an indicator of whether a trainee was able to complete the training in the initial MOS assignment or had to be "recycled" into another MOS.

Comprehensiveness: This is a measure of one component of training performance (finishing the training regimen that was started).

Discriminability: The dichotomous scoring yields very little discriminatory information.

Practicality/feasibility: This measure can be gathered from archival records; it does not require any development efforts.

Susceptibility to contamination: Other factors besides individuals' ability may enter here; for example, they may not desire to do well at the assigned MOS so that they can be reassigned.

Correlations with other Criteria: These analyses have not yet been completed.

Variables that predict it best: Validation analyses have not been completed; these will be run when the database is completed.

Measure: Total Tries in Fort Bragg Training (FBTRIES)

Short Description of Measure: This variable is contained in the SFQC Longitudinal Database. It indicates the total number of tries an individual participated in an uninterrupted SFQC class at Ft. Bragg. (This does not include medics who split their training; these trainees would have a value of zero.) For non-medics, this variable is a count of the number of times they attempted the SFQC before graduating or being relieved.

Psychometrics:

Scoring: 0 for medics who made it through the whole sequence once during the FY89 - 90 classes
1 for all trainees (including FY91 - 92 medics) who made it through the course in their first and only try
2 through 5 for all trainees who made it through the course in 2 - 5 tries

Relevance: Relevant to training domain; this is a summary level measure indicating the number of times required to make it succesfully through a course which should be completed in the first try.

Comprehensiveness: This is one component of training performance -- essentially a measure of whether they could make it through in one pass or had to repeat the course some number of times. This type of information could be added to other information (e.g., ratings, final outcome, etc.) to get a more "complete" picture for each subject.

Discriminability: This measure yields a little more information than just a "GO/NO GO" in one try measure.

Practicality/feasibility: This measure can be gathered from archival records; it does not require any development or administrative time.

Susceptibility to contamination: Factors other than ability can potentially affect this measure, such as a personal problem beyond the individual's control (e.g. death, illness of family member).

Correlations with other Criteria: These analyses have not yet been completed.

Variables that predict it best: Validation analyses have not been completed; these will be run when the database is completed.

Measure: Peer Rankings of Leadership Potential (to be collected during Q Course)

Short Description of Measure: This ranking will be done by all members of every team during the field phase of the Q-course, at three points in time: 1) at the end of the first three weeks, 2) two weeks later, and 3) at the end of the course. Each team member will rank every other member (but not themselves) from Most to Least in terms of their contribution to the team's overall performance. They will also write a short justification for the decisions they make about the highest and lowest ranking individuals. A very low ranking may be used (by SF decision-makers) as an indication that an individual would not get along well with team members.

Psychometrics:

Scoring: [We expect that these rankings could be computed the same as the SFAS rankings] The peer rank is computed from the variables: 1) size of the peer-rated group and 2) raw score ranking.

Relevance: Relevant to the training domain and the interpersonal domain.

Comprehensiveness: Rating scales (if constructed to do so) can measure performance more comprehensively than most other measurement methods; performance ratings have been shown to be determined by declarative knowledge, procedural knowledge and skill, and motivation (McCloy, 1990).

Discriminability: The ranking procedure forces team members to make discriminations (more so than a rating procedure might).

Practicality/feasibility: This type of measure does not require a lot of resources to develop and does not require a lot of time to collect.

Susceptibility to contamination: These rankings of contribution to team performance can be subject to personal views about the individuals on the rater's team. To make this measure useful for research purpose, the raters would need to be given training in how to avoid bias.

Correlations with other Variables: Peer ratings of leadership potenial collected during first tour predict performance as an NCO in second tour (Campbell et al., 1994).

Variables that predict it best: Validation analyses have not been completed; these can be run when the data are added to the database.

Measure: SFQC Land Navigation Written Test

Short Description of Measure: This is a measure of the ability to "maintain one's bearings with respect to points on a compass and to maintain appreciation of one's location relative to landmarks in the environment." It contains 20 items and has a 12 minute time limit. The test gives a schematic map that contains familiar landmarks; the examinee is given the direction of one landmark to another and must figure out the direction from a given third landmark to a specified fourth landmark. (Teplitzky, 1994). The examinee must correctly use and interpret maps containing detailed contour, relief, and location information.

Psychometrics:

Scoring: This test is scored as Pass or Fail.

Relevance: This test is relevant to training (Q-course) performance, specifically, performance using maps and associated information.

Comprehensiveness: This test could be a comprehensive test of orienting oneself in relation to landmarks if the questions capture all critical aspects of performing this task. The written mode of administration may limit the content that can be covered.

Discriminability: Students in the Q-course are given a dichotomous score. This scoring procedure does not yield as much discrimination information as a percentage correct (out of the total of 20) score.

Practicality/feasibility: This is a test that has already been developed; an alternate form might be useful to develop to be used for research purposes

Susceptibility to contamination: This measure is probably free from most typical sources of criterion contamination. For those who re-take the exam, practice effects may influence performance.

Correlations with other Criteria: Scores on this measure correlate -.04 with scores on the land navigation field exam (LNFLD) (Teplitsky, 1994). (The LNFLD score is a summary level score; it indicates whether the examinee passed the test on the first try or had one or more failures.)

Variables that predict it best: In multiple regression analyses, the Wonderlic was the best predictor of the land navigation written test (pass/fail) and the Map test was the next best predictor, and the Assembling Objects is the third of the three.

Measure: Land Navigation Field Exam

Short Description of Measure: This exam is a "Go"/"No-Go" practical exercise that requires an examinee to cover 18 kilometers of varying terrain. The examinee has to find four points in nine hours, starting at 2 a.m. (which requires some navigation under conditions of darkness). If an examinee fails on the first attempt, he takes remedial training and is retested one to two weeks later. If an examinee also fails the retest, he is either dropped from the Q-course or recycled into another class.

Psychometrics:

Scoring: This score is a count of the number of times that an examinee took and failed the exam (within a certain time interval in the database):
0 = no failures (passed on first try)
1 = failed first test but passed retest
2 = failed both initial test and retest
3, 4 etc. = failed subsequent tests after recycling into a later class

Relevance: This test is relevant to training (Q-course) performance.

Comprehensiveness: This test could be a comprehensive measure of spatial skills if the course requires the examinees to use most/all critical navigation skills. The mode of administration makes the test face valid to examinees.

Discriminability: The categorical nature of the scoring obscures some discriminatory information -- the performance on this test could be scored so that it yields more information (e.g., a scoring scheme based on number of wrong decisions, number of times traveled in the wrong direction, time required to locate each point, etc.).

Practicality/feasibility: This test is already developed. It takes a long time (9 hours) to administer but archival scores could be retrieved.

Susceptibility to contamination: Testing conditions (e.g., temperature, length of time in darkness -- time of year, etc.) may affect performance. Practice effects may occur for those who take the exam more than once.

Correlations with other variables: Scores on this measure correlate -.04 with scores on the written land navigation exam (Teplitzky, 1994). The written test is more a measure of map use skills. Preliminary data suggest that individuals from conventional army combat arms MOS are more likely to pass the land navigation field exam (personal communication, Wilderman).

Variables that predict it best: In multiple regression analyses, the only predictor that was significant was the Assembling Objects (AO) Test, and it explained only 7% of the variance in the criterion (Teplitzky, 1994). Higher scores on the AO test were associated with fewer failures on the land navigation test.

Proposed Measure	10

Measure: Promotion Rate

Short Description of Proposed Measure: We propose to develop a measure of promotion rate to indicate the soldiers' progress in the enlisted ranks. Using this as a measure of success is based on the assumption that soldiers performing at higher levels progress more quickly thorugh the enlisted ranks. We propose to develop the Promotion Rate scores using two sources: 1) soldiers' self-report of the number of recommendations they got for promotion before having been in grade for the required time period, and 2) from the available data in the Enlisted Master File (EMF).
For the Project A/Career Force data for first-tour job proficiency, the deviation score was included in the Maintaining Personnel Discipline (MPD) variable combined with Army-wide Ratings of Personnel Discipline and an Administrative Index for for Number of Article 15's and Flag Actions. For the second-tour NCO data, Promotion Rate was included in the Leadership (LEAD) variable combined with Army-wide Ratings of Leading/Supervisory skills, the Situation Judgment Test, and role play exercises.

