LEVEL II

ARI TECHNICAL REPORT
TR-78-B11

AD A077999

EVALUATION OF LEADERSHIP AT THE FORT BENNING ASSESSMENT CENTER

by

Kay H. Smith
3273 North Shadowbrook Circle
Provo, Utah 84601

November 1978

Contract DAHC 04-72-A-0001 New for this source

Prepared for:

ARI Field Unit, Fort Benning, Georgia
Personnel & Training Research
Laboratory, ARI

DDC FILE COPY

DDC
RECEIVED
DEC 12 1979
D

U.S. ARMY RESEARCH INSTITUTE
for the BEHAVIORAL and SOCIAL SCIENCES
5001 Eisenhower Avenue
Alexandria, Virginia 22333

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

408010

79 22 5 092

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

JOSEPH ZEIDNER
Technical Director

WILLIAM L. HAUSER
Colonel, US Army
Commander

Research accomplished
under contract to the Department of the Army

DAHC 04-72-A-0001
Task 75-222

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-P, 5001 Eisenhower Avenue, Alexandria, Virginia 22333.

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.

ARI TECHNICAL REPORT
TR-78-B-11

EVALUATION OF LEADERSHIP AT THE FORT BENNING ASSESSMENT CENTER

by

Kay H. Smith
3273 North Shadowbrook Circle
Provo, Utah 84601

November 1978

Contract DAHC 04-72-0001

Prepared for:

ARI Field Unit, Fort Benning, Georgia
Personnel & Training Research
Laboratory, ARI

Accession For
NTIS GRA&I
DDC TAB
Unannounced
Justification
By
Distribution/
Availability Codes
Dist. | Avail and/or special
A

U.S. ARMY RESEARCH INSTITUTE
for the BEHAVIORAL and SOCIAL SCIENCES
5001 Eisenhower Avenue
Alexandria, Virginia 22333

Army Project Number 2Q163101A755

This is an advanced development research project designed to meet military management requirements for research bearing on a specific management problem. A limited distribution is made--primarily to the operating agencies directly involved.

PREFACE

Previous research by the Army Research Institute for the Behavioral and Social Sciences (ARI) provided reliable and valid techniques and methods for assessment of leadership potential and performance. The experimental Officer Evaluation Center at Fort McClellan in 1963-65 demonstrated that valid factors of leadership behavior, notably Combat Leadership and Technical-Managerial Leadership, could be assessed in a 3-day sequence of tasks in a simulated combat environment.

A survey of possible applications of behavioral science to military problems resulted in a recommendation by the Office of the Special Assistant for Training, U.S. Army, to establish a pilot assessment center at the U.S. Army Infantry School at Fort Benning, applying assessment techniques to personnel from three Infantry School courses. R&D support for development of the assessment exercises, training of assessors, and evaluation of effectiveness of the assessment process was delegated to the ARI Field Unit at Fort Benning. The Assessment Center completed its mission in December 1974.

The present report describes assessment processes used in the Center and analysis of initial results. Follow-on work by ARI will provide a validation of the assessment concept as applied to junior officers and enlisted men, as well as a continuation of officer career development research.

Many people at ARI and the Infantry School cooperated in the research reported herein and in preparation of this report. Dr. William H. Helme of ARI, Washington has provided invaluable guidance and support from the project's inception. Special thanks are due to Mr. Richard Hilligoss of ARI, Fort Benning, and to Dr. James Caviness, who were deeply involved in the research at all stages. The contribution of Col. Wallace Veaudry and Majors John Campbell and Richard Davis of the USAIS Assessment Center is also gratefully acknowledged; without them research would have been impossible.

EVALUATION OF LEADERSHIP AT THE FORT BENNING ASSESSMENT CENTER

CONTENTS

Chapter		Page
1.	EXECUTIVE SUMMARY	1
2.	INTRODUCTION	3
3.	DEVELOPMENT OF THE ASSESSMENT CENTER	5
	Behavioral Dimensions	5
	Exercises and Questionnaires	5
	Psychometric Tests	6
	Training of Assessors and Counselors	9
	Data Processing	10
	Pilot Testing	10
4.	OPERATION OF THE ASSESSMENT CENTER	11
	Assessment Scheduling	11
	Data Processing	12
	Feedback Counseling	12
	Assessment Samples	13
5.	THE VALIDATION AND FOLLOW-UP RESEARCH PROGRAM	15
6.	ASSESSEES' SELF-EVALUATIONS AND REACTIONS TO ASSESSMENT PROCESS	17
	Self-Evaluation of Skill Levels	17
	Evaluation of the Assessment Process	18
7.	RELIABILITY OF ASSESSMENT MEASURES	31
	Non-Interactive Exercises	31
	Appraisal Interview	32
	Conglomerate Exercise	32
	Leaderless Group Discussion	34
	Assigned Leader Group Exercise	34
	Leader War Game	34
	Summary	39
8.	THE MODIFIED ASSESSMENT PROGRAM	41
	Brief Description of Modified Assessment Program	41
	Reliability Measures from MAP Pilot Test	42

CONTENTS (continued)

Chapter		Page
9.	INTERNAL STRUCTURE OF THE ASSESSMENT EXERCISES	51
	Entry Interview	51
	In-Basket Exercise	53
	Writing Exercise	53
	Simulate Exercise Global Ratings	55
	Appraisal Interview	57
	Leaderless Group Discussion	57
	Leaderless Group Discussion--IOAC Officers	58
	Assigned Leader Group Exercise	60
	Leader War Game--IOAC Officers	63
	Conglomerate Exercise	65
	Paper and Pencil Tests	65
10.	DIMENSIONAL STRUCTURE OF THE OVERALL ASSESSMENT PERFORMANCE OF BIOCC CANDIDATES	71
	Factors Primarily Defined by a Single Exercise	71
	Factors Defined by Substantial Loadings on Two or More Exercises	75
	Factors Primarily Defined by Paper-and-Pencil Tests	83
	Summary	87
11.	RELATIONSHIP OF MEASURED DIMENSIONS TO MEASUREMENT GOALS	89
	Summary	93
APPENDIX		95

TABLES

Table 1.	Performance Exercises Used at U. S. Army Infantry School Assessment Center	7
2.	Standardized Paper-and-Pencil Tests Used at U. S. Army Infantry School Assessment Center	9
3.	Self-ratings on Six Skill Areas given by various assessment groups	19
4.	Self-ratings on Six Skill Areas by Two Captain Subgroups	20
5.	Self-rankings of Relative Performance in Six Skill Areas Given by Various Assessment Groups	21

TABLES (continued) Page

Table 6. Self-rankings of Relative Performance in Six Skill Areas Given by Two Captain Subgroups 22

7. Ratings of Overall Conduct of the Assessment and of Interest in Feedback by Assessment Groups 23

8. Evaluation of Individual Exercises by Assessment Groups 25

9. Evaluation of Interactive Exercises by Assessment Groups 26

10. Evaluation of Individual Exercises by two Captain Subgroups 28

11. Evaluations of Interactive Exercises by Two Captain Subgroups 29

12. Inter-rater Reliability of Assessor Ratings on the BIOCC Conglomerate Exercise 33

13. Inter-rater Reliability of Assessor Ratings on the BIOCC Leaderless Group Discussion 35

14. Inter-rater Reliability of Assessor Ratings on the IOAC and 197th Inf. Bde. Leaderless Group Discussion 36

15. Inter-rater Reliability of Assessor Ratings on the BIOCC Assigned Leader Group Exercise 37

16. Inter-rater Reliability of Assessor Ratings on IOAC, 197th Inf. Bde. Leader War Game 38

17. Inter-rater Reliability of Scales Used in the Modified Assessment Program Leaderless Group Discussion 43

18. Inter-rater Reliability of Scales Used in the Modified Assessment Program Assigned Leader Exercise 45

19. Inter-rater Reliability of 11 Assessors' Ratings of Two Video-tape Recorded Assigned Leader Exercises in the Modified Assessment Program 46

20. Inter-rater Reliability of Scales Used in Modified Assessment Program In-basket Exercise 49

TABLES (continued) Page

Table 21. Inter-rater Reliability of Scales used in modified assessment program writing exercise 50

22. Factor analysis of Entry Interview for BIOCC Candidates 52

23. Factor Analysis of In-basket Exercise for BIOCC Candidates 54

24. Factor Analysis of Writing Exercise for BIOCC Candidates 55

25. Factor Analysis of Global Ratings by Controller/Assessors in Simulate Exercise for BIOCC Candidates 56

26. Factor Analysis of Appraisal Interview Exercise for BIOCC Candidates 58

27. Factor Analysis of Leaderless Group Discussion Exercise for BIOCC Candidates 59

28. Factor Analysis of Leaderless Group Discussion Exercise for IOAC Officers 61

29. Factor Analysis of Assigned Leader Group Exercise for BIOCC Candidates 62

30. Factor Analysis of Leader War Game Exercise for IOAC Officers 64

31. Factor Analysis of Conglomerate Exercise for BIOCC Candidates 66

32. Factor Analysis of Paper-and-Pencil Tests for BIOCC Candidates 67

33. Factors from Overall Analysis Defined Primarily by Loadings From a Single Exercise 72

34. Factors from Overall Analysis Defined by Substantial Loadings from Two or More Exercises 76

35. Factors from Overall Analysis Defined Primarily by Loadings from Paper-and-Pencil Tests. 84

EVALUATION OF LEADERSHIP AT THE FORT BENNING ASSESSMENT CENTER

Chapter 1

EXECUTIVE SUMMARY

Early in FY 1973 the Fort Benning Assessment Center was established as a research and demonstration project. It was jointly sponsored by the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) and the Office of the Special Assistant for Training (OSAT). In order to implement the project, ARI established a field unit at Fort Benning staffed by three research psychologists. A military unit was established at the directorate level within the U.S. Army Infantry School (USAIS) with 21 officers and enlisted men. Chief of the ARI field unit was Dr. Kay Smith. Col. Wallace F. Veaudry was director of the military unit.

Although the full complement of personnel did not arrive until March of 1973, the assessment process was developed and assessment began, as scheduled, in July of 1973. The assessment process was an intensive three-and-one-half day experience during which the assessee took a battery of standardized paper and pencil tests and participated in three group exercises, an intensive interview, a combat simulate, an appraisal exercise (in which the assessee acted as interviewer), an in-basket exercise, and a writing exercise.

The research project had two major purposes. One was to determine the validity of the assessment process for predicting performance in Officer Candidate School and also in junior officer assignments. For this purpose 143 Branch Immaterial Officer Candidate Course (BIOCC) candidates were assessed just prior to beginning their BIOCC training. Ninety-six newly commissioned second lieutenants were also assessed immediately prior to entering the Infantry Officer Basic Course (IOBC) at the USAIS.

The second purpose of the Assessment Center was to test the value of assessment and counseling as a personal and career development aid. For this purpose 88 Infantry Officer Advanced Course (IOAC) captains and 87 Advanced NonCommissioned Officer Educational System (ANCOES) NCO's were assessed immediately prior to entering their respective courses at the USAIS. Some ten days to two weeks following assessment, these men returned to the Assessment Center for an intensive three to four hour feedback and counseling session. In this session they were given the results of their assessment performance including videotaped examples of their performance in various exercises. In addition, various means they might use to improve on their weaknesses and to capitalize on their strengths were discussed.

To test the value of assessment in an operational setting, 35 company commanders from the 197th Infantry Brigade were also assessed and counseled. In addition, 45 volunteers from ROTC units in the Georgia-Alabama area were assessed during the operation of the Assessment Center in order to determine the applicability of the assessment exercises to an ROTC population.

The analysis of Assessment Center measurements reported herein yielded the following results.

Reliability. The Reliability of various observational ratings was adequate. In some cases, the reliability was remarkably high, considering the difficulty of behavioral observation and evaluation.