Psychometrics:

Scoring: Scores are calculated within MOS, using a two step process. First, a grade deviation score was calculated from data available in the EMF; this adjusts the soldier's paygrade to the mean of those who have the same time in service. Then, the grade deviation score was combined with an indicator of whether the soldier had been recommended for promotion in the secondary zone.

Relevance: Relevant to the job performance domain; this is a summary-level indicator of how well one performs on the job relative to others within the same MOS.

Comprehensiveness: This is a comprehensive measure of how well a soldier advances within his/her MOS, but does not take into account differences across MOS (e.g., not all MOS offer similar opportunities for promotion).

Discriminability: This measure provides a way to compare all soldiers within the same MOS on the same scale.

Practicality/feasibility: This measure can be obtained through two sources: archival records and individual soldiers (e.g., self-report). No additional efforts are required to develop the measure; administrative time would be required to collect these data for SF subjects.

Susceptibility to contamination: EMF data is not always up-to-date; soldiers are relied on for accuracy in reporting the promotion recommendations.

Correlations with other variables: In Project A. the Administrative Index for Promotion Rate -- used as a predictor -- correlated with other Administrative Indices (Ns = 817 to 1,035); significant at $p < .01$): Awards $r = .31$; Article 15s/Flag Actions $r = -.19$; Physical Readiness $r = .14$; M16 Qualification $r = .14$; Military Training $r = .39$.

Variables that predict it best: Temperament was the best predictor (mean validity for composite scores = .32) of Maintaining Personal Discipline across nine Army enlisted jobs (corrected for range restriction and adjusted for shrinkage) (Campbell & Zook, 1991).

Proposed Measure 11

Measure: Number of Article 15s and Flag Actions (Disciplinary Actions)

Short Description of Proposed Measure: We propose to collect from individuals a self-report of the number of Articles 15 and Flag Actions they have received. For the convential Army, an index such as this was included in the Maintaining Personal Discipline variable, along with the Personal Discipline ratings (Army-wide).

Psychometrics:

Scoring: The value was the sum of the Flag Actions and Articles 15 that the soldier self-reported (in pay grades E-4 and above) (Campbell & Zook, 1990).

Relevance: Relevant to the job performance domain; this reflects negative aspects of performance.

Comprehensiveness: There may be other negative aspects of performance that are not captured by these formal recorded actions. Some negative performance instances/examples in the conventional Army might be useful for **prediction** of performance in the SF context (e.g., inability to function well as a team player or losing composure in stressful situations).

Discriminability: This measure distinguishes between soldiers who have had disciplinary problems and those who haven't; however, Riegelhaupt et. al. (1987) report a low base rate (11%) for this variable.

Practicality/feasibility: No additional effort is required to develop this measure: administrative time would be required to collect this for SF subjects.

Susceptibility to contamination: Soldiers are relied on to provide accurate numbers of their Article 15s and Flag Actions.

Correlations with other variables: Other variables also used as **predictors** in Project A correlated significantly at $p < .01$ with: Awards ($r = -.08$); Physical Readiness ($r = -.11$); Military Training ($r = -.16$); and Promotion Rate ($r = -.19$) (Campbell & Zook, 1990).

Variables that predict it best: Temperament was the best predictor (mean validity for composite scores = .32) of Maintaining Personal Discipline across nine Army enlisted jobs (corrected for range restriction and adjusted for shrinkage) (Campbell & Zook, 1991).

Proposed Measure	12

Measure: Number of Awards, Memorandum, and Certificates

Short Description of Proposed Measure: We propose to develop this measure of personal achievement. A form could be developed that contains a section for soldiers to report the number of Awards, Memorandum, and Certificates they have received.

The Personnel File Form was a self-report measure used during Project A data collections to obtain personnel file data. This index was included in the Maintaining Personal Discipline variable, along with the Personal Discipline ratings (Army-wide).

Psychometrics:

Scoring: The value was the sum of all awards, certificates, and memoranda that the soldier self-reported.

Relevance: Relevant to the job performance domain; this is a summary level indicator that reflects outstanding aspects of performance.

Comprehensiveness: There may be other critical aspects of high-end performance that are not captured by these formal recorded actions. Some performance instances/examples from the Special Forces context, if recorded, could be useful criterion measures (e.g., memorandum filed for outstanding interactions/negotiations with representatives from another culture).

Discriminability: This variable allows differences in performance levels (at the high end) to be captured.

Practicality/feasibility: Additional effort could be invested to revise this measure to reflect unique performance elements for the SF contect. Administrative time would also be required to collect these data for SF subjects.

Susceptibility to contamination: In using a self-report form, soldiers must be relied upon to provide accurate data. However, Army regulations do not require all letters, certificates to be placed in soldiers' 201 files, so the official records are not necessarily complete and up-to-date.

Correlations with other variables: Number of awards and certificates (a similar variable) was combined with ratings on effort and MOS proficiency to form a criterion composite for the convential Army -- Effort and Leadership (ELS). ELS during first tour was a good predictor of NCO performance.

Variables that predict it best: Validation analyses have not been completed; these will be run when the database is completed.

Operational Measure	13

Measure: NCOER and OER (Non-Commissioned Officer Evaluation Report) and Officer Evaluation Report

Short Description of Measure: These are the formal performance evaluation tools for operational use in the Army. The sections of the forms are:

Authentication - name and rank of rater, senior rater, and reviewer
Duty description - description of daily duties and scope, including reference to people, equipment, facilities, and dollars, as appropriate; description of areas of special emphasis; description of appointed duties
Values - answers to questions about adherence to Army ethics (loyalty, duty, selfless service, integrity), and to personal values (commitment, competence, candor, courage)
Responsibilities - dimensions are: competence, physical fitness and military bearing, leadership, training, responsibility and accountability
Overall Performance and Potential - ratings made by both the rater and senior rater

Psychometrics:

Scoring: Range of 1 - 125; mean of approximately 123 or 124

The Values section has 7 items, each with a Y/N rating
The Responsibilities section has 5 parts, each with 4 rating categories of "Excellence" (exceeds standards), "Success" (meets standards), "Needs (SOME) Improvement," and "Needs (MUCH) Improvement."

Relevance: Relevant to on-the-job performance; the total score is a summary of the scores on the dimensions that are considered important for NCO and Officer on-the-job performance .

Comprehensiveness: Rating scales inherently (if constructed to do so) can measure performance more comprehensively than most other measurement methods; performance ratings have been shown to be determined by declarative knowledge, procedural knowledge and skill, and motivation (McCloy, 1990).

Discriminability: Ratings collected for research purposes only (rather than for operational purposes) are thought to be useful for criterion measures. Rater training is considered to be useful for reducing the effects of rater errors. Communcation from Francis Grafton: there is almost no variability in scores (see "Scoring" above).

Practicality/feasibility: In general, ratings are a very practical criterion measurement method, assuming that care is taken in developing scales and training raters.

Susceptibility to contamination: Rater errors can contaminate these measures (communication from Francis Grafton). Leniency is the main problem; operational use of this performance appraisal system has inflated the ratings.

Correlations with other Criteria: No knowledge of research studies using NCOER/OER as criterion measure.

Variables that predict it best: No knowledge of research studies using NCOER/OER as criterion measure.

Proposed Measure 14

Measure: Hands-On (Common Task) Performance Tests

Short Description of Proposed Measure: We propose to develop hands-on measures for SF common tasks. SF common tasks are those that all SF soldiers are required to be able to perform regardless of their individual MOS, such as land navigation, small unit tactics, etc.)

The common task tests for Project A were developed as follows: 1) the task domain was defined on the basis of Army Manuals (Common Tasks), Army Occupational Survey Program (AOSP) data, and SME judgments of task characteristics; 2) tasks were selected to represent the domain of common tasks; and 3) tests were constructed through a process of: determining which tasks were conducive to the hands-on format, determining test conditions, listing performance measures, stating instructions for examinees, and developing scorer instructions.

Each hands-on test contains a set of activities to be performed to set the conditions for testing each examinee, instructions to be read to each examinee, instructions for administering the test, and a series of performance measures for scoring examinee behavior.

Psychometrics:

Scoring: The score on each hands-on test is the percent of total steps scored "Go."

Relevance: Relevant to the job performance domain; the process of defining the job universe and testing domain are critical to ensure that the final set of tested tasks are relevant

Comprehensiveness: The nature of hands-on testing (extensive time requirements) restricts the potential coverage of the performance domain, but the set of tests will be comprehensive to the extent that the behaviors tapped represent the critical aspects of performance.

Discriminability: In Project A, the degree of discriminability in test scores varied across tasks and MOS, but there was enough variation in the scores to show significant correlations with ASVAB scores (Knapp & Campbell, 1992).

Practicality/feasibility: Hands-on tests require extensive commitment of time and resources to administer.

Susceptibility to contamination: Potential sources of error are: scorer errors and differences across testing sites (if applicable).

Correlations with other variables: Hands-on test scores are highly correlated with other job proficiency measures such as written job knowledge tests (Campbell & Zook, 1991).

Variables that predict it best: Cognitive variables such as ASVAB scores and spatial test scores are good predictors of hands-on test performance (Campbell & Zook, 1991).

Proposed Measure	15

Measure: Hands-on (MOS-Specific) Performance Tests

Short Description of Proposed Measure: We propose to develop hands-on measures for each SF MOS (18B, 18C, 18D, 18E, and 18A/180A).
In Project A, the specific hands-on tests for each MOS were developed as follows: 1) the task domain was defined on the basis of Army Manuals (Soldiers' Manuals), Army Occupational Survey Program (AOSP) data, and SME judgments of task characteristics; 2) tasks were selected to represent each MOS; and 3) tests were constructed through a process of: determining which tasks were conducive to a hands-on format, determining test conditions, listing performance measures, stating instructions for examinees, and developing scorer instructions.
Each hands-on test contains a set of activities to be performed to set the conditions for testing each examinee, instructions to be read to each examinee, instructions for administering the test, and a series of performance measures for scoring examinee behavior.

Psychometrics:

Scoring: The score on each hands-on test is the percent of total steps scored "Go."

Relevance: Relevant to the job performance domain; the process of defining the job universe and testing domain are critical to ensure that the final set of tested tasks are relevant.

Comprehensiveness: The nature of hands-on testing (extensive time requirements) restricts the potential coverage of the performance domain, but the set of tests will be comprehensive to the extent that the behaviors tapped represent the critical aspects of performance.

Discriminability: In Project A, the degree of discriminability in test scores varied across tasks and MOS, but there was enough variation in the scores to show significant correlations with ASVAB scores (Knapp & Campbell, 1992).

Practicality/feasibility: Hands-on tests require extensive commitment of time and resources to administer.

Susceptibility to contamination: Potential sources of error are: scorer errors and differences across testing sites (if applicable).

Correlations with other variables: Intercorrelations of Project A criteria show that performance on a standardized job sample is a significant component of performance, but not all of it (Campbell & Zook, 1991). Total hands-on score (corrected for attenuation) correlated .34 with overall performance rating. Total hands-on score correlated more highly with Core Technical Performance (.74) and with General Soldiering Proficiency (.72) than with Effort & Leadership (.26), Personal Discipline (.15), and Fitness and Military Bearing (.07). [These data were reported for Batch A MOS in Project A.]

Variables that predict it best: Cognitive variables such as ASVAB scores and spatial test scores are good predictors of hands-on test performance (Campbell & Zook, 1991).

Proposed Measure 16

Measure: End of Training Hands-on (MOS-Specific) Performance Tests

Short Description of Proposed Measure: We propose to develop end-of-training hands-on measures for each SF MOS (18B, 18C, 18D, 18E, and 18A/180A).
In Project A, the specific hands-on tests for each MOS were developed as follows: 1) the task domain was defined on the basis of Army Manuals (Soldiers' Manuals), Army Occupational Survey Program (AOSP) data, and SME judgments of task characteristics; 2) tasks were selected to represent each MOS; and 3) tests were constructed through a process of: determining which tasks were conducive to a hands-on format, determining test conditions, listing performance measures, stating instructions for examinees, and developing scorer instructions.
Each hands-on test contains a set of activities to be performed to set the conditions for testing each examinee, instructions to be read to each examinee, instructions for administering the test, and a series of performance measures for scoring examinee behavior.

Psychometrics:

Scoring: The score on each hands-on test is the percent of total steps scored "Go."

Relevance: Relevant to the training performance domain; the process of defining the training universe and testing domain are critical to ensure that the final set of tested tasks are relevant.

Comprehensiveness: The nature of hands-on testing (extensive time requirements) restricts the potential coverage of the performance domain, but the set of tests will be comprehensive to the extent that the behaviors tapped represent the critical aspects of training.

Discriminability: In Project A, the degree of discriminability in test scores varied across tasks and MOS, but there was enough variation in the scores to show significant correlations with ASVAB scores (Knapp & Campbell, 1992).

Practicality/feasibility: Hands-on tests require extensive commitment of time and resources to administer.

Susceptibility to contamination: Potential sources of error are: scorer errors and differences across testing sites (if applicable).

Correlations with other variables: Intercorrelations of Project A criteria show that performance on a standardized job sample is a significant component of performance, but not all of it (Campbell & Zook, 1991). Total hands-on score (corrected for attenuation) correlated .34 with overall performance rating. Total hands-on score correlated more highly with Core Technical Performance (.74) and with General Soldiering Proficiency (.72) than with Effort & Leadership (.26), Personal Discipline (.15), and Fitness and Military Bearing (.07). [These data were reported for Batch A MOS in Project A.]

Variables that predict it best: Cognitive variables such as ASVAB scores and spatial test scores are good predictors of hands-on test performance (Campbell & Zook, 1991).

Measure: MOS-Specific SF Behaviorally Anchored Rating Scales

Short Description of Measure: In the job analysis for SF jobs (Delivery Order #1), 5 MOS- Specific behavioral dimensions or "performance categories" were developed for each SF MOS through the steps of collecting critical incidents, forming initial dimensions, conducting a retranslation exercise, and revising dimensions and constructing anchors. The performance categories are:

MOS-Specific Performance Categories

18B Weapons Expert	Operating and Maintaining DIrect-Fire Weapons
	Employing Indirect-Fire Weapons and Techniques
18C Engineer	Employing Demolitions Techniques
	Constructing for Mission-Related Requirements
18E Communications	Following Communications Procedures and Policies
Expert	Assembling and Operating Commo Equipment
18D Medic	Evaluating and Treating Medical Conditions and Injuries
	Determining and Administering Medications and Dosages
	Ensuring Standards of Health-Related Facilities, Conditions, and Procedures
18A/180A Leader	Considering Subordinates
	Providing Direction

Psychometrics:

Scoring: Raters are asked to rate performance on each scale using a 7-point rating scale with the scale points 1 and 2 (Needs Improvement); 3, 4, and 5 (Effective); and 6 and 7 (Highly Effective). Each scale point has one or two critical incidents listed to illustrate performance examples.

Relevance: Relevant to the job performance domain.

Comprehensiveness: These performance categories were formed on the basis of a comprehensive job analysis. At a more general level -- the method of performance ratings -- ratings have been shown to be determined by declarative knowledge, procedural knowledge and skill, and motivation (McCloy, 1990).

Discriminability: As with all rating methods, there is a tendency for rater errors (halo, central tendency, and leniency) to affect the rating distributions. However, rater training programs can be used in conjunction with pledges of confidentiality of the data to counteract rater tendencies (Pulakos and Borman, 1986).

CONTINUED NEXT PAGE

Practicality/feasibility:	Ratings are relatively practical in terms of ease of developing and collecting them. Care must be taken in the developmental stages and data quality should be ensured by providing rater training.
Susceptibility to contamination:	In general, raters' evaluations are affected by generalizations/stereotyping, personal beliefs about the ratee, personal standards for performance, carelessness, etc. (Pulakos, 1984).
Correlations with other variables:	In Project A, MOS-specific ratings were pooled with ratings of effort and technical skill to form the Effort and Leadership (ELS) criterion variable.
Variables that predict it best:	The ASVAB is a good predictor of ELS and personality measures yield good incremental validity over and above the ASVAB for predicting ELS (Campbell & Zook, 1991).