Behavioral Dimensions Measured. The eight exercises were each analysed to determine the dimensions of behavior measured. The paper-and-pencil tests were similarly analysed as a group. These analyses revealed that some 62 separate dimensions had been measurd in the individual exercises. These 62 measures were then analysed. This overall analysis revealed that 21 separate dimensions of behavior had been assessed.

Comparison of the 21 dimensions of behavior actually measured at the Assessment Center with the 12 characteristics set up as a priori goals of measurement revealed that the measurement aims had largely been accomplished.

Dimensional analysis showed the strength of the Assessment Center process for measuring communication skills, forcefulness, social skills, and the application of mental ability. These characteristics are not readily measured, with accuracy, in any other known way.

Validity. Later reports will show the relationship of attributes measured at the Center to performance in school and on duty assignment. They will also outline the short and long term effects of development counseling.

Chapter 2

INTRODUCTION

In the past decade, considerable advances have been made in the assessment of those elusive skills, attitudes, and behavior patterns which comprise the general concept of leadership. Since even the most sophisticated of the standard psychological tests have fallen short of satisfactory assessment of these complex skills, industrial firms have increasingly relied on a combination of interpersonal exercises and individual work samples administered in an assesssment center. Today many of the largest corporations, and smaller organizations as well, are using assessment centers successfully for selection of leaders, all the way from first line supervisors to top executives.

Military applications of assessment centers date back to a period prior to World War II. The U.S. Army Research Institute for Behavioral and Social Sciences (ARI) and its predecessor organizations pioneered in the development of sophisticated simulations for measurement of combat leadership in the early 1960s. (See Willemin, Louis P. Prediction of Officer Performance. Technical Research Report 1134, 1964. New Report Nos. 1172, 1173, and 1182.) The British Army has relied for many years almost exclusively upon asssessment centers for selection of men for training programs leading to a commission.

A few companies and other organizations have used assessment centers as leadership training and development tools. Notable among these is the Center for Creative Leadership operated by the Richardson Foundation in Greensboro, North Carolina. Here intensive assessment is followed by detailed feedback and personal counseling along with training in leadership skills. Two groups of Army officers--12 Lieutenant Colonels, and 12 Brigadier General Designees--participated in this program and found it to be of considerable benefit to themselves as leaders.

Objectives. In 1972 the Office of the Special Assistant for Training (OSAT) and ARI jointly planned the Assessment Center Demonstration Project. This project was not designed to test the validity of assessment centers for measurement of leadership skills. This validity has been amply demonstrated in industrial assessment centers, notably in the American Telephone and Telegraph Co., Standard Oil of Ohio, General Electric and J. C. Penney. The feasibility of measuring military leadership was shown in the ARI experimental project at Fort McClelland. The task of the Assessment Center Demonstration Project was to further the development and cost effectiveness of assessment-center technology as it applies to Army selection and training. Specific goals were:

1. Development of performance tests which are military in nature, have face validity for the military assessee, and are directly relevant to real-life tasks of the military leader.

2. Standardization and objectification of exercises and measurement devices to reduce training demands on assessors and to make assessment results interchangeable from time to time and place to place.

3. Rationalization of scoring and score combination procedures to reduce the necessity for intuitive judgment in dimensions of behavior.

4. Development of counseling and feedback procedures which are effective in providing self-insight, yet can be implemented by military men who are not professionals in behavioral science.

If it is determined that these goals were achieved, the Army will have at its disposal a powerful set of new tools for selection and development of leaders.

Project Phases. The Assessment Center Demonstration Project was designed as a three-phase program.

Phase 1. Development, including all work necessary to produce an operational assessment center. Development was completed as scheduled during the second half of FY 73.

Phase 2. Operation. The center operated during FY 74 and the first half of FY 75 assessing samples of the original target populations--captains from IOAC classes, candidates from BIOCC classes, and NCOs from ANCOES classes. During the development phase, the Center assumed the additional tasks of assessing samples of 2nd lieutenants from IOBC classes and ROTC cadets.

Phase 3. Follow-up. ARI scientists are obtaining criterion performance data from the Infantry School and from the units to which assessees, and comparable control ratees are assigned, for at least a two-year period following assessment.

This report includes a description of how the Assessment Center was developed and operated. The overall research plan is described and data given on assessees' reactions to assessment, reliability of measurement, and the dimensions being measured as revealed by factor analysis. Later reports will describe the results of validity and follow-up research. Details of development and operation of the Assessment Center as well as actual exercise materials and measurement devices can be found in the Assessment Center After Action Report published by the United States Army Infantry School.

The reliability and dimensionality analyses discussed in chapters seven, nine, and ten were conducted for the most part with the BIOCC group. The reason for this was that this group, with 143 assessees, provided the most adequate data base. In some cases where exercises were substantially different, data were analyzed for the IOAC group as well. The reliability data given in chapter eight were computed for the 72-man Modified Assessment Program pilot test. These men were not part of the 492 man sample assessed in the original Assessment Center research program, and described in chapter four.

Chapter 3

DEVELOPMENT OF THE ASSESSMENT CENTER

The developmental sub-phases of activity at the Assessment Center were directed towards actualizing the plans prepared by OSAT and ARI. The developmental sub-phases, conducted during the second half of FY 72, consisted of 1) selection of behavioral dimensions to be assessed, 2) development of assessment exercises and questionnaires, 3) selection of psychometric tests, 4) training of staff in assessing and counseling, 5) development of data processing techniques and 6) pilot testing the assessment process.

BEHAVIORAL DIMENSIONS

Job functional leadership behavior was systematically analyzed into component knowledges, skills, and personality traits, with the goal of explicating those dimensions of behavior that are the mark of superior leading. Negative dimensions that marked failure or inferiority were not stressed. The dimensions that surfaced as a result of systems analysis were reviewed to determine the practicability and practicality of objective measurement. Throughout this sub-phase, intuitive, empirical, and consensual verification of the dimensions was consistently sought. The resulting list is similar in nature to the products of other research into leadership dimensions such as projects conducted by Development Dimensions, Inc., and HumRRO.

The twelve dimensions of leader behavior judged appropriate for measurement in the Assessment Center were:

Adaptability	Motivation
Administrative Skills	Organizational Leadership
Communication Skills	Physical Fitness
Decisionmaking Skills	Social Skills
Forcefulness	Supervisory Skills
Mental Ability	Technical and Tactical Competence

EXERCISES AND QUESTIONNAIRES

A review of assessment exercises and questionnaires currently in general use (war games, competitive discussions, etc.) revealed a plethora of material. However, when evaluated against the needs of the Assessment Center, available "off the shelf" exercises were judged less than adequate. In rare cases an exercise could be recast into a military context. In other cases, completely different exercises needed to be developed.

The adaptation, creation, and revision of exercises turned out to be a major challenge, with rigorous demands. The exercises which finally emerged are in Table 1.

The three versions of the Simulate, exercise 8 in Table 1, were produced by HumRRO Division 4 contractors under the direction of Joseph Olmstead as Principal Scientist. Military personnel from the Center as well as members of the ARI field unit worked closely with the HumRRO scientists in the development of these simulates.

In addition to these exercises, three questionnaires were developed. These were designed to be filled out by the assessees at various times during the assessment experience. The first questionnaire provided information descriptive of the background of the individual. The second required a series of associate evaluations (peer rankings) and the final questionnaire was targeted at obtaining reactions to, and measuring acceptance of, the assessment and counseling conducted at the center.

The exercises and questionnaires were pretested as data collection devices. In accordance with lessons learned, instructions were rewritten, data recording sheets were revised, exercises were modified to facilitate interaction, and assessors received additional preparation.

PSYCHOMETRIC TESTS

In selecting standardized psychological tests for inclusion in the program, the primary criterion was relevance of the variables to these dimensions of leadership: administrative skills, communication skills, supervisory skills, forcefulness, adaptability, decisionmaking, and mental ability.

The secondary criteria used in selecting tests were: non-offensiveness of test items, suitability in content and format for mature adults, adequacy of data and theoretical discussions, recency of publication or revision, and shortness of the time required for test administration.

The nine tests selected to provide data for a psychometric profile are listed in Table 2. These tests were all paper and pencil tests, objectively scored.

Psychometric profiles compared the assessee's individual scores with the average scores of his group and with the published norms for the tests. In addition, test scores were used to augment the assessments made in the exercises. Results of the psychometric testing also were used in the feedback counseling sessions.

Table 1
(Page 1 of 2)

PERFORMANCE EXERCISES USED AT US ARMY INFANTRY SCHOOL ASSESSMENT CENTER

Group exercises

1. Leaderless Group Discussion (LGD). Required each participant to attempt to persuade the other members of 6-man group to decide in favor of the position he represents while also acting as a responsible group member in helping the group reach the best overall decision. Two versions: A. Resource allocation discussion--used with IOAC captains, B. Soldier-of-the-Month selection board--used with all other groups. Three raters, 9 rating scales, 6 peer ranking scales.

2. Competitive Stock Trading Exercise (Conglomerate). Three leaderless 6-man teams traded stock, attempting to gain control of conglomerates while blocking other teams from doing so. Three raters, 8 rating scales, 5 peer ranking scales.

3. Rotating Assigned Leadership Exercise. Two versions: A. Assigned Leader Group Exercise (ALGE). Used with all groups except captains and some IOBC lieutenants. Outdoor exercise with 6 lanes. A different member of each 6-man group was assigned as leader for each lane. Each lane had an obstacle to be surmounted and a task to be performed. Three raters, 9 rating scales, 4 peer ranking scales. B. Leader War Game. Used with captains and some IOBC lieutenants. The game was played in six periods with a different member of a 6-man group as leader in each period. The groups allocated budget for arms purchase, modification, and deployment against an enemy group. Three raters, 9 rating scales, 5 peer ranking scales.

Interpersonal interaction--one-on-one

4. Entry Interview (EI). Each man was interviewed on the morning of his first full day. The interview was aimed at assessing his attitudes toward himself and his environment as well as his composure and style in handling such a communication situation. One rater, 14 rating scales.

5. Appraisal Interview (AI). Each man was required to plan and conduct two interviews to appraise the qualifications of candidates for a specified assignment in his unit. Persons interviewed were also assessees. Each man interviewed twice and was interviewed twice. His written plan and selection decision were graded along with video tapes of his interview. Two raters, 8 rating scales.

Table 1
(page 2 of 2)

Work Samples.

6. In-Basket (IB). Each man was assigned to role play a newly assigned leader. He was given three hours to react to the items in a full in-basket. He was required to write letters, plan meetings, delegate actions, etc. His written work was thoroughly reviewed and partially graded. Grading was completed after he was interviewed on his reasons and motives for actions taken. Three versions: A. New first sergeant--used with NCOs, B. New company commander, used with BIOCC, IOBC, and ROTC, C. New battalion commander, used with IOAC. One rater, 15 rating scales.

7. Writing Exercise (WE). Each man was required to write a report based on information given to him. Three versions: A. Staff action paper--used with IOAC, B. Officer's statement on discharge action--used with BIOCC, IOBC, and ROTC, C. Noncommissioned officer's statement on discharge action--used with ANCOES. One rater, 4 rating scales.

Leadership in a simulated situation.

8. Simulate. Each man led a "group" of men in a simulated organization under stressful conditions. Combat conditions were simulated for NCOs and captains, who had considerable Army experience, while IOBC, BIOCC, and ROTC assessees were required to lead in a civilian emergency. The assessee received instructions from and reported to an assessor who role-played his superiors. The assessee gave orders to and received information from another assessor who role-played the assessee's subordinates. Three versions: (A) Company commander, combat simulation--used with IOAC, (B) Temporary platoon leader, combat simulation--used with ANCOES, (C) Platoon leader, civilian emergency simulate--used with BIOCC, IOBC, ROTC. Two raters, 11 rating scales. Ratings from the two raters were based to a large extent on separate observations of behavior 100-130 performance ratings on specific simulate behaviors.

Physical Training.

9. Physical Training Exercise. This was Army Conditioning Drill One, with a mile run replacing the run-in-place. It was given early each morning to BIOCC groups only. Three raters, 5 rating scales.

Table 2

STANDARDIZED PAPER AND PENCIL TESTS USED AT U. S. ARMY INFANTRY SCHOOL ASSESSMENT CENTER

Leadership Opinion Ouestionnaire (LOQ)
2 scores: Structure, Consideration

Watson-Glaser Critical Thinking Appraisal (WG)
1 score

Nelson-Denny Reading Test (ND)
4 scores: Vocabulary, Comprehension, Total, Reading Rate

Henmon-Nelson Test of Mental Ability (HN)
3 scores: Verbal, Ouantitative, Total

Leadership O-Sort Test (LQS)
7 scores: Leader Potential (total), Personal Integrity, Consideration, Mental Health, Technical Information, Decisionmaking, Teaching/Communication

Social Insight Test (Chapin) (SI)
1 score

Work Environment Preference Schedule (Gordon) (WEPS)
1 score

Strong Vocational Interest Blank (SVIB)
Administered but scores not analyzed

Edwards Personal Preference Schedule (EPPS)
14 scores: Achievement, Deference, Order, Exhibition, Antonomy, Affiliation, Intraception, Succorance, Dominance, Abasement, Nurturance, Change, Endurance, Aggression

TRAINING OF ASSESSORS AND COUNSELORS

The Assessment Center staff consisted of persons with varying backgrounds and training in the knowledges and skills prerequisite to assessing and counseling. Differences in education, training, and other general background parameters necessitated a comprehensive training program to bring each person within an acceptable range of competence.

Training consisted of formal presentations of principles and techniques of asssessment by members of the ARI Field Unit; an assessment workshop conducted by a consultant affiliated with Development Dimensions, Inc.; a performance counseling workshop conducted by the U. S. Army Human Research Unit from Fort Bliss, Texas; a counseling interview workshop conducted by Baker Consulting Associates; and numerous rehearsals, with analysis and review, of the various exercises as they were developed.

DATA PROCESSING

The final sub-phase of development was the formulation of data processing procedures for recording, reducing, and analyzing test data. These procedures had to be such that scores and comparison group information would be available for use in counseling the assessee shortly after his stay at the Center.

Records of the behavioral data obtained by observing the assessees' performance at the Assessment Center were retained in three modes: the raw data sheets were filed, data were transposed onto IBM punch cards, and a master file was recorded on magnetic tape.

Analyses performed on the data had three functions. First, summary descriptive statistics were computed that showed averages and standard deviations of performance, exercise by exercise, across individuals within one category of assessee (e.g., NCOs from ANCOES). These data established norms.

Second, data were analyzed for reporting on an individual profile sheet. An individual's test scores and dimension ratings were computed, and these, along with comparison indices for his group, were printed out on his profile sheet. This sheet was used in preparing and presenting the feedback counseling.

The third purpose served by analyzing the data was evaluative, not of the assessees, but rather of the assessment program itself, its acceptability, reliability and validity. Some relevant analyses are described in this report.

PILOT TESTING

The assessment process, with its newly developed exercises and questionnaires, its newly assembled and trained staff, and its tentative schedule for administration of the exercises and paper-and-pencil tests had its shakedown cruise in June, July, and August, 1973. Pilot tests were run, utilizing seven samples of assessees, representative of the military leaders to be assessed during the operational phase of the demonstration project.

The experience of pilot testing the individual exercises and combining them into a scheduled program brought the assessment process into sharp focus. Lessons were learned, revisions were made to the exercises, and the schedule was adjusted. In addition, the pilot test runs provided opportunity for the assessors to practice their assessing and counseling skills.

Chapter 4

OPERATION OF THE ASSESSMENT CENTER

In the operational phase of the Assessment Center Demonstration Project, 492 men in 31 groups were assessed and counseled at a rate of approximately 2 groups a month. This phase occupied all of FY 74, and the first half of FY 75.

ASSESSMENT SCHEDULING

The assessment schedules varied for assessees from IOAC, IOBC, and ANCOES classes. However, a typical schedule for one man in an in IOAC class (2-74) conveys the flavor of the "total immersion" assessment schedule:

DAY 1	1530-1600	Welcome and Orientation
	1605-1630	In-processing
	1635-1700	Psychometric Tests I
	1705-1730	Entry Questionnaire
	1730-1825	Evening Meal
	1825-1925	Writing Exercise (staff action paper)
	1930-2055	Appraisal Interview (interviewer)
DAY 2	*0700-0825	Entry Interview
	0830-1130	In-Basket Exercise
	1130-1200	Lunch
	**1200-1555	War Game I
	1600-1700	Psychometric Test II
	1800-1900	Evening Meal
	1900-2100	Conglomerate Exercise
DAY 3	0700-0955	Psychometric Test III
	1000-1130	In-Basket Interview
	1130-1200	Lunch
	1200-1325	Simulate Preparation
	**1330-1625	War Game II
	1630-1730	Psychometric Test IV
	1800-1900	Evening Meal
	1900-1920	Appraisal Interview (as candidate)
	1925-2015	Army War College Leadership Questionnaire
DAY 4	0700-0750	Appraisal Interview (as candidate)
	0800-0900	Psychometric Test V
	0905-1130	Leaderless Group Discussion
	1130-1200	Lunch
	1200-1700	Simulate
	1705-1735	Exit Questionnaire

END

*For BIOCC, a Physical Training Test was given at 0530.

**For IOBC and ANCOES, the Assigned Leader Group Exercise replaced WAR GAME I and II.

This schedule approximated a "total immersion" test situation in which a maximum of information is collected in the shortest time. The schedule also had the effect of fully occupying the assessees and limiting to a minimum the effect of outside influences during the 74-hour test period.

The staff was fully occupied both during and between assessment periods. During an assessment period, the assessors expended over 413 man hours in actually conducting assessment exercises and preparing reports.

Betwen assessment periods, the assessors prepared data for reduction and analysis, and prepared for the counseling sessions when the assessees returned for feedback.

DATA PROCESSING

Data processing began on Day 2, the first full day of assessment. Ratings information from the questionnaire, profiles of psychometric-test scores, and exercise scores were reported to the Computer Group. This group continued to receive data until the day after the last assessment day. All data, except those from the Post-Counseling Questionnaires and Post-Counseling Assessor Report were transcribed to master data sheets. These sheets were taken to the Management Information Systems Office (MISO) for keypunching, recording on a master computer tape, and analyses. With a turn-around time of one day, the information was returned, having been analyzed and reported in graphic percentiles. This information was turned over to the counselors, who utilized it in writing their feedback counseling reports.

FEEDBACK COUNSELING

Counseling was mainly limited to the two leadership development groups, IOAC and ANCOES, although some members of IOBC and ROTC groups were counseled on an experimental basis. The schedule called for each man to receive 3 1/2 hours of individual feedback counseling. The counseling took place over a three-day period, scheduled from 1 to 4 weeks after assessment.

The counseling was given individually, one counselor to one assessee. The counseling session was organized around the list of behavioral dimensions. It began with the assessee's strongest dimension and closed with one of his stronger dimensions. The weakest dimension was sandwiched in the middle of the session. The session was arranged to be both non-threatening and credible. The goal was to allow the assessee to gain a clear picture of himself and his performance at the Assessment Center. One of the most effective counseling helps was a 20-minute video tape showing selected samples of the assessee's performance in Assessment Center exercises.

At the close of the counseling session, the assessee filled out a Post Counseling Questionnaire and the counselor completed a Counselor's Reaction Scale, both targeted at recording the assessee's acceptance of the Assessment Center, including the counseling.

ASSESSMENT SAMPLES

The 492 men who were assessed were distributed among the following samples:

1. IOAC (Infantry Officer's Advanced Course), 88 men assessed in five groups.

2. 197th Infantry Brigade Company Commanders, 33 men assessed in two groups. These men were not part of the research sample. Their assessment and counseling were carried out as a small pilot test of the applicability of assessment and counseling to officers currently serving in command positions.

3. IOBC (Infantry Officer's Basic Course), 96 men assessed in seven groups. Of the seven groups, two contained only six men and one contained only twelve. In the six-man groups there was no opportunity to rotate assignments in the interactive exercises, and the nature of the Conglomerate exercise was changed considerably. Data from these groups are excluded from the dimensionality and validity analyses.

4. BIOCC (Branch Immaterial Officer Candidate Course), 143 men assessed in eight groups.

5. ANCOES (Advanced Non-Commissioned Officer Education System), 87 men assessed in five groups.

6. ROTC, 45 men assessed in four groups. Assessment of these men demonstrated the feasibility of the process as applied to ROTC students. However, since the sample was not random and contained two undersized groups of six men, and was heterogeneous (several colleges with widely varying student bodies were represented, and assessees ranged from first-to third-year students), no attempt was made to analyze the data further.

Chapter 5

THE VALIDATION AND FOLLOW-UP RESEARCH PROGRAM

Validation of the assessment measures and tests of the value of feedback counseling were part of the original research plan. With the BIOCC group, the purpose of assessment was to obtain performance data which would allow prediction of performance in the officer candidate course, as well as performance as a commissioned officer. The testing of IOBC officers had a similar purpose, prediction of performance in infantry officer assignments.

For the two junior officer groups, both proximal and ultimate criteria will be correlated with the assessment dimensions described later in this report. The proximal criteria are performance scores of the assessees obtained from the BIOCC and IOBC courses at the Infantry School. Since the Assessment Center is being considered as a possible selection device, the prediction of BIOCC performance has considerable significance. In the case of both BIOCC candidates and IOBC officers, ratings are being collected periodically for a two-or three-year period to determine the predictive validity of assessment measures for performance in actual officer assignments.

Matched control groups of non-assessed BIOCC and IOBC men who go through the Center will also be examined to determine the possible effects on performance of simply having gone through the assessment process.

For two career groups, IOAC Captains and ANCOES NCOs, the main purpose of assessment was personal development. All assessees in these groups were given intensive feedback and counseling sessions. These sessions included not only information on the assessee's performance at the Center but also a discussion of possible programs which might help the assessee improve on his weak points and capitalize on his strengths.

For these career development groups, measures of post-assessment behavior at the Infantry School and also in subsequent duty assignments are being gathered. Of interest here is the effect of assessment and counseling on school and duty performance as well as on other kinds of behavior. The school and duty performance of a carefully matched group of IOAC and ANCOES classmates who were not assessed and counseled is also being gathered. Data will be sought on personally initiated self-development activities among the two groups. Questionnaires will also seek to discover the long lasting effects of assessment and counseling on self-concept, career aspirations, attitudes toward Army training and career development programs, and the like.

Subsequent reports on this research project will present results of these validation and follow-up efforts.

PRECEDING PAGE NOT FILMED
BLANK

Chapter 6

ASSESSEES' SELF-EVALUATIONS AND REACTIONS TO ASSESSMENT PROCESS

In the entry and exit questionnaires, assessees were asked to make judgments about (1) themselves and their skills in leadership functions, and (2) the assessment center, their interest in feedback on their performance, and the relative value of the various assessment activities.

A complete analysis of all four of the systematically sampled populations follows. The data allow comparison of the response averages and proportions from population to population.

SELF EVALUATION OF SKILL LEVELS

The entry questionnaire asked the assessees to evaluate the personal skills in six general areas, described as follows:

1. Skills in getting along with coworkers.
2. Skills in getting ideas across face-to-face.
3. Skills in getting ideas across in writing.
4. Leadership skills.
5. Skills in dealing with the administrative procedures for getting an outfit set up correctly and keeping it running well.
6. Skills in understanding problems and figuring out answers.