Measure: SF-Common Behaviorally Anchored Rating Scales

Short Description of Measure: In the job analysis for SF jobs (Delivery Order #1), 15 SF- Common behavioral dimensions or "performance categories" were developed through the steps of collecting critical incidents, forming initial dimensions, conducting a retranslation exercise, and revising dimensions and constructing anchors. The performance categories are:

SF-Common Performance Categories

Teaching Others
Building and Maintaining Effective Relationships with Indigenous Populations
Handling Difficult Interpersonal or Intercultural Situations
Using and Enhancing Own Language Skills
Troubleshooting and Solving Problems
Decision Making
Planning and Preparing for Missions
Contributing to the Team Effort and Morale
Showing Initiative and Extra Effort
Displaying Honesty and Integrity
Confronting Physical and Environmental Challenges
Navigating in the Field
Being Safety Conscious
Administering First Aid and Treating Casualties
Handling Administrative Duties

Psychometrics:

Scoring: Raters are asked to rate performance on each scale using a 7-point rating scale with the scale points 1 and 2 (Needs Improvement); 3, 4, and 5 (Effective); and 6 and 7 (Highly Effective). Each scale point has one or two critical incidents listed to illustrate performance examples.

Relevance: Relevant to the job performance domain.

Comprehensiveness: These performance categories were formed on the basis of a comprehensive job analysis. At a more general level -- the method of performance ratings -- ratings have been shown to be determined by declarative knowledge, procedural knowledge and skill, and motivation (McCloy, 1990).

Discriminability: As with all rating methods, there is a tendency for rater errors (halo, central tendency, and leniency) to affect the rating distributions. However, rater training programs can be used in conjunction with pledges of confidentiality of the data to counteract rater tendencies (Pulakos and Borman, 1986).

CONTINUED NEXT PAGE

Practicality/feasibility:	Ratings are relatively practical in terms of ease of developing and collecting them. Care must be taken in the developmental stages and data quality should be ensured by providing rater training.
Susceptibility to contamination:	In general, raters' evaluations are affected by generalizations/stereotyping, personal beliefs about the ratee, personal standards for performance, carelessness, etc. (Pulakos, 1984).
Correlations with other Criteria:	In Project A, factor analyses of Army-wide rating scales yielded three ratings factors: 1) Effort and Leadership, 2) Personal Discipline, and 3) Physical Fitness and Military Bearing (Peterson, Hough, Dunnette, Rosse, Houston, & Toquam., 1990).
Variables that predict it best:	Project A data showed that the ASVAB does predict performance in each of the three ratings factors. Personality measures have been incremental validity over the ASVAB (Campbell & Zook, 1991).

Proposed Measure 19

Measure: Written (Common Task) Job Knowledge Tests

Short Description of Proposed Measure: We propose to develop written job knowledge tests to cover the SF common tasks -- those tasks that all members of SF are expected to perform, regardless of the specific MOS they are also trained in.

For Project A, written tests for the common tasks were developed as follows: 1) the task domain was defined on the basis of the Army Common Task manual, Army Occupational Survey Program (AOSP) data, and SME judgments of task characteristics; 2) tasks were selected to represent the domain; and 3) tests were constructed to emphasize performance knowledge, through a process of item construction, review, pilot testing, and revision by test development experts.

Psychometrics:

Scoring: The score on each test is the percent of correct responses.

Relevance: Relevant to the job performance domain; the process of defining the job universe and testing domain are critical to ensure that the final set of tested tasks are relevant.

Comprehensiveness: Specifically, in Project A, the Army sampled twice as many tasks for written testing as for hands-on testing. In general, more tasks can be tested through written testing than through hands-on testing, so there is the potential for more comprehensive coverage of the performance domain. Also, tasks that may be difficult/infeasible to test in the hands-on mode can be tested with a written format.

Discriminability: Nine Army tests ranged in difficulty from 56% to 70% correct (Knapp & Campbell, 1992). Similar types of tests for other services varied from 44 to approx. 74% correct.

Practicality/feasibility: The Army tested both the hands-on content (performance-based items) and other content just in written mode. The quality of the performance-based items is dependent on the effort put into development and pilot testing. Both the administration and scoring of these tests are straightforward, convenient, and economical. SF would need to develop their own versions of this test type or revise common task tests already developed.

Susceptibility to contamination: Scores for less "verbal" examinees may not reflect their true scores (due to the amount of reading involved). Performance-based items (with pictures) help to reduce this potential effect.

Correlations with other variables: Hands-on test scores are highly correlated with other job proficiency measures such as written job knowledge tests (Campbell & Zook, 1991).

Variables that predict it best: Cognitive variables such as ASVAB and spatial test scores are good predictors of hands-on test performance (Campbell & Zook, 1991).

Proposed Measure	20

Measure: Written (MOS-Specific) Job Knowledge Tests

Short Description of Proposed Measure: We propose to develop written job knowledge tests for each SF MOS (18B, 18C, 18D, 18E, and 18A/180A).
In Project A, the specific written tests for each MOS were developed as follows: 1) the task domain was defined on the basis of the MOS-specific Soldiers' Manuals), Army Occupational Survey Program (AOSP) data, and SME judgments of task characteristics; 2) tasks were selected to represent the domain; and 3) tests were constructed to emphasize performance knowledge, through a process of item construction, review, pilot testing, and revision by test development experts.

Psychometrics:

Scoring: The score on each test is the percent of correct responses.

Relevance: Relevant to the job performance domain; the process of defining the job universe and testing domain are critical to ensure that the final set of tested tasks are relevant.

Comprehensiveness: This test could be comprehensive if the 100 questions capture all critical aspects of performance in each MOS. The written mode of administration may limit the content that can be covered, e.g., the behavioral aspects of training and leadership.

Discriminability: Nine Army tests ranged in difficulty from 56% to 70% correct (Knapp & Campbell, 1992). Similar types of tests for other services varied from 44 to approx. 74% correct.

Practicality/feasibility: The Army tested both the hands-on content (performance-based items) and other content just in written mode. The quality of the performance-based items is dependent on the effort put into development and pilot testing. Both the administration and scoring of these tests are straightforward, convenient, and economical. SF would need to develop their own versions for each MOS.

Susceptibility to contamination: Scores for less "verbal" examinees may not reflect their true scores (due to the amount of reading involved). Performance-based items (with pictures) help to reduce this potential effect.

Correlations with other variables: Intercorrelations of Project A criteria show that performance on a standardized job sample is a significant component of performance, but not all of it (Campbell & Zook, 1991). Total hands-on score (corrected for attenuation) correlated .34 with overall performance rating. Total hands-on score correlated more highly with Core Technical Performance (.74) and with General Soldiering Proficiency (.72) than with Effort & Leadership (.26), Personal Discipline (.15), and Fitness and Military Bearing (.07). [These data were reported for Batch A MOS in Project A.]

Variables that predict it best: Cognitive variables such as ASVAB scores and spatial test scores are good predictors of hands-on test performance (Campbell & Zook, 1991).

Proposed Measure 21

Measure: Task Performance Ratings

Short Description of Proposed Measure: We propose to develop rating scales to be used for rating performance on critical SF tasks and guidelines for selecting for which tasks performance will be rated. These scales can be used for rating both MOS-specific and SF common tasks; supervisors and/or peers can make these ratings.

For the Project A data collection, these ratings were of performance on tasks also tested in hands-on mode. Both peers and supervisors were asked to rate a soldier's performance on each task.

Psychometrics:

Scoring: A rating for performance on each task was made using a 7-point rating scale.

Relevance: Relevant to the job performance domain. Rating scales inherently (if constructed to do so) can measure performance more comprehensively than most other measurement methods; performance ratings have been shown to be determined by declarative knowledge, procedural knowledge and skill, and motivation (McCloy, 1990).

Comprehensiveness: The general advantage of the method applies here: rating scales inherently (if constructed to do so) can measure performance more comprehensively than most other measurement methods; performance ratings have been shown to be determined by declarative knowledge, procedural knowledge and skill, and motivation (McCloy, 1990).

Discriminability: As with all rating methods, there is a tendency for rater errors (halo, central tendency, and leniency) to affect the rating distributions. However, rater training programs can be used in conjunction with pledges of confidentiality of the data to counteract rater tendencies (Pulakos and Borman, 1986).

Practicality/feasibility: Ratings are relatively practical in terms of ease of developing and collecting them. Care must be taken in the developmental stages and data quality should be ensured by providing rater training.

Susceptibility to contamination: These ratings are subject to raters' generalizations/stereotyping, personal beliefs about the ratee, and personal standards for performance (Pulakos, 1984).

Measure: Situational Judgment Test

Short Description of Measure: This is a multiple-choice paper-and-pencil test developed as a measure of supervisory skill for NCOs (second-tour soldiers). There are 35 items (comprising ten behavior dimensions); each requires the soldier to read a scenario describing a problem a supervisor might face, then select the most and least effective response alternatives from a set of five. This test is essentially a job knowledge test for supervisory job content. [A similar test has been developed for the Army in the ECQUIP project and has been piloted but not yet keyed or used in the large-scale data collection.]