The assessees were asked to make two kinds of judgments concerning their own skills in these six areas. The first judgment involved placing themselves in comparison with others at their own grade and experience level. The assessees could indicate that their own skills placed them in (1) the bottom third, (2) the middle third, (3) the top third but not the top 10%, or (4) the top 10%. The second judgment involved selecting, from among the six areas, their strongest and their weakest skill areas. They then ranked the remaining areas from second to fifth.

Table 3 presents a tabulation of the ratings. All four groups rated themselves highest in getting along with others, lowest in written communication or adminstration. In each skill-group category, the ratings appear to be unrealistically high, 50% to 99% of the raters placing themselves in the upper third of their peers, only 0 to 10% in the lower third. On all six skills, the BIOCC group provided the highest self-ratings. On the average across skills, 57% of this group placed themselves in the upper 10%. For all skills but written communication, the next highest ratings were provided by the ANCOES group. This tendency of the enlisted men to rate themselves higher than did the officers in relation to others of their grade and experience has some justification. Both enlisted groups are select groups. The officer candidates, in particular, represent a small and select percentage of enlisted persons.

PRECEDING PAGE NOT FILMED
BLANK

An additional reason for high self-ratings among the officer candidate may have been a desire to impress any superios's who might be in a position to affect their success in the officer candidate course. Also, the selection of candidates may have favored those who tended to display self-confidence, warranted or not. The relatively low (though more-or-less inflated) self-ratings of the BIOC group may be attributable in part to a substantial proportion low in commitment to an Army career and hence under less pressure to make a good impression.

Table 4 gives the self-ratings on the six skill areas by the two subgroups of Army captains assessed--the active company commanders from the 197th Infantry Brigade and the IOAC students. Few substantial differences are shown between the two groups. There was a fairly consistent tendency for the 197th company commanders to evaluate their skills more highly. This was most pronounced in the area of written communication but was also shown in all other areas but leadership.

Table 5 shows how the various assessment groups evaluated their individual skills in order of merit. Skill in getting along with others was clearly perceived as the strongest by all four groups. In every case more than half the men ranked this skill as first or second. For the ANCOES group the tendency to rank this skill area highly is most marked. Oral communication and leadership are also frequently highly ranked by men in all four groups. At the other end of the scale, written communication and administrative skills tend most often to be ranked in the fifth or sixth position. Overall the BIOCC group shows the least uniformity in assigning ranks to the skill areas, the ANCOES group the most.

Table 6 shows the skill rankings for the two captain groups. Getting along with others tends to be ranked highest, by both groups, with leadership a close second. Written communication and adminitration skills tend to be ranked lowest. There are few substantial differences between the two captain groups. There is a slight tendency for the IOAC group to rank their leadership skills higher and the 197th group is higher for problem solving. On the other end of the scale, the 197th officers tend somewhat more often to denigrate their administrative skills but not so much their leadership and written communication skills.

EVALUATIONS OF THE ASSESSMENT PROCESS

On the exit questionnaire, the assessees in all groups were asked a number of questions designed to evaluate their reactions to individual exercises as well as to the assessment process as a whole.

Table 7 regarding the assessment shows that in all the groups, feelings about how well the assessment process was conducted were overwhelmingly positive. This response is a tribute to the planning and construction efforts that went into development of the exercises, as well as to the concerned, professional efforts of the military assessors.

Table 3

SELF-RATINGS ON SIX SKILL AREAS GIVEN
BY VARIOUS ASSESSMENT GROUPS

Skills	Groups	Self Rating Percentages			
		Bottom	Middle	Top 1/3 but not Top 10%	Top 10%
Skill in getting along with others	IOAC	0	8	35	57
	IOBC	2	15	40	43
	BIOCC	0	1	17	82
	ANCOES	0	2	24	74
Skill in oral communcation	IOAC	1	12	60	29
	IOBC	0	30	45	25
	BIOCC	0	4	35	61
	ANCOES	1	8	45	46
Skill in written communication	IOAC	10	39	34	16
	IOBC	5	28	50	17
	BIOCC	1	11	44	44
	ANCOES	9	38	38	15
Skill in leadership	IOAC	1	9	53	36
	IOBC	3	32	42	23
	BIOCC	1	9	41	49
	ANCOES	1	7	45	47
Skill in Administration	IOAC	2	22	53	23
	IOBC	7	27	52	15
	BIOCC	0	14	39	47
	ANCOES	6	24	49	21
Skill in problem solving	IOAC	0	17	55	28
	IOBC	5	17	52	27
	BIOCC	0	3	38	59
	ANCOES	1	5	55	39

Table 4

SELF-RATINGS ON SIX SKILL AREAS BY
TWO CAPTAIN SUBGROUPS

Skills	Groups	Self Rating Percentages			
		Bottom	Middle	Top 1/3 but not Top 10%	Top 10%
Skill in getting along with others	IOAC	0	8	35	57
	197th	0	3	33	64
	Total	0	7	35	59
Skill in oral communication	IOAC	1	12	60	29
	197th	0	15	46	39
	Total	1	12	55	31
Skill in written	IOAC	10	39	34	16
	197th	3	27	42	27
	Total	8	36	37	19
Skill in leadership	IOAC	1	9	53	36
	197th	3	12	49	36
	Total	2	10	52	36
Skill in administration	IOAC	2	22	53	23
	197th	0	27	39	33
	Total	2	23	49	26
Skill in problem solving	IOAC	0	17	55	28
	197th	0	12	50	38
	Total	0	16	53	31

Table 5

SELF-RANKINGS OF RELATIVE PERFORMANCE IN SIX SKILL AREAS
BY VARIOUS ASSESSMENT GROUPS

Skills	Groups	Skill Ranking Percentages					
		1st	2nd	3rd	4th	5th	6th
Skill in getting along with others	IOAC	44	14	9	8	13	12
	IOBC	42	25	12	8	7	5
	BIOCC	41	12	7	7	13	20
	ANCOES	56	20	7	8	5	2
Skill in oral communication	IOAC	10	20	20	18	22	7
	IOBC	12	18	23	32	5	10
	BIOCC	16	23	13	21	10	17
	ANCOES	12	36	18	18	12	5
Skill in written communication	IOAC	6	8	5	13	21	49
	IOBC	8	12	18	7	23	32
	BIOCC	13	22	11	28	10	16
	ANCOES	1	2	18	8	16	55
Skill in leadership	IOAC	19	31	26	16	3	4
	IOBC	15	13	28	20	15	8
	BIOCC	13	20	16	18	12	21
	ANCOES	15	24	33	20	7	2
Skill in administration	IOAC	10	14	17	16	19	22
	IOBC	7	12	10	10	32	31
	BIOCC	11	15	17	13	25	19
	ANCOES	6	9	35	14	29	7
Skill in problem solving	IOAC	10	14	23	29	21	5
	IOBC	17	20	8	23	28	24
	BIOCC	6	9	35	14	29	7
	ANCOES	10	12	21	34	17	5

Table 6

SELF-RANKINGS OF RELATIVE PERFORMANCE IN SIX SKILL AREAS BY TWO CAPTAIN SUBGROUPS

Skills	Groups	Skill Ranking Percentages					
		1st	2nd	3rd	4th	5th	6th
Skill in getting along with others	IOAC	44	14	9	8	13	12
	197th	36	27	6	18	3	12
	Total	42	17	8	11	10	12
Skill in oral communication	IOAC	10	20	20	18	22	7
	197th	3	28	12	22	15	19
	Total	8	22	18	19	20	10
Skill in written communication	IOAC	6	8	5	13	21	49
	197th	9	9	15	15	12	43
	Total	7	8	7	13	19	48
Skill in leadership	IOAC	19	31	26	16	3	4
	197th	12	18	30	15	18	6
	Total	17	27	27	16	8	4
Skill in administration	IOAC	10	14	17	16	19	22
	197th	18	6	12	12	33	18
	Total	12	12	16	15	23	21
Skill in problem solving	IOAC	10	14	23	29	21	5
	197th	22	12	25	19	19	3
	Total	13	13	23	26	20	4

Table 7

RATINGS OF OVERALL CONDUCT OF THE ASSESSMENT AND OF INTEREST IN FEEDBACK BY ASSESSMENT GROUPS

Rating Scales	Groups				
	IOAC	197th	IOBC	BIOCC	ANCOES
I. Feelings about how the assessment process was conducted					
(4) Very smoothly and professionally conducted	94.3%	96.9%	90.0%	96.5%	88.4%
(3) Generally well conducted but with some lapses	4.5%	3.1%	10.0%	2.8%	11.6%
(2) Some things done well others not	1.1%	0.0%	0.0%	0.7%	0.0%
(1) Generally poorly conducted	0.0%	0.0%	0.0%	0.0%	0.0%
Mean Rating	3.93	3.97	3.90	3.96	3.88
II. Interest in receiving feedback					
(4) A high degree of interest	96.6%	100.0%	93.3%	90.0%	95.3%
(3) A moderate degree of interest	3.4%	0.0%	5.0%	8.6%	3.5%
(2) A low degree of interest	0.0%	0.0%	1.7%	0.7%	1.2%
(1) No interest at all	0.0%	0.0%	0.0%	0.7%	0.0%
Mean Rating	3.97	4.00	3.92	3.88	3.94

Table 7 also gives results of a question on interest in receiving feedback regarding the assessee's performance. It can be seen that interest in feedback ran very high, even in the two groups that had been told from the beginning that they would not be receiving feedback--the IOBC and BIOCC groups.

At one point on the exit questionnaire, all the exercises were listed. The assessee was asked to rate each exercise on its interest, its challenge, how well it was run, its realism, and its overall value. These ratings were made using a 5-point scale, 5 being very good, 4 good, 3 marginal, 2 poor, and 1 very poor. After the assessee had made these ratings, he was asked to indicate which of the exercises he felt was best and which he felt was worst, giving reasons for his choices.

Tables 8 and 9 give the results of the ratings and choices. In Table 8 are the results of exercises which did not involve interaction with other assessees. Thus, the simulate is included here. While the simulate was highly interactive, all the interaction was with role-playing assessors. Table 9 gives the responses for exercises which did involve interaction with other assessees.

Ratings were generally very positive. In almost all cases, the average was above 4, somewhere between good and very good. It is not surprising that the standardized tests fared less well than the exercises. They were tedious and time consuming to a great extent, and lacked the interest and challenge inherent in the other exercises. The only potential reward for the assessee lay in the promise of some interesting information about abilities and personality which might come from the feedback session. Even this gain was missing for the IOBC and BIOCC groups.

The radio simulate was generally preferred as the best exercise. This result is more pronounced for the IOAC and IOBC groups than for the BIOCC and ABCOES groups. The latter two groups showed somewhat stronger preferences than the former two for the In-Basket and the field exercise (Assigned Leader Group Exercise). That there were three different Radio Simulate exercises may account in part for the differences between groups in percent judging Radio Simulate as best or as least moderate. However, two groups, IOBC and BIOCC, went through the same simulate yet differed substantially in percentage of "best" responses.

The different backgrounds of assessees in these two groups and their differing levels of military experience may account for the discrepancy.

The Radio Simulate, the In-Basket, and the assigned leader exercises (Leader War Game for IOAC, and Field Exercise for the other three groups) account for over 80% of the choices for best exercise in all four groups. For the ANCOES group the choices are spread almost evenly among the three exercises. The same is true of the BIOCC group, although they did show a somewhat stronger preference for the Simulate.

Table 8

EVALUATION OF INDIVIDUAL EXERCISES BY ASSESSMENT GROUPS

Rating Scale	Groups	Mean Ratings by Exercise					
		Entry Interview	In-Basket Exercise	In-Basket Interview	Writing Exercise	Radio Simulate	Standard-ized Test
Interest	IOAC	4.63	4.76	4.64	4.40	4.80	3.87
	IOBC	4.28	4.53	4.60	4.04	4.83	2.73
	BIOCC	4.58	4.68	4.76	4.35	4.76	3.50
	ANCOES	4.59	4.86	4.81	4.51	4.86	4.15
Challenge	IOAC	3.97	4.68	4.14	4.76	4.73	4.28
	IOBC	3.57	4.85	4.18	4.21	4.92	3.46
	BIOCC	3.92	4.82	4.29	4.48	4.88	3.93
	ANCOES	4.22	4.88	4.59	4.63	4.83	4.48
How well run	IOAC	4.81	4.80	4.70	4.66	4.77	4.52
	IOBC	4.63	4.75	4.43	4.40	4.80	3.86
	BIOCC	4.71	4.76	4.68	4.61	4.83	4.35
	ANCOES	4.63	4.75	4.81	4.65	4.85	4.25
Realism	IOAC	4.53	4.43	4.44	4.22	4.62	3.95
	IOBC	4.29	4.62	4.25	4.24	4.83	3.04
	BIOCC	4.52	4.52	4.60	4.47	4.63	3.74
	ANCOES	4.39	4.45	4.67	4.54	4.69	4.02
Overall value	IOAC	4.56	4.64	4.49	4.48	4.79	4.25
	IOBC	4.22	4.85	4.56	4.13	4.90	3.17
	BIOCC	4.47	4.73	4.62	4.46	4.74	3.78
	ANCOES	4.48	4.80	4.75	4.46	4.80	4.16
Choices	Groups	Percent of Choices by Exercise					
Percent of choices of best exercise	IOAC	5	13	0	1	55	0
	IOBC	2	16	9	0	54	0
	BIOCC	4	27	2	1	35	0
	ANCOES	1	27	4	0	29	1
Percent of choices as least valuable exercise	IOAC	3	7	5	8	3	17
	IOBC	0	0	4	5	2	60
	BIOCC	3	4	1	10	4	45
	ANCOES	4	2	0	6	5	36

Table 9

EVALUATION OF INTERACTIVE EXERCISES BY ASSESSMENT GROUPS

Rating Scales	Groups	Mean Ratings by Exercise				
		Leader War Game	Field Exercise	Leaderless Group	Conglomerate	Appraisal Interview
Interest	IOAC	4.66	a	4.61	4.65	4.34
	IOBC	a	4.67	4.15	4.65	4.20
	BIOCC	a	4.75	4.50	4.61	4.37
	ANCOES	a	4.86	4.63	4.67	4.54
Challenge	IOAC	4.86	a	4.61	4.47	4.14
	IOBC	a	4.67	4.18	4.43	3.97
	BIOCC	a	4.75	4.43	4.63	4.38
	ANCOES	a	4.88	4.58	4.70	4.58
How well run	IOAC	4.66	a	4.73	4.60	4.65
	IOBC	a	4.50	4.33	4.58	4.28
	BIOCC	a	4.65	4.45	4.61	4.59
	ANCOES	a	4.84	4.46	4.71	4.62
Realism	IOAC	4.27	a	4.44	4.09	4.29
	IOBC	a	4.45	4.33	4.09	4.08
	BIOCC	a	4.31	4.29	4.15	4.32
	ANCOES	a	4.68	4.34	4.37	4.52
Overall value	IOAC	4.45	a	4.55	4.28	4.25
	IOBC	a	4.55	4.19	4.26	4.02
	BIOCC	a	4.58	4.33	4.38	4.34
	ANCOES	a	4.76	4.38	4.51	4.48
Choices	Groups	Percent of Choices by Exercises				
Percent of choices as best exercise	IOAC	20	a	4	2	0
	IOBC	a	12	3	3	0
	BIOCC	a	25	3	4	0
	ANCOES	a	26	1	9	2
Percent of choices as least valuable exercise	IOAC	14	a	6	20	16
	IOBC	a	0	12	9	9
	BIOCC	a	2	11	8	11
	ANCOES	a	7	8	21	11

[a] Captains did not participate in the Field Exercise. Other groups did not participate in the Leader War Game.

In the IOBC group, Simulate preference was by far the strongest followed by the In-Basket and then the Field Exercise. In the IOAC group the Simulate was also most strongly preferred, the War Game came second followed by the In-Basket.

There is a similar spread of choices by the ANCOES group for least valuable exercise but with a much higher percentage for standardized tests. In the BIOCC and IOBC groups, the percentage is still higher, probably due to the lack of opportunity to receive feedback.

Tables 10 and 11 allow comparison of the exercise evaluations by the IOAC captains and the 197th Infantry Brigade Company Commanders. The results are very similar, with the exception that the 197th captains showed a somewhat lower preference for the Simulate along with a higher preference for the In-Basket. In selecting the least valuable exercise, the 197th captains tended more often to choose the Simulate and somewhat less often to choose the War Game and the standardized tests.

In summary, the exercises were evaluated quite positivly by all the groups, but with some differences among groups attributable to differing interest patterns.

Table 10

EVALUATIONS OF INDIVIDUAL EXERCISES BY TWO CAPTAIN SUBGROUPS

Rating Scales	Groups	Mean Rating by Exercise					
		Entry Interview	In-Basket Exercise	In-Basket Interview	Writing Exercise	Radio Simulate	Standard-ized Tests
Interest	IOAC	4.63	4.76	4.64	4.40	4.80	3.87
	197th	4.36	4.58	4.72	4.12	4.58	3.64
	Total	4.56	4.71	4.66	4.33	4.74	3.80
Challenge	IOAC	3.97	4.68	4.14	4.76	4.73	4.28
	197th	3.50	4.58	4.22	4.48	4.58	4.06
	Total	3.84	4.65	4.16	4.68	4.69	4.22
How well run	IOAC	4.81	4.80	4.70	4.66	4.77	4.52
	197th	4.67	4.91	4.78	4.73	4.79	4.55
	Total	4.77	4.83	4.72	4.68	4.78	4.53
Realism	IOAC	4.53	4.43	4.44	4.22	4.62	3.95
	197th	4.48	4.21	4.56	4.21	4.39	4.22
	Total	4.52	4.37	4.47	4.22	4.56	4.03
Overall value	IOAC	4.56	4.64	4.49	4.48	4.79	4.25
	197th	4.27	4.64	4.64	4.15	4.52	4.30
	Total	4.48	4.64	4.53	4.39	4.72	4.27
Choices	Groups	Percent of Choices by Exercise					
Percent of choices as best exercise	IOAC	5	13	0	1	55	0
	197th	6	21	0	0	36	0
	Total	5	16	0	1	50	0
Percent of choices as least valu-able exercise	IOAC	3	7	5	8	3	17
	197th	6	9	6	12	15	9
	Total	4	8	5	9	7	15

Table 11

EVALUATIONS OF INTERACTIVE EXERCISES BY TWO CAPTAIN SUBGROUPS

Rating Scales	Groups	Mean Ratings by Exercise			
		Leader War Game	Leaderless Group	Conglom-erate	Appraisal Interview
Interest	IOAC	4.66	4.61	4.65	4.34
	197th	4.67	4.56	4.58	4.15
	Total	4.66	4.60	4.63	4.29
Challenge	IOAC	4.86	4.61	4.47	4.14
	197th	4.79	4.34	4.42	4.00
	Total	4.84	4.54	4.46	4.10
How well run	IOAC	4.66	4.73	4.60	4.65
	197th	4.55	4.69	4.42	4.00
	Total	4.63	4.72	4.55	4.67
Realism	IOAC	4.27	4.44	4.09	4.29
	197th	4.27	4.06	4.00	4.46
	Total	4.27	4.34	4.07	4.33
Overall Value	IOAC	4.45	4.55	4.28	4.25
	197th	4.51	4.25	4.27	4.09
	Total	4.47	4.47	4.28	4.20
Choices	Groups	Percent of Choices by Exercise			
Percent of choices as best exercise	IOAC	20	4	2	0
	197th	24	3	6	3
	Total	21	3	3	1
Percent of choices as least valu-able exercise	IOAC	14	6	20	16
	197th	6	0	21	15
	Total	12	4	20	16

Chapter 7

RELIABILITY OF ASSESSMENT MEASURES

Reliabilities are given, not just procedures. Report of reliability measurement always requires speification of procedures. The pay-off is the calculated reliabilities. One of the major concerns in constructing the exercises for the Assessment Center was reliability of measurement. ARI personnel spent many hours with the military personnel of the center discussing the theory of measurement, various kinds of common rating errors, and ways of avoiding rating errors. In constructing rating scales for the various exercises, a concerted attempt was made to tie the scale points to readily observable behavior and to make them descriptive rather than evaluative. As each exercise reached the point where it was ready for trial, first with center personnel and later with representatives of the assessment populations, as many assessors as possible observed and rated performance. This procedure served the dual purpose of giving the assessors experience and checking on the adequacy of the scales. These experiences led to extensive scale revisions in almost every exercise.

In the interactive exercises, it was possible to have behavior observed and rated by more than one assessor, and to compute indexes of reliability. Limitations in time and personnel, as well as the nature of the exercises, made it impossible to obtain multiple ratings. Attempts were made to check periodically on rating reliability and scale adequacy. When the bulk of assessment had been completed, and the Modified Assessment Program was being constructed and pretested, personnel were available to obtain multiple ratings on some exercises other than the interactive ones. The results of the reliability analysis on those exercises will be discussed in Chapter Eight where the Modified Assessment Program is described.

Each of the standardized paper and pencil tests had a published test manual which provided adequate evidence of test reliability. Great care was taken to assure that published instructions were followed in administering these tests. Research psychologists from the ARI staff administered all standardized tests during the Assessment Center operation.

NON-INTERACTIVE EXERCISES

Four exercises were characterized as individual or non-interactive. These were the Entry Interview, the In-Basket, the Writing Exercise, and the Radio Simulate. It was not possible in any of these exercises to obtain multiple ratings of performance routinely throughout the operation of the Assessment Center. For all of them however, during the construction and assessor training phases of development, ratings were obtained on common performance samples. This was accomplished either by having all assessors present during the actual performance (with the Radio Simulate), by use of video tape (with the

PRECEDING PAGE NOT FILMED
BLANK

Entry Interview and In-Basket Interview), or by use of duplicated written materials (with the In-Basket and Writing Exercise). The rating scales for all these exercises were revised somewhat as a result of the experience in multiple rating. In addition, the assessors developed a more uniform approach as a result of group discussions of the ratings they had given and their reasons for giving them. It was felt that the final rating scales for these exercises all had at least minimally acceptable reliability.

APPRAISAL INTERVIEW

The Appraisal Interview was a unique exercise in that it produced a written product (an interview plan and recommendations on the candidate to be selected) as well as a performance sample (a video tape of the actual interviews). Since both of these were available for evaluation at any time, rating of Appraisal Interview performance was not rated until after the assessment period was over. When the rating was performed, two assessors met for a three-or four-hour session and rated seven or eight, assessees on the Appraisal Interview. Since these assessors were viewing the video tapes together, and since discussion between them would not disturb the assessee, it was decided that a pooled rating might be most appropriate for this exercise. The two assessors would each make independent judgments on the rating scales and then, through discussion, resolve any differences of opinion. If necessary, the video tape could be replayed to provide another look at the behavior which had produced different ratings. Raters involved in this process reported that large discrepancies in ratings were rare, and in most cases were resolved to the complete satisfaction of both assessors.

CONGLOMERATE EXERCISE

Table 12 shows the interrater reliability averages for the BIOCC groups. Since all assessment groups participated in the same conglomerate exercise, the average interrater correlation coefficients shown here should be fairly good estimates of reliability of observation for all groups.

As might be expected, the lowest interrater agreement occurs on two scales not very directly tied to concrete behaviors--receptiveness to ideas and sensitivity to the feelings of others. Another scale, decision quality, was most frequently not completed. The reliability of judgments on decision quality, when these were made, was fairly high however. Overall the interrater reliability of these scales is quite high. Even with only two raters, the reliability of all the scales except receptiveness and sensitivity would fall above .7.