Psychometrics:

Scoring: After investigating five alternative scoring strategies, the "M-L Effectiveness" score was selected as the most promising. This is a composite of the two effectiveness scores for each item (obtained by subtracting the mean effectiveness of the response chosen as the least effective from the mean effectivenss of the response chosen as the most effective), averaged across items.

Relevance: Relevant to the job performance domain. The process of defining the supervisory domain is critical to ensure that the final set of problem scenarios is relevant.

Comprehensiveness: In general, more supervisory situations can be presented through written testing than through role play (hands-on) testing, so there is the potential for more comprehensive coverage of the supervisory performance domain. Also, problems that may be difficult/infeasible to test in the role play mode may be testable with a written format.

Discriminability: Five alternative scoring strategies all resulted in scores with "reasonable variance" (Campbell & Zook, 1991).

Practicality/feasibility: The quality of the performance-based items is dependent on the effort put into development and pilot testing. Both the administration and scoring of these tests are straightforward, convenient, and economical once the test is developed and scoring procedures are identified.

Susceptibility to contamination: Respondents' answers may indicate the judgments that they have been trained to adopt (by their superiors) rather than their own personal opinions.

Correlations with other variables: In one study (Motowidlo, Dunette, & Carter, 1990) of 120 subjects, aptitude test measures did not correlate with the situational judgment (SJ) test scores, except for GPA in major ($r = .30$, $p < .05$). However, SJ ratings did correlate significantly with interview ratings of interpersonal skills ($r = .21$), communication skills ($r = .16$) and negotiation ratings ($r = .50$).

Variables that predict it best: Supervisory Experience and How Often Required to Supervise correlated significantly with SJT scores ($r = .14$ and $r = .15$, respectively; $p < .05$). for the preferred scoring procedure.

Proposed Measure	23

Measure: Computer-Based Simulations of SF Mission-Planning and Decision-Making Tasks/Scenarios

Short Description of Proposed Measure: We propose to develop simulations of complex scenarios to measure behaviors tapped by the behavioral dimensions developed during the Special Forces Job Analysis. These might require subjects to make decisions at various levels and the results of those decisions would affect what options were made available to them at each decision point. The simulations would be developed so that they are feasible to administer on the computer, and such that they could be used for training, as well as testing, purposes. We expect these exercises to focus on mission planning and decision making behaviors.

Measure: End-of-Training Written School Knowledge Test

Short Description of Proposed Measure: We propose to develop end-of-training written school knowledge tests for each SF MOS.
In Project A, these tests are paper-and-pencil achievement tests designed to assess soldiers' level of knowledge after they finish MOS-specific Advanced Individual Training. The number of multiple choice items per test ranges from 97 to 180; the items measure both technical (MOS-relevant) knowledge and Army-wide knowledge.
To develop these "School Knowledge" tests, an initial item pool was developed and reviewed by job incumbents and school trainers. The items were pilot tested and revised, field tested and revised, then used in the Concurrent Validation and Longitudinal Validation (Campbell & Zook, 1990).

A comparable test could be developed to test Q-course students after they have completed all phases of their training.

Psychometrics:

Scoring: Basic composite scores for examinees were derived for each MOS test.

Relevance: Relevant to the training performance domain.

Comprehensiveness: The written format allows for more of the training domain (potentially more tasks and types of tasks) to be tested (vs. hands-on mode).

Practicality/feasibility: The quality of the items depends on the effort put into development and pilot testing. Both the administration and the scoring of these tests are straightforward, convenient, and economical.

Susceptibility to contamination: To the extent that trained performance is on hands-on aspects of the jobs, the written format may penalize those who are less verbal but still very knowledgeable.

Correlations with other variables: Information about job knowledge tests is relevant here, due to the similarity between school knowledge and hands-on tests. Intercorrelations of Project A criteria show that performance on a standardized job sample is a significant component of performance, but not all of it (Campbell & Zook, 1991). Total hands-on score (corrected for attenuation) correlated .34 with overall performance rating. Total hands-on score correlated more highly with Core Technical Performance (.74) and with General Soldiering Proficiency (.72) than with Effort & Leadership (.26), Personal Discipline (.15), and Fitness and Military Bearing (.07). [These data were reported for Batch A MOS in Project A.]

Variables that predict it best: Again, information about job knowledge tests is relevant here, due to the similarity between school knowledge and hands-on tests. Cognitive variables such as ASVAB scores and spatial test scores are good predictors of hands-on test performance (Campbell & Zook, 1991).

Proposed Measure	25

Measure: Peer and Instructor Ratings in SFQC

Short Description of Proposed Measures: These ratings will be collected from peers and from instructors during the SF Qualification Course, the formal training course taken by those who are selected to attend after completing SFAS (the SF Assessment and Selection process).

Rating scales will be developed to cover important aspects or dimensions of performance in the course. There will be scales that cover the MOS-specific segments of the Q-course and scales that cover the segments of the course that all trainees complete (e.g., land navigation and small-unit tactics).

Peers and instructors will receive training in how to make their ratings more objective, rather than subjective.

Proposed Measure	26

Measure: Cadre Ratings During Robin Sage Role-Play Exercise

Short Description of Proposed Measures: We propose to develop a structured rating system for cadre members to use during the Robin Sage exercise. ("Cadre members" are those in SF who conduct SFAS and Q-course. In the Q-course, they train the students and then, during the Robin Sage exercise, they rate the performance of the students.)

Robin Sage is the name given to the lengthy exercise conducted at the end of the Q-course. The scenario is as follows: a team begins the exercise by jumping into the deep forest (in N.C.). They must conduct an unconventional warfare exercise; local townspeople role play guerilla forces parts.

Rating scales will be developed to cover important aspects or dimensions of performance in the Robin Sage exercise, e.g., making decisions, negotiating with guerrillas, building rapport with guerillas.

Cadre members will receive training in how to make objective observations of performance incidents, how to record performance incidents, and how to use the recorded objective information to make objective, performance-based ratings of individual students' performance.

Measure: Self Development Test (SDT)

Short Description of Measure: The self-development program for NCOs (levels E5 - E7) consists of individual study, research, professional reading, application and self-assessment. The SDT is a three-part test designed to measure and guide growth in the skills and competencies needed by NCOs to develop as leaders. The three parts of the test are:

1) Army Leadership (developed by the Center for Army Leadership)
2) Training Management Principles (developed by the US Army Sergeants Major Academy)
3) MOS Knowledge (developed by the MOS Proponent)

This is a formally adminstered written test that contains approximately 100 questions and requires about 2 hours to complete. A different test is developed every year for each MOS (approx. 650) and skill level. It has been used so far to help NCOs evaluate self development progress and to focus future development and training efforts in any deficiency areas. The SDT will be implemented for school selection and promotion decisions in FY94 for the active component and FY95 for the reserve component. The sections break out as follows:

1) 20 questions on Leadership section (taken from 3 Lp manuals)
2) 20 questions on Training Management (taken from 1 manual)
3) approximately 60 questions on MOS knowledge (taken from SM)

The sections covering Lp and TM cover the same content within a rank and differ across some SDT versions only for test security purposes. The 18 series MOS (Special Forces) do use these tests.

Psychometrics:

Scoring: The possible score range is 0 - 100%.
The average score varies across the 650 MOS, but the grand mean (average average) score is about 78%.

Relevance: Relevant to the job performance domain; the tests are likely to be relevant to the extent that the questions are faithful to task content requirements.

Comprehensiveness: This test could be comprehensive if the 100 questions capture all critical aspects of performance in each MOS and in the leadership and training areas. The written mode of administration may limit the content that can be covered, e.g., the behavioral aspects of training and leadership. In addition, the comprehensiveness can change from year to year as the test itself changes.

Discriminability: Up to this point, performance differences are probably due more to differences in time spent to prepare for the test than to true performance variability. This is due to the fact that it was not used operationally until FY94. Of 80,000 test takers, three-quarters reported studying less than 10 hours and about a third didn't study at all.

Practicality/feasibility: The yearly renewal of these tests necessarily requires development time, but should improve test security and reduce practice effects over the years.