Table 12

INTER-RATER RELIABILITY OF ASSESSOR RATINGS
ON THE BIOCC CONGLOMERATE EXERCISE

Scales	Average r[a]	Number of r's[b]	Estimated Reliability of 3-Rater Sum[c]
Receptiveness	.345	62	.612
Group Facilitation	.742	68	.896
Emergent Leadership	.782	72	.915
Sensitivity	.340	66	.607
Oral Communication	.659	72	.853
Energy	.799	72	.923
Decision Quality	.657	58	.859
Overall Effectiveness	.782	70	.915

[a] All correlation coefficients were transformed to Fisher's z before averaging.

[b] Three raters observed all 24 6-man groups. Thus 72 inter-rater correlation coefficients were possible. In some cases a rater did not feel able to use a scale for one or more men in the group. Correlation coefficients were excluded when based on less than 4 ratings or when one rater assigned the same rating to all the men he rated.

[c] The reliability of the 3-rater sum was estimated using the Spearman-Brown formula, $r_{33} = 3r_{xx}/(1 + 2r_{xx})$, where r_{xx} is the average r for all groups.

LEADERLESS GROUP DISCUSSION

The Leaderless Group Discussion exercise involved the same content for the BIOCC, IOBC, and ANCOES groups, but was a different problem for IOAC groups. Table 13 gives the interrater reliability of the BIOCC group and Table 14 gives the reliability of the same scales for the IOAC Leaderless Group Discussion. The reliability of all the rating scales is fairly high. Only one scale in each of the two tables has such low reliability that the estimated reliability of the 3-rater sum is below .70. It is not surprising that the rating of social concern has low reliability for the BIOCC Group. The behaviors represented by this scale are subtle and somewhat rare. It may be that the greater complexity of the IOAC problem produced more conflict and thus more opportunity for group members to show concern for the feelings of other group members. It is more surprising that the conveying information scale showed low reliability in the IOAC-197th Groups. Perhaps there was less variability in verbal skills among these more advanced officers, reducing the ratio of true variance to error variance in the ratings given. The rating of leadership and of Group facilitation has high reliability for groups on which reliability was calculated.

ASSIGNED LEADER GROUP EXERCISE

Table 15 gives the interrater reliability of the scales used with the Assigned Leader Group Exercise for the BIOCC group. Once again two scales, Flexibility and Stress Tolerance, show markedly lower reliability than the others. As with the Conglomerate, these two scales deal with characteristics less directly tied to clearly observable behavior. In both cases, a man could be well endowed with the characteristic but not show it, either because of no opportunity (no stress or no frustration), or because he plans ahead and reacts to stress so smoothly that it is not apparent to the rater that there was any stress or any need for flexibility. Once again, the scales dealing with assigned leadership, emergent leadership, and group facilitation produced high reliability.

LEADER WAR GAME

The IOAC and 197th Infantry Brigade captains were the only groups who did not participate in the Assigned Leader Group Exercise in the field. Instead, they participated in the Leader War Game which was also divided into six periods, each with a different assigned leader. Table 16 shows the interrater reliability of the Leader scales. Once again the flexibility rating shows low reliability. This characteristic is very difficult to tie to concrete behavior. The evidence suggests that flexibility as a characteristic cannot be reliably measured in exercises like those used at the Assessment Center. If flexibility needs to be measured, a specially designed exercise should be created with frustrations, changes in requirements, or changes in rules. It might then be possible to specify behaviors which show flexibility and others which do not.

Table 13

INTER-RATER RELIABILITY OF ASSESSOR RATINGS ON THE BIOCC LEADERLESS GROUP DISCUSSION

Scales	Average r[a]	Number of r's[b]	Estimated Reliability of 3-Rater Sum[c]
Formal Oral Communication	.566	70	.796
Oral Organization	.555	70	.789
Presentation Impact	.503	70	.752
Participation	.825	67	.934
Leadership & Group Facil.	.745	70	.898
Persuasiveness	.545	67	.782
Social Concern	.340	65	.607
Avoiding a Negative Impression	.573	61	.801
Conveying Information	.548	70	.784

[a]All correlation coefficients were transformed to Fisher's z before averaging.

[b]Three raters observed 23 of the 24 6-man groups. Only two observers were present for one group. Thus 70 inter-rater correlation coefficients were possible. In some cases a rater did not feel able to use a scale for one or more men in the group. Correlation coefficients were excluded when based on less than four ratings or when one rater assigned the same rating to all the men he rated.

[c]The reliability of the 3-rater sum was estimated using the Spearman-Brown formula, $r_{33} = \frac{3r_{xx}}{1 + 2r_{xx}}$, where r_{xx} is the average for all groups.

Table 14

INTER-RATER RELIABILITY OF ASSESOR RATINGS ON THE IOAC AND 197th INF BDE LEADERLFSS GROUP DISCUSSION

Scales	Average r[a]	Number of r's[b]	Estimated Reliability of 3-Rater Sum[c]
Formal Oral Communication	.551	59	.786
Oral Organization	.646	61	.846
Presentation Impact	.570	63	.799
Participation	.690	59	.870
Leadership & Group Facil.	.708	63	.879
Persuasiveness	.449	61	.710
Social Concern	.635	55	.839
Avoiding a Negative Impression	.552	53	.787
Conveying Information	.289	61	.549

[a]All correlation coefficients were transformed to Fisher's z before averaging.

[b]Three raters observed all 21 of the 6-man groups. Thus 63 inter-rater correlation coefficients were possible. In some cases a rater did not feel able to use a scale for one or more men in the group. Correlation coefficients were excluded when based on less than four ratings or when one rater assigned the same rating to all the men he rated.

[c]The reliability of the 3-rater sum was estimated using the Spearmen-Brown formula, $r_{33} = 3r_{xx}/(1 + 2r_{xx})$, where r_{xx} is the average r for all the groups.

Table 15

INTER-RATER RELIABILITY OF ASSSESSOR RATINGS ON
BIOCC ASSIGNED LEADER GROUP EXERCISE

Scales	Average r[a]	Number of r's[b]	Estimated Reliability of 3-Rater Sum[c]
Planning	.522	51	.766
Leadership	.624	49	.833
Decisiveness	.539	48	.778
Flexibility	.222	44	.461
Motivation	.510	49	.757
Physical Ability	.639	40	.842
Stress Tolerance	.300	49	.562
Emergent Leadership	.600	51	.818
Group Facilitation	.682	49	.865

[a]All correlation cofficients were transformed to Fisher's z before averaging.

[b]Due to inclement weather, 2 of the 24 6-man groups did not participate in this exercise. Three of the groups were osberved by only one rater, three were observed by three raters. Thus 51 inter-rater correlation cofficients were possible. In some cases, a rater did not feel able to use a scale for one or more men in the group. Correlation coefficients were excluded when based on less than four ratings or when one rater assigned the same rating to all the men he rated.

[c]The reliability of the 3-rater sum was estimated using the Spearman-Brown formula, where $r_{33} = 3r_{xx}/(1 + 2r_{xx})$, where r_{xx} is the average r for all groups.

Table 16

INTER-RATER RELIABILITY OF ASSESSOR RATINGS ON IOAC, 197th INF BDE LEADER WAR GAME

Scales	Average r[a]	Number of r's[b]	Estimated Reliability of 2-Rater Sum[c]
Organization	.538	14	.700
Leadership	.463	19	.633
Planning	.616	18	.762
Flexibility	.250	15	.400
Supervisory Skill	.718	18	.836
Participation	.809	20	.894
Problem Comprehension	.782	19	.878
Emergent Leadership	.683	21	.812
Overall Effectiveness	.856	21	.922

[a]All correlation coefficients were transformed to Fisher's z before averaging.

[b]Two raters observed all 21 of the 6-man groups. Thus 21 inter-rater correlation coefficients were possible. In some cases, a rater did not feel able to use a scale for one or more men in the group. Correlation coefficients were excluded when based on less than four ratings or when one rater assigned the same rating to all the men he rated.

[c]The reliability of the 2-rater sum was estimated using the Spearman-Brown formula, with $r_{22} = 2r_{xx}/(1 + 2r_{xx})$, where r_{xx} is the average r for all groups.

Rating of assigned leadership was less reliable in the Leader War Game than in the Assigned Leader Group Exercise. The probable reason is that in the latter exercise, a different problem is presented to each member of the group, whereas, in the Leader War Game, the situation remains the same and only the first and, possibly, the second assigned leader are fully challenged. But some coefficients are higher than in the assigned Leader Group Exercise. Because of this weakness, a new Assigned Leader Exercise was developed for the Modified Assessment Program.

SUMMARY

The spot checks and measured interrater reliabilities indicated that reliability of measurement was, in general, adequate and, in some cases remarkably high for observational measurement of the kind employed.

Chapter 8

THE MODIFIED ASSESSMENT PROGRAM

The original Assessment Center plan had called for completion of scheduled assessments by 1 July 1974. Early in FY 74, a six-month extension of Assessment Center activities was granted for two purposes: One was to expand the data base on BIOCC candidates from 72 to 144 assessees. The other was to design a Modified Assessment Program (MAP) which could be administered by the Infantry School in the IOAC within constraints on available personnel and time. During the first half of FY 75, the MAP was designed and pilot-tested on 72 IOAC captains. During the second half of FY 75 and the first half of FY 76, the MAP was implemented, on a trial basis, as an elective in the IOAC course at the Infantry School. A brief description of MAP is presented below together with reliability measures computed on the scores for the 72 pilot assessees. A more complete description of the MAP program can be found in the Assessment Center After Action Report published by the U. S. Army Infantry School.

BRIEF DESCRIPTION OF MODIFIED ASSESSMENT PROGRAM

The foremost aim in developing the MAP was to retain as much of the essence of the Assessment Center as possible, while reducing time and assessor personnel requirements.

Only two of the original eight behavioral exercises, the Appraisal Interview and the Conglomerate exercise, were completely excluded from the MAP, although several others were modified and one, the Leader War Game, was replaced by a newly designed assigned leader exercise. Only two of the original eight paper-and-pencil tests, the Nelson-Denney Reading Test and the Strong Vocational Interest Blank, were retained in the MAP. Most of the paper-and-pencil, standardized tests had been used in the original Assessment Center as research devices in order to help identify the nature of behavioral dimensions being measured. The two retained seemed to have particular value for career and self-development counseling. Each of the MAP exercises will be described in the following paragraphs. Indication will be given of the ways in which the original exercises were modified for use in the MAP. The changes made in the exercises were based partly on the experience of the assessors in conducting the exercises and in using the rating scales, and partly on preliminary reliability and internal structure analyses of the results of the original exercises.

Entry Interview (1 hour). This exercise was essentially unchanged from the original Assessment Center form, except for a slight reduction in time allotted. There were some revisions in the rating scales used.

Leaderless Group Discussion (2 hours). Again, the essential nature of the exercise was unchanged. A small reduction was made in allotted time. There were also revisions in the scales used.

PRECEDING PAGE NOT FILMED
BLANK

Assigned Leader Exercise (3 hours). This was a new exercise, specially constructed to replace the Leader War Game. Experience with Leader had shown that because the same situation was used throughout the organizing, directing and deciding abilities of leaders assigned in later sequence within the group were not adequately tested. Consequently the MAP Assigned Leader Exercise was designed, as had been the Assigned Leader Group Exercise, with six separate and distinct problems, each requiring its own approach. The problems, for the most part, required sifting and analysis of a large amount of data followed by a synthesis and an eventual decision. Since these problems dealt with written materials, much of the group behavior involved thinking, reading, and writing. This requirement led to some problems of reliability in rating, and continuing revision of the exercise through the pilot testing period.