CONTINUED NEXT PAGE

Susceptibility to contamination:	The differences in preparation time can contaminate the scores -- the scores may reflect preparation time more than actual knowledge. NCOs are to prepare for these tests on their own time rather than on unit time.
Correlations with other Criteria:	SDT scores probably correlate highly with SQT scores for those who have SQT scores. These two types of tests cover some similar content (e.g., MOS knowledge) but the administration methods and support differ (e.g., time was allocated to unit level to study for and administer the SQT; preparation is done on individuals' own time for the SDT).
Variables that predict it best:	Validities have not been calculated since the SDT has been used only for self-evaluation purposes until the end of FY93. Operational testing is taking place during FY94; data will be analyzed starting in late fall of 1994.

Operational Measure	28

Measure: Defense Language Proficiency Test (DLPT)

Short Description of Measure: This is a test of language proficiency that includes sections on: a) reading comprehension (1-1/2 hours), b) speaking proficiency (2-1/2 hours), and c) one-on-one interview (45 min.). The test is specific to the language trained and is taken when the language training has been completed.

Psychometrics:

Scoring: An examinee receives a category score denoting his level of proficiency with a specific language, e.g., level 2 in Spanish.

Relevance: Relevant to [lanuage] training performance.

Comprehensiveness: Three essential parts of language performance are tapped: listening, reading, and speaking, which the military now considers to be important parts of the language criterion domain.

Discriminability: This test yields scores which allow individuals to be assigned to a language that they will be capable of learning.

Practicality/feasibility: This test takes a lot of time to administer - approximately 5 hours per examinee.

Susceptibility to contamination: Unknown

Correlations with other variables: DLPT test scores could not be correlated with academic attrition because those who did not pass the course also did not take the exam.

Variables that predict it best: In the Silva and White (1993) study, correlations of the ASVAB subtests, g, and the Defense Language Aptitude Battery (DLAB) with the Listening and Reading scores on the DLPT ranged from .43 to .73. The measure of g was the best predictor and the DLAB was the next best predictor. The correlations of the predictors with the Speaking criterion ranged from .16 to .42. DLAB was a better predictor for Speaking than g was.

Proposed Measure	29

Measure: Language School Grades

Short Description of Proposed Measure: We propose to collect grades recorded for students during their language training courses. We propose to interview language training instructors to learn about how they assess the students' level of learning and what types of measures they use. We will decide whether to use aggregated measures (e.g., sum of quiz and oral response grades) or summary measures (e.g., final written and oral test grades, or overall course grade).

Proposed Measure	30

Measure: Language School Instructor Ratings

Short Description of Measure: We propose to develop a structured rating system for language school instructors to use to rate performance of students in language training.

We propose to work with language school instructors to first develop rating scales to cover important aspects or dimensions of performance in language training.

Language school instructors will receive training in how to make objective observations of language performance, how to record performance incidents, and how to use the recorded objective information to make objective, performance-based ratings of individual students' performance.

Criterion Description Form 31

Measure: Performance on Exercises at the Joint Readiness Training Center (JRTC)

Short Description of Measure: The current emphases in exercises run at the JRTC are on training and the team level. The specific purposes are to: provide realistic training for units, leaders, and soldiers, and to provide unit-level (team) performance feedback. Teams perform several cycles of mission planning, isolation, preparation, execution, and after-action review; during this process, they are observed by Observer-Controllers (OCs). Teams are given feedback at the team, not the individual, level. The OCs record qualitative information in "gray books" and everything is videotaped (Dyer, 1994).

Archival information is available -- at the unit level -- from these direct (but often incomplete) sources:
1) Task Force After Action Reviews (TF AARs) - conducted by the senior OC after mission/phase is complete, to provide training feedback (2.5 hours length). Videotapes and slide copies are archived. The content is: a) short summary of mission from various viewpoints plus critique, b) mission planning and preparation in each battlefield operating system (BOS) plus discussion, c) mission execution summary plus critique, and d) description of mission planning, preparation, and execution -- by opposing force.
2) Company After Action Reviews (AARs) - conducted by company OC after each mission phase (1 -3/4 hours length); they serve as a discussion and learning session. Videotapes of the AARs are archived but no paper records are kept. The format and topics discussed are not standardized, but usually include mission planning, preparation, and execution phases, plus the opposing force critique of the unit performance.
3) Take Home Packages (THPs) - a report written by the OC about unit performance, provided to each task force at the end of the rotation. Hard and soft copies are archived. Sections of the report cover brigade task force trends, battalion task force missions, and detail on each mission and its outcomes. Also included are strengths, areas for improvement, and training recommendations.
4) Training and Evaluation Outline (T&EO) Data Base - the purpose of this data base is to provide an archival record of performance ratings on units and echelons; these ratings of performance on tasks are made by OCs for the unit and echelon level.

To develop measures of individual performance, we propose to work with JRTC instructors to develop individual rating scales and to collect observer ratings of individuals' performance. If live performance will not be available, we could score videotapes of performance.

Psychometrics:

Scoring: Dyer (1994) transcribed the audio portions of the tapes, developed coding procedures (or used already developed coding schemes), and content analyzed the AARs and the THPs. The three main archical sources (AARs, THPs, T& EO) were compared in terms of their adequacy of coverage of four content areas (irrelevant for our purposes).

CONTINUED NEXT PAGE

Relevance:	Relevant to training performance -- at the unit level. Perhaps measures could be developed for individual level performance so that videotaped performance or real-time performance could be scored. Another possibility is that measures of team performance could be used to reflect the leader's performance (e.g., the leader is able to influence the team to do X).
Comprehensiveness:	To get a full picture of the missionand summaries of the outcomes, Dyer (1994) recommends using all three sources of information. However, there was a lot of missing data in the T & EO data base. Not all information expected to be available on videotape was actually available.
Discriminability:	As with all rating methods, there is a tendency for rater errors (halo, central tendency, leniency) to affect the rating distributions. However, rater training programs can be used in conjunction with pledges of confidentiality of the data to counteract rater tendencies (Pulakos & Borman, 1986).
Practicality/feasibility:	Archival data is available but may not be useful; special permission would have to be obtained from those who run JRTC to collect individual-level measures.
Susceptibility to contamination:	The methods for recording the data provide many opportunities for errors, inconsistencies, and incomplete data.

Proposed Measure	32

Measure: "Client" Ratings

Short Description of Measure: In reponse to a suggestion from SF individuals, we propose to develop a structured rating system for "clients" to use to rate performance of SF individuals. "Clients" can be defined as all those who the SF individuals may "work for" (such as the American Ambassador when in another country or the host nation ambassador when in a host nation) or work with (such as the officers or NCOs of the host nation forces or the guerilla chief) when on a mission.

We propose to interview a representative sample of these "clients" to first develop rating scales to cover important aspects or dimensions of performance.

There are, however, a variety of potential problems associated with trying to develop and conduct this type of rating system. For example, intercultural issues could be a concern -- people from other cultures may not react favorably to participating in either developing or using such an approach. The concept of rating SF performance may be too "foreign" to them.

If this is determined to be a viable measurement option, we would train "clients" how to make objective observations of performance, how to record performance incidents, and how to use the recorded objective information to make objective, performance-based ratings of individual SF members' performance.

Appendix J
Criterion Description Bibliography

Campbell, J.P., Peterson, N.G., & Johnson, J. (1994). The prediction of future performance from current performance and from training performance. In J.P. Campbell & L.M. Zook Building and Retaining the Career Force: New Procedures for Accessing and Assigning Army Enlisted Personnel Annual Report, 1993 Fiscal Year. Alexandria VA: U. S. Army Research Institute for the Behavioral and Social Sciences

Campbell, J.P.,& Zook, L.M. (Eds.) (1990). Building and retaining the career force: New procedures for accessing and assigning Army enlisted personnel. (ARI TR 952). Alexandria VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Campbell, J.P.,& Zook, L.M. (Eds.) (1991). Improving the selection, classification, and utilization of Army enlisted personnel: Final report on Project A (ARI RR 1597). Alexandria VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Dyer, J.L. (1994). A comparison of information in the joint readiness training center archival records (ARI RR 1659). Alexandria VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knapp, D.J.,& Campbell, J.P. (1992). Building a joint-Service classification roadmap: Criterion-related issues (AL/HR-TP-1993-0028) Brooks Air Force Base TX: Armstrong Laboratory.

McCloy, R.A. (1990). A new model of job performance: An integration of measurement, prediction, and theory. Unpublished Ph.D. dissertation. University of Minnesota, Minneapolis, MN.

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.

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.

Pulakos, E.D. (1984). A comparison of rater training programs: Error training and accuracy training. Journal of Applied Psychology, 69, 581-588.

Pulakos, E.D., & Borman, W.C. (Eds.) (1986). Development and field test of Army-wide rating scales and the rater orientation and training program (ARI Technical Report 716). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Riegelhaupt, B.J., Harris, C.D., & Sadacca, R. (1987). The development of administrative measures as indicators of soldier effectiveness (ARI Technical Report 754). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Silva, J.M. (1994). Examining the self-development test for race and gender fairness (ARI Project No. 20465803D730, Technical Report). Alexandria VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Silva, J.M., & White, L.A. (1993). Relation of cognitive aptitudes to success in foreign language training. Military Psychology, 5, 2, 79-93.

Teplitsky, M.L. (1991). Physical performance predictors of success in Special Forces Assessment and Selection (ARI RR 1606). Alexandria VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Teplitsky, M.L. (1994, June). Spatial abilities and land navigation performance. Paper presented at the Army Science Conference.

Appendix K
Criterion Expert Judgment Exercise Instructions

INSTRUCTIONS FOR COMPLETING CRITERION MEASURE RATING EXERCISE

Materials

You should have hard copies of two important sets of information in front of you:

1. (Potential) Criterion Measure Descriptions -- there are 32 measures included in this set. There are three major kinds of measures in this set:

   (1) Proposed -- measures that we are considering for development for Special Forces.

   (2) Experimental -- measures that have been developed and field tested but are not currently in use.

   (3) Operational -- measures that are currently in use.

   The measure description gives a brief review of what the measure is, and additional psychometric information that was available.

2. SF Performance Categories -- there are 26 performance categories; these were formed during the SF job analysis study. You have these in your copy of the final report from that project (or in one of the briefing packets).

You should also have a soft copy of the rating form -- I will give you a copy of this file. You will make your ratings directly into the QUATRRO-PRO file while sitting at your computer, rather than on paper. The file for you to use has your initials on it plus "rat.wb1" (example "jcrat.wb1"). The file contains the matrix in which you will record your judgments about the extent to which each of the measures would "measure" each of the performance categories. Performance categories are listed in the rows of the matrix. Measures are listed in the columns. You will type your numeric ratings into each of the cells.

Specific Instructions

Please follow these steps to make your ratings:

1. Scan through the descriptions of the measures to get an idea of the type and level of information available about each one.

2. Read the list of performance categories carefully.

3. Now start with the first measure and read the information. Consider the first performance category. To what extent does the first measure "measure" the first performance category? Use the following scale to quantify your judgment:

0	1	2	3	4	5	6	7	8
This perf. cateogry is not at all measurable by the criterion measure (it is almost useless)			This perf. category is measured partly by the criterion measure (it is of some use)				This perf. category is entirely measured by the criterion measure (it is very useful)	

Factors to Consider in Making Your Extent-of-Measurement Judgments

What does the measure "measure"? The description of the measure and psychometric information are intended to help you better understand what the measure actually measures.

What if there isn't much information about a measure? The amount of information available for each measure varies greatly depending on whether it is an operational, experimental, or proposed measure. Your job as an expert judge will be to make the best judgment you can given the amount of available information and your expertise with the performance categories.

4. Make your judgment for each measure in the same manner; you will be completing one COLUMN AT A TIME. You will probably want to save your work at various intervals while you work. You should also make a back-up copy for safety.

5. When you have completed your ratings, save the file under the same name (your initials plus "rat.wb1). Give your copy of the file to Teresa or Jennifer.

Appendix L
Mission Performance Expert Judgment Exercise Instructions

Importance Ratings of Performance Categories to SF Missions

Instructions

As you know, the SF job analysis resulted in job performance categories (such as *Teaching Others*, *Building Effective Relationships with Indigenous Populations*, and *Decision Making*). When SF and SWC personnel have looked at those performance categories, some have noted that the categories are not equally important and that the importance of each performance category depends on the mission. For example, *Building Effective Relationships with Indigenous Populations*, is more important for some of SF's primary missions than it is for others.

The purpose of this survey is to gather judgements about the importance of the performance categories for SF's primary missions. It will take 15 to 20 minutes of your time. The ratings you and other raters provide will be used to develop weights for the performance categories according to the five primary missions. If you have any questions please call Teresa Russell at (703) 706-5666. Your input is greatly appreciated.

Please follow these steps to make your judgments:

(1) Read the definitions provided below of the performance categories provided on the following three pages.

(2) Consider the first performance category, *A. Teaching Others*; also consider what is required on the first type of SF mission, FID. How important is *Teaching Others* for the effective accomplishment of a FID mission?

(3) Record the importance rating that best represents how important you believe *Teaching Others* is for accomplishing a FID mission.

(4) Consider the next type of SF mission, UW, how important is *Teaching Others* for the effective accomplishment of a UW mission?

(5) Consider each SF mission in turn regarding how important the first performance category is for completing these missions, until all the missions have been rated.

(6) Next, consider the second performance category, *Building and Maintaining Effective Relationships with Indigenous Populations*, and make ratings in the same manner, and so on for each performance category.

Privacy Act Statement - This is an experimental personnel data collection activity conducted by the U.S. Army Research Institute for the Behavioral and Social Sciences pursuant to its research mission as prescribed in AR 70-1. When identifiers (e.g., name) are requested, they are to be used for administrative and statistical control purposes only. Full confidentiality of the responses will be maintained in the processing of these data. Although your participation is voluntary, we encourage you to provide complete and accurate information in the interests of the research. There will be no effect on you for not providing all or any part of the information.

(7) Return the rating form to Teresa Russell (HumRRO, 66 Canal Center Plaza, Suite 400, Alexandria VA 22314), in the enclosed stamped envelope. Thank you for your time and participation in this phase of data collection.

In the boxes below each SF mission please rate how important each of the performance categories are for effective performance of the mission using the following rating scale:

How important is this performance category for the effective accomplishment of this SF Mission (i.e., FID, UW, DA, CT, SR)?

1 = Unimportant

2 = Minor Importance

3 = Important

4 = Very Important

5 = Extremely Important

Definitions of 21 Performance Categories:

A. **Teaching Others**. Conveying knowledge and skill to others; developing POI and tailoring material to the target audience's needs and capabilities; obtaining audience interest and involvement; presenting material in an orderly fashion; using handouts, aids, or tools; finding appropriate ways around language barriers; demonstrating own proficiency.

B. **Building and Maintaining Effective Relationships with Indigenous Populations**. Demonstrating respect for and engaging in behavior appropriate to indigenous culture, values, and customs; providing services and assistance to develop rapport with indigenous people and build respect for SF.

C. **Handling Interpersonal Situations**. Dealing with others constructively, persuading rather than forcing own way; remaining composed, even when provoked; using non-verbal communication skills to interpret behaviors; resolving disputes; allowing others to "win" confrontations.

D. **Using and Enhancing Language Skills**. Using foreign language skills to communicate with Host Nation/Guerilla (HN/G) or other foreign personnel; practicing and developing language skills.

E. **Contributing to the Team Effort and Morale**. Motivating others; communicating effectively with team members; enhancing new and existing team members' skills and readiness; building team spirit through personal interactions.

F. **Showing Initiative and Extra Effort**. Putting forth the effort to produce high-quality work in a timely fashion; actively pursuing self-improvement goals; volunteering for demanding tasks or extra responsibility; taking initiative; presenting a positive image of SF.

G. **Displaying Honesty and Integrity**. Adhering to laws or rules of conduct; knowing when to put aside personal beliefs to follow policy requirements/SOPs, but taking a more difficult, morally correct course of action when appropriate; owning up to own mistakes; being truthful and genuine with others.

H. **Planning and Preparing for Missions**. Developing mission plans that are technically sound, well-coordinated, and likely to lead to mission accomplishment; obtaining complete information needed for planning; drawing on team members' experiences; anticipating enemy movement or other obstacles; weighing alternative courses of action; determining and preparing resources needed for mission accomplishment.

I. **Decision Making.** Assessing the situation and determining an appropriate course of action within a reasonable time frame; digesting information and drawing conclusions; using time, personnel, equipment, and tactics effectively; acting swiftly and decisively when needed; remaining level-headed and task-oriented in stressful situations.

J. **Confronting Physical and Environmental Challenges**. Defeating odds and environment to survive an ordeal; maintaining team standard of performance in physically challenging situations; preparing physically for challenge; following field survival guidance; taking steps to ensure own health and endurance.