In-Basket and Writing Exercise (2.5 hours). The In-Basket exercise was modified to a considerable extent. The time was reduced substantially, and the In-Basket Interview was dropped in favor of a written questionnaire completed near the end of the exercise. The shortened time required some simplification of the problems included. The modification resulted in a loss of popularity of the In-Basket, according to responses on the exit questionnaire used with the pilot testing groups. This result was probably due to the dropping of the interview, which provided opportunity for some feedback on performance as well as a chance to explain and justify actions taken. Possibly the men in the pilot groups felt better about the In-Basket after their feedback counseling sessions, although this effect was not measured. Since the MAP In-Basket was an entirely written exercise it was possible to have it scored independently by two assessors to estimate reliability of rating.

The Writing Exercise became an integral part of the In-Basket exercise in the MAP. It was scored separately, however. The MAP pilot study also made possible a test of the reliability with which performance in this exercise was rated.

The Combat Simulate (2 hours). The Combat Simulate was drastically revised, reducing the required time by more than half. A number of different alternatives were tried in an attempt to reduce the requirement of three assessors for each two assessees. None of these efforts proved successful. It was concluded that in order to fulfill adequately the requirements of conducting the problem role-playing, and performance observation and rating, the three-to-two ratio would have to be maintained. Personnel restrictions and the nature of this problem precluded the gathering of data for interrater reliability.

RELIABILITY MEASURES FROM MAP PILOT TEST

The Map Leaderless Group Discussion. In this exercise two assessors observed and rated each group of six assessees. Table 17 presents the results of an interrater reliability analysis on the 72-man pilot test sample. These results are presented by groups in order to convey the extent of variability in the interrater correlations. The

Table 17

INTER-RATER RELIABILITY OF SCALES USED IN THE MODIFIED ASSESSMENT PROGRAM LEADERLESS GROUP DISCUSSION

Scales	Groups 1 & 2		Groups 3 & 4		Total	
	Avg. r[a]	No. of r's[b]	Avg. r[a]	No. of r's[b]	Avg. r[a]	Estimated Reliability of 2-Rater Sum
Communication Skills						
Speaking	.435	6	.665	6	.561	.719
Organization	.225	6	.511	6	.379	.550
Informal	.665	6	.438	6	.561	.719
Social Skills						
Participation	.580	6	.698	6	.644	.783
Interaction	.370	6	.450	6	.412	.584
Forcefulness						
Self Confidence	.605	6	.769	6	.696	.821
Emergent Leadership	.825	6	.743	6	.787	.881
Adaptability	.330	6	.380	6	.355	.524
Motivation						
Task Orientation	.840	6	.555	6	.728	.843
Personal Gain	.510	5	.352	4	.434	.605

[a] All correlation coefficients were transformed to Fisher's z for averaging.

[b] In some cases a rater did not feel able to use a scale for one or more men in the group. Correlation coefficients were excluded when based on less than 4 ratings or when one rater assigned the same rating to all the men he rated.

[c] The reliability of the 2-rater sum was estimated using the Spearman-Brown formula, where

$r_{22} = 2r_{xx}/(1 + 2r_{xx})$, where r_{xx} is the average for all groups.

estimated reliability of the two-rater sum falls below .70 for four of the six scales--Organization in Oral Communication, Social Interaction skill, Adaptability, and motivation for Personal Gain. Some of these reliability coefficients undoubtedly can be improved by scale revision and assessor training. A comparison of Table 17 with Tables 13 and 14 indicates together with scale changes, some loss in average reliability for the MAP. This result may be due to the scale revisions, time reduction, and changes in the conduct of the exercise. Most of the MAP Leaderless Group Discussion scales show at least minimally acceptable reliability.

The MAP Assigned Leader Exercise. In this newly constructed exercise, two assessors observed and rated the performance of each of the six assessees in the exercise groups. Table 18 shows the interrater reliability for the various pilot-test assessment groups. The scale reliability coefficients for groups 1 and 2 are unacceptably low. Between the assessment of groups 1 and 2 and that of groups 3 and 4, considerable work was done to improve reliability. Exercise time was lengthened to allow more time for planning and work assignment by assigned leaders, and the exercise instructions were modified. In addition, some scale anchors were changed and additional training was given to assessors. Overall, reliability improved with groups 3 and 4, in some cases dramatically. Still, with one or two exceptions, reliability remained marginal to unacceptable.

Following the 72-man pilot test, additional training was given to assessors, and 11 of them rated a video tape recording (VTR) of one group's Assigned Leader Exercise. Table 19 shows the results of this rating. Over the 11 assessors, the average reliability was still unacceptably low. It was possible, however, to separate the four assessors who had the highest average interrater correlation. Table 19 shows the average reliability among these four assessors. For every scale but Administrative Skills, these coefficients were quite respectable. With this result in mind, a second VTR was rated by the 11 assessors. They were given clear instructions designed to motivate the highest level of vigilance in observation and the utmost seriousness in making rating judgments. Table 19 shows that the average reliability for all scales, except Communication Skills, improved dramatically. As a check to see whether this improved reliability resulted from an increase in the halo effect, correlation coefficients were computed for ratings on each scale with a randomly selected rating by a different assessor on a different scale. These average cross-assessor, cross-scale correlation coefficients are given in Table 19. They show that a considerable halo effect was operating. For most scales however, the within scale interrater reliability coefficients are higher than the cross-assessor, cross-scale values.

Adequate reliability of measurement has not been clearly established for the Assigned Leader Exercise. Careful watch needs to be maintained on reliability of measurement of this exercise as it is implemented in the Infantry School. Additional revisions in the exercise and the scales may be necessary. It may even be necessary to

Table 18

INTER-RATER RELIABILITY OF SCALES USED IN THE MODIFIED ASSESSMENT PROGRAM ASSIGNED LEADER EXERCISE

Scales	Groups 1 & 2		Groups 3 & 4		Total	
	Avg. r[a]	No. of r's[b]	Avg. r[a]	No. of r's[b]	Avg. r[a]	Estimated Reliability of 2-Rater Sum
Communication Skills	-.035	6	.486	6	.221	.362
Mental Ability	.452	6	.220	6	.324	.489
Administrative Skills	.235	6	.282	6	.253	.404
Supervisory Skills	-.470	6	.813	6	.170	.291
Forcefulness	.200	6	.330	6	.259	.411
Decision Making	.230	6	.155	6	.190	.319
Adaptability	.090	6	.418	6	.249	.399
Social Skills	.200	5	.198	6	.196	.328
Over-all Effectiveness	.280	6	.667	6	.441	.612

[a]All correlations were transformed to Fisher's z for averaging.

[b]In some cases a rater did not feel able to use a scale for one or more men in the group. Correlation coefficients were excluded when based on less than four ratings or when one rater assigned the same rating to all the men he rated.

[c]The reliability of the two rater sum was estimated using the Spearman-Brown formula,

where $r_{22} = 2r_{xx}/(1 + 2^{r}_{xx})$, where r_{xx} is the average for all groups.

Table 19

INTER-RATER RELIABILITY OF 11 ASSESSORS' RATINGS OF TWO VIDEO-TAPE RECORDED ASSIGNED LEADER EXERCISES IN THE MODIFIED ASSESSMENT PROGRAM

Scales	1st VTR		2nd VTR	
	Avg. r[a]	Avg. r[b] Best 4	Avg. r[a]	Cross, Scale, Cross Assessor Avg. r[c]
Communication Skills	.105	.594	.031	.140
Mental Ability	.183	.836	.545	.449
Administrative Skills	.200	.361	.854	.624
Supervisory Skills	.242	.742	.795	.653
Forcefulness	.516	.879	.572	.458
Decision Making	.211	.572	.790	.571
Adaptability	.093	.544	.572	.488
Social Skills	.277	.892	.450	.455
Over-all Effectiveness	.267	.715	.662	.563

[a]All correlation coefficients were transformed to Fisher's z for averaging. Averages in these columns are for the 55 coefficients for all possible pairs of 11 assessors. Each individual coefficient is based on an N of six.

[b]These averages represent the six correlation coefficients between all possible pairs of the four assessors who showed the best average reliability when compared with all other assessors.

[c]These averages represent 11 correlation coefficients between the ratings on each scale given by the 11 assessors and the ratings given on a randomly selected other assessor.

devise a new exercise requiring more overt behavior by assessees, or to re-institute the Assigned Leader Group Exercise from the original assessment center program.

The MAP In-Basket Exercise. The MAP In-Basket produced only a written output, since the interview was replaced by a paper and pencil questionnaire. This change made multiple scoring of the In-Basket materials. This multiple scoring was also rendered by a reduction in the feasible practicable length of the exercise and the greater availability of assessor time in the MAP development and pilot testing.

Table 20 gives the estimated reliability of the MAP In-Basket scales. The correlation coefficients and averages shown are estimates only since, for any one group, the X scores and likewise the Y scores were not all ratings by the same rater. The X and Y groups were nonoverlapping in that no assessor contributed ratings to both sets of scores. On the other hand different combinations of raters could have been used to produce the correlations shown. For example, suppose that six raters were involved in rating six assessees. The list of X and Y scores, for correlation purposes, might look like this:

Assessee	X score rater	Y score rater
1	A	B
2	A	B
3	C	D
4	C	D
5	E	F
6	E	F

Other arrangements are possible as well. Each possible arrangement would produce a somewhat different correlation.

The correlation is likely to be underestimated since the combination of raters in any one score group introduces error. It seems safe to assume that the reliability of the MAP In-Basket scores is uneven, and in most cases, could stand improvement. If multiple ratings could be made in the MAP In-Basket implemented in the Infantry School, the two-rater sum would have at least marginally acceptable reliability for most scales.

It seems probable that the reliability of In-Basket ratings made in the original Assessment Center was somewhat higher, since ratings in that test were based on more written behavioral data and also on the In-Basket Interview. Moreover, the behavioral data may have been more accurately appraised because of explanations provided in the interview for actions that were taken.

The MAP Writing Exercise. The Writing Exercise was performed as an integral part of the In-Basket exercise. It was scored separately, however, by different raters. Table 21 shows the reliability averages for the exercise. Possibly due to greater experience, the raters, seemed to do better with the second set of 35 assessees in groups 3 and 4. Overall, the reliability seems at least marginally acceptable since the estimated reliability of all the two-rater sums was at least .60.

Table 20

INTER-RATER RELIABILITY OF SCALES USED IN MODIFIED ASSESSMENT PROGRAM IN-BASKET EXERCISE

Scales	Estimated Inter-Rater r's[a]								
	Group 1	N[b]	Group 2	N[b]	Group 3	N[b]	Group 4	N[b]	Total[c]
Written Communication	.017	18	.476	17	.706	18	.290	18	.404
Planning and Organizing	.421	16	.699	17	.723	18	.260	18	.549
Directing Ability	.133	15	.334	17	.094	18	.277	18	.189
Task Orientation	.078	13	.667	16	.401	18	.093	18	.336
Decisiveness	-.022	13	.750	14	.564	18	.009	18	.340
Sensitivity	.098	13	.630	17	.355	18	.090	18	.314
Working with Subordinates and Superiors	-.122	13	.382	17	.404	18	.311	18	.251
Human Awareness	.247	13	.683	17	.387	18	.306	18	.424
Problem Solving	-.080	12	.480	13	.301	18	.247	18	.246
Problem Analysis	-.107	14	.227	15	.319	18	.387	17	.213
Supervision of subordinates	.552	15	.699	11	.713	17	.227	18	.574

[a]The correlations shown are estimates only. Since no two raters graded enough common In-Baskets to allow computation of correlation coefficients on individual pairs of raters, non-overlapping groups of assessors were assigned to the X and Y variables and a single correlation was computed for each 18-man assessment group.

[b]In some cases a rater did not feel able to use a scale for some men in the group.

[c]This total represents the unweighted mean of the four individual group correlations. The individual correlations were transformed to Fisher's z for averaging.