K. **Navigating in the Field**. Maintaining correct direction of movement in diverse/demanding conditions; orienting self/team members using navigational aids and terrain features; noticing and taking into account map or environmental details to aid in navigating.

L. **Troubleshooting and Solving Problems**. Thinking of alternative ways to solve a problem; using the resources at hand to fabricate needed items; improvising from own technical knowledge of mechanical and electrical principles.

M. **Being Safety Conscious**. Being alert to safety at all times; rigorously following safety guidelines and instructions for weapons/explosives or other hazardous materials; monitoring others to ensure compliance with SOP when using weapons/dangerous equipment; being alert to potential threat; maintaining noise/light discipline.

N. **Administering First Aid and Treating Casualties**. Applying emergency life-saving techniques and skills when accidents or injuries occur; treating ailments/conditions caused by the environment; following SOP for treating conditions and injuries.

O. **Managing Administrative Duties**. Keeping accurate, up-to-date, organized records; processing paperwork in a timely fashion; establishing SOP; obtaining and ensuring maintenance of supplies and equipment; coordinating with others to share resources or

work on projects; finding the source of administrative problems; using computers; handling classified materials.

P. **Weapons Skills.** Operating and maintaining direct-fire weapons; loading, disassembling, assembling, clearing, reducing stoppage in weapons; emplacing, laying, and aligning mortars and their ammunition; executing FDC procedures.

Q. **Engineering Skills.** Emplacing mines or charges in appropriate area(s); using firing systems correctly and clearing misfires appropriately, electric and non-electric; improving the environment of operations through construction; building necessary structures; using rigging devices; overseeing construction.

R. **Communications Skills.** Planning and preparing communication requirements; following SOP in communication procedures; using cryptic message format to send and receive messages; coordinating communication efforts; configuring and operating equipment, using knowledge of equipment; managing equipment problems.

S. **Medic Skills.** Obtaining medical records and treatment histories and using this in prescribing/administering medications; investigating and evaluating symptoms; performing or assisting doctor in surgical procedures; conducting laboratory tests; treating and monitoring patients; testing and monitoring environmental conditions; providing guidance to HN in preventive health.

T. **Team Leader Skills.** Noticing when subordinates are experiencing personal problems or are demoralized or injured; listening; uplifting others; taking the time and effort to research and correct subordinates' problems (e.g., problems receiving mail while on deployment); establishing a direction; defining tasks clearly; setting specific, challenging, but attainable goals; giving praise when due and discipline as appropriate.

U. **Intelligence Skills.** Planning and directing intelligence collection, analysis, and dissemination; preparing area studies; conducting interrogation and briefing/debriefing patrols.

Importance of Performance Categories for SF Missions

Level of Importance How important is the performance category for the effective accomplishment of this SF Mission?

1 = Unimportant
2 = Minor Importance
3 = Important
4 = Very Important
5 = Extremely Important

	Primary Special Forces Missions				
Performance Categories	1. FID	2. UW	3. DA	4. CT	5. SR
A. Teaching Others					
B. Relations with indigenous people					
C. Interpersonal Situations					
D. Enhancing Language Skills					
E. Team Effort and Morale					
F. Initiative and Extra Effort					
G. Honesty and Integrity					
H. Planning and Preparing					
I. Decision Making					
J. Physical and Environmental Challenges					
K. Navigating in the Field					
L. Troubleshooting and Solving problems					
M. Safety Conscious					
N. First Aid and Treating Casualties					
O. Administrative Duties					
P. Weapons Skills					
Q. Engineering Skills					
R. Communications Skills					
S. Medic Skills					
T. Team Leader Skills					
U. Intelligence Skills					

Appendix M
Recommendations for the Development of the SF Biographical, Interest, and Temperament Survey (SF BITS)

Recommendations for the Development of the SF Biographical, Interest, and Temperament Survey (SF BITS)

Measures and Scales	SF Job Analysis Attribute (Mean Extent of Measurement Rating from 0 to 8)
Army Biodata Inventory:	
Academic Performance	Math Ability, Writing Ability, Achievement and Effort (3.64)
Formal Leadership	Leadership (5.36), Enterprising Interests (4.09), Achievement and Effort (4.00), Motivating Others (4.00), Team Playership (3.64), Supervising (3.45)
Ruggedness	Interest in Adventure and Outdoor Activities (6.55), Physical Fitness and Military Bearing (3.91)
Mechanical Activities	Mechanical Ability (5.55)
Work Experience	Initiative (3.18)
Nondelinquency	Personal Discipline (5.27), Dependability (4.45), Maturity (3.91), Moral Courage (3.36)
Team Sports/Group Orientation	Team Playership (5.09)
Ranger Biodata Inventory:	
Cognition Under Stress	Adaptability (3.36), Maturity (3.09)
Mature Team Commitment	Team Playership (5.55), Leadership (4.36), Dependability (3.36), Motivating Others (3.18)
Self Esteem	Autonomy (4.09) Maturity (3.36)
Need for Achievement	Achievement and Effort (5.82), Initiative (5.64), Perseverance (5.00), Enterprising Interests (3.00)
Outdoor Orientation	Interest in Adventure and Outdoor Activities (6.82)
Physical Endurance	Physical Endurance (6.55), Physical Fitness and Military Bearing (3.73)
Physical Strength	Physical Strength (6.36)
Object Belief	Team Playership (3.91), Interest in People (4.00)
New Biodata Items:	
Family History	Adaptability
Cross-Cultural Sensitivity	Cultural and Interpersonal Adaptability, Interest in Other Cultures, Interest in People
Forced-Choice Assessment of Background and Life Experiences (FCABLE):	
Work Orientation	Initiative (5.64), Perseverance (5.09), Achievement and Effort (4.64)
Dominance	Leadership (5.36), Enterprising Interests (5.00), Persuasiveness/Diplomacy (4.55), Motivating Others (3.64), Supervising (3.27)
Dependability	Dependability (6.73), Personal Discipline (3.82), Maturity (3.73)
Agreeableness	Team Playership (4.45), Interest in People (3.55)
Emotional Stability	Maturity (5.45)
Army Vocational and Occupational Interest Career Examination (AVOICE):	
Rugged/Outdoors	Interest in Adventure and Outdoor Activities (6.64)
Skilled Technical	Interest in Skilled Trades (4.73)
Structural/Machines	Interest in Skilled Trades (4.73)
Interpersonal	Interest in People (3.91)
Job Orientation Blank:	
Autonomy	Autonomy (6.00)
Organizational Identity:	Team Playership (3.00)

Recommendations for the Development of the SF Biographical, Interest, and Temperament Survey (SF BITS)

SF BITS will have three parts:

(1) Biographical- merged items from the Army Biodata Inventory (ABI) and the Ranger Biodata Inventory (RBI) along with new items targeted at family history and cross-cultural sensitivity

(2) Temperament- the Forced-Choice Assessment of Background and Life Experiences (FCABLE)

(3) Interests and Preferences- scales from the Army Vocational And Occupational Interest Career Examination (AVOICE), Job Orientation Blank (JOB), and Organizational Identity

If all of the instruments constituting SF BITS were administered in their entirety. They would take about two hours of time. The target time frame is 30-35 minutes for each part to total 1 and 1/2 hours.

Recommended Development Steps:

(1) Use the FCABLE as is. It currently takes about 30 minutes to complete.

(2) Use the interest and preference measures "as is." Together they take about 30 minutes to complete.

(3) Eliminate verbatim redundancies in ABI and RBI items.

(4) Consider the items on ABI Ruggedness and RBI Outdoor Orientation, eliminate items that are conceptually redundant. Form a merged scale.

(5) Consider the items on ABI Team Sports/Group Orientation and RBI Mature Team Commitment; eliminate items that are conceptually redundant. Form a merged scale.

(6) If there are still too many items, consider dropping ABI Academic Performance and ABI Mechanical Activities. Those constructs are also measured by cognitive test scores.

(7) Write 8-10 items for two new scales, Family History and Cross-Cultural Sensitivity. For Cross-Cultural Sensitivity, draw on items that have been tested on peacekeeping troops who deal with indigenous peoples. Family History items should tap whether the family moved frequently and the type of hardships individuals were exposed to as children.

(8) Present results of steps 1-7 to Len White, Fred Mael, and Bob Kilcullen.

(9) Make final revisions based on their comments.