Table 21

INTER-RATER RELIABILITY OF SCALES USED IN MODIFIED ASSESSMENT PROGRAM WRITING EXERCISE

Scales	Groups 1 & 2		Groups 3 & 4		Total	
	Avg. r[a]	No. of r's	Avg. r[a]	No. of r's	Avg. r[a]	Estimated Reliability of 2-Rater Sum[b]
Written Communication	.452	4	.546	4	.461	.631
English Errors	.104	4	.813	4	.429	.600
Organization	.490	4	.730	4	.544	.705
Comprehensiveness	.830	4	.923	4	.705	.827
Problem Analysis	.755	4	.774	4	.644	.783

[a]All correlation coefficients were transformed to Fisher's z for averaging.

[b]The reliability of the 2-rater sum was estimated using the Spearman-Brown formula, $r_{22} = \frac{2r_{xx}}{1 + r_{xx}}$,

where r_{xx} = the average r for all groups.

Chapter 9

INTERNAL STRUCTURE OF THE ASSESSMENT EXERCISES

Previous research by Helme et al[1] has shown that in the overall analysis of assessment centers the strongest factors to emerge tend to be specific to individual exercises. In the present research, scores for each exercise were factor analyzed in an attempt to reduce this effect somewhat.

For the most part analyses were conducted for the BIOCC group since this group provided the largest homogeneous sample. For two exercises, Leaderless Group Discussion and Leader War Game, analyses were done with IOAC officers, since these exercises were unique to that group. In every case, the analysis was done using the Statistical Package for the Social Sciences (SPSS). Principal factors were extracted using the squared multiple correlation of each variable with the rest of the variables as the initial communality estimate. An iterative solution was obtained using re-estimates of the communalities with each iteration. Factors with eigenvalues of .40 or greater were rotated orthogonally by the varimax procedure. In a few cases factors were dropped from the final solution when, after rotation, they had insufficient loadings to allow meaningful interpretation.

The appendix tables show results of preliminary analyses with these and other assessment groups. The analyses were done with the Biomed program. They used multiple R's for communality estimates and one per cent of variance accounted for as the criterion for rotation.

ENTRY INTERVIEW

Table 22 shows the rotated factors which emerged from analysis of the Entry Interview. Analysis of the 14 rating scales produced six relatively clear factors. The simple structure approximation of the rotation is quite good.

Factor I seems clearly to indicate that a major dimension of the interviewer's impression was the extent to which the assessee had seriously adopted Army goals as his own (goal congruence), and his determination to prepare himself to make personal contribution to goal achievement (task motivation and task orientation). It is significant that this impression makes a substantion contribution to the interviewer's overall impression.

---

[1]Helme, William H., Willemin, Louis P., and Grafton, Francis C. Dimensions of Leadership in a Simulated Combat Situation (U). ARI Technical Research Report 1172. July 1971.

Table 22

FACTOR ANALYSIS OF ENTRY INTERVIEW
FOR BIOCC CANDIDATES

Scales	Factor Loadings[a]					
	I	II	III	IV	V	VI
Conveying Information	.34			.63		
Sense of Humor			.73[b]			
Verbal Fluency				.75[b]		
Enthusiasm			.72			
Expressing Opinions	.31			.45	.44	.33
Interest Range					.74[b]	
Task Orientation	.61					
Asset Evaluation		.72		.34		
Liability Evaluation		.87[b]				
Goal Congruence	.74[b]					
Self Development	.38				.53	
Task Motivation	.72					
Creativity						.65[b]
Overall Impression	.43		.30	.58	.30	

[a]The ten factors rotated accounted for 82% of the common variance in the matrix. N = 143. Loadings less than .30 have been omittd for clarity.

[b]Scale selected to represent the factor in the overall analysis.

Factor II, realistic self-appraisal, and factor III, humor and enthusiasm, are quite clear and distinct. They appear to make relatively little contribution to overall impression. Factor IV, verbal facility, makes the most substantial contribution to overall impression. This is to be expected.

Factor V, diversity of interest, makes a relatively small contribution to overall impression. Its high loading on self-development is probably due to the contribution of discussions of various training courses taken by the assessee along with self-training in various hobby areas. Factor VI, creativity, is difficult to interpret. This scale is probably the most difficult to define among all those in the exercise. It does not substantially influence overall impression.

A scale was selected to represent each of the six factors in the overall analysis, as indicated in the table. These scales were selected with the aim of maximizing the loading on the factor they represent while minimizing loadings on other factors.

IN-BASKET EXERCISE

Although only 15 scales were used for evaluating in-basket performance, nine moderately strong group factors emerged from the analysis. Table 23 shows 10 factors but Factor IX does not appear substantial. The simple structure is quite good even though the highest loadings on some factors are only moderate. Four scales had loadings above .30 on at least three factors.

Perhaps the clearest factor is Factor I, use of detailed information. It is interesting that planning and analysis, Factor VI, separates out from decision quality, Factor III. There is some overlap on the problem analysis scale. Factor IV, self confidence, and Factor VIII, initiative and task orientation, seem clear and interpretable although they share loadings on working with superiors.

Factor II takes its emphasis from supervision and sensitivity and appears to be a human relations factor. This interpretation is modified somewhat by a moderate loading on task orientation. Factor X has its highest loading on written communication but this loading is not much higher than those of three other scales. This factor may simply reflect the fact that in this exercise, the major way of taking decisive, task-oriented action was by written communication.

A scale was selected to represent each of the nine interpretable factors in the overall analysis. In view of the moderate loadings of some of these, and of diffusion of loadings for several scales, it might be expected that one or more factors specific to this exercise would emerge from the overall analysis.

WRITING EXERCISE

Table 24 shows results of the analysis of the Writing Exercise. Only four scales were used in evaluating the assessee's written product. Three of these appear substantially interrelated. Grammar contributed heavily to both Factor I, spelling and grammer, and Factor II, completeness and grammar. Accuracy seems to be the main component of Factor III, which is quite weak.

One scale represents each of these factors in the overall analysis.

Table 23

FACTOR ANALYSIS OF IN-BASKET EXERCISE
FOR BIOCC CANDIDATES

Scales	Factor Loadings[a]									
	I	II	III	IV	V	VI	VII	VIII	IX	X
Written Communication										.50[b]
Planning & Organization						.65[b]				
Problem Analysis			.31			.70				
Decision Quality			.74[b]					.32		
Directing Skill					.59[b]					
Supervision		.43			.40					.40
Use of Information	.67						.32		.35	
Attention to Detail	.78[b]									
Sensitivity		.49[b]								
Task Orientation		.44						.46		.31
Self Confidence				.51[b]						
Decisiveness								.58		.42
Working with Superiors		.36		.40				.48		
Initiative								.84[b]		
Noting Racial Implications							.62[b]			

[a]The ten factors rotated accounted for 90% of the common variance in the matrix. N = 143. Loadings less than .30 have been omitted for clarity.

[b]Scale selected to represent the factor in the overall analysis.

Table 24

FACTOR ANALYSIS OF WRITING EXERCISE
FOR BIOCC CANDIDATES

Scales	Factor Loadings[a]		
	I	II	III
Accuracy			.36[b]
Grammar	.56	.48	
Spelling	.58[b]		
Completeness		.53[b]	

[a]The three factors rotated accounted for 86% of the common variance in the matrix. N = 143. Loadings less than .30 have been omitted for clarity.

[b]Scale selected to represent the factor in the overall analysis.

SIMULATE EXERCISE GLOBAL RATINGS

Extensive analysis of the three simulates did not reveal substantial clusters of individual action ratings. As a result, this analysis of BIOCC performance involved only global ratings made by by the assessor who role-played superiors and staff agencies (company assessor) and the assessor who role-played subordinates (platoon assessor). A sufficient number of assessors felt unable to rate assessee's performance in specific areas that two scales, organizational identification, and technical and tactical competence, were dropped from the analysis for both platoon and company assessors. The supervisory skills scale was dropped from the company assessor ratings only. The simulate used for BIOCC candidates was based on a civil emergency and not on a military combat situation. This probably accounts, at least in part, for difficulty in using the two scales. The company assessors in this simulate had little opportunity to observe the assessee functioning in a supervisory role.

It is readily apparent from Table 25 that platoon assessors were unable to discriminate among the behaviors represented by the several scales. With minor exceptions, their ratings cluster into one global factor of supervisory performance.

Table 25

FACTOR ANALYSIS OF GLOBAL RATINGS BY CONTROLLER/ASSESSORS IN SIMULATE EXERCISE FOR BIOCC CANDIDATES

Scales	Factor Loadings[a]						
	I	II	III	IV	V	VI	VII
Platoon Assessor Ratings							
Social Skills	.71				.30		
Communication Skills	.52						.68
Adaptability	.72						
Motivation	.66		.42				
Forcefulness	.54		.46				
Decision Making	.67			.30			
Administration Skills	.68						
Supervisory Skills	.88[b]						
Organizational Role Effectiveness	.71						
Company Assessor Ratings							
Social Skills					.76[b]		
Communication Skills				.30		.66[b]	
Adaptability	.31	.37		.44			
Motivation		.36	.52				
Forcefulness			.76[b]				
Decision Making				.72[b]			
Administration Skills		.73[b]					
Organizational Role Effectiveness		.53	.30				

[a]The seven factors rotated accounted for 81% of the common variance in the matrix. N = 143. Loadings less than .30 have been omitted for clarity.

[b]Scale selected to represent the factor in the overall analysis.

Company assessors, on the other hand, produced ratings that separated statistically into five relatively independent factors. Factor II, administrative ability and Factor VI communication skill had no corroborating loadings from the platoon assessor ratings. A separate communication skills factor (Factor VII) emerged, loaded only on the platoon assessor rating. This scale was not selected for the overall analysis because of its high loading on Factor I.

Factor III, forcefulness and motivation, shows relatively strong evidence from both company and platoon assessor ratings. Factor IV, decisionmaking, and Factor V, social skills have only weak corroboration from loadings of platoon assessor scales.

APPRAISAL INTERVIEW

Table 26 shows that the eight rating scales of the appraisal interview separate clearly into three factors. The first two relate to the written documents prepared by the assessee while the third deals generally with actual performance in the interview. Factor I is planning of the interview. Factor II is a factor dealing with the quality of the written plan and the final recommendations. Factor III seems to be primarily a global rating of the self-confidence of the assessee and his smoothness in the interview.

LEADERLESS GROUP DISCUSSION

The 15 ratings of performance in this exercise fall into three categories. The first three scales are ratings of the formal oral presentations each assessee made at the beginning of the discussion. The next six scales are ratings of his performance in the discussion itself. And the last six scales are average rankings, on the scales listed, given by the other assessees in the discussion group.

The factors shown in Table 27 reflect to a large extent the scale categories discussed above. Factor I is clearly a dimension reflecting overall quality of the oral presentation. Factor II seems to be one of task-oriented leadership. This is shown primarily by the loadings on assessor ratings, but is quite strongly supported by the loadings on leadership and idea quality as rated by the other participants. Factor III seems to be a group maintenance or social concern factor. It appears only to be seen by assessors, as it is not supported by any substantial loadings on participant ratings. Factor IV apparently is a global impression made on other participants. Factor V did not have sufficiently high loadings for clear interpretation or for representation in the overall analysis.

Table 26

FACTOR ANALYSIS OF APPRAISAL INTERVIEW EXERCISE FOR BIOCC CANDIDATES

Scales	Factor Loadings[a]		
	I	II	III
Planning	.81[b]		
Topic Selection	.81		
Written Communication		.81[b]	
Written Organization		.76	
Self Confidence			.85[b]
Oral Communication			.81
Use of Information			.63
Accommodation			.68

[a]The three factors rotated accounted for 75% of the common variance in the matrix. N = 143. Loadings less than .30 have been omitted for clarity.

[b]Scale selected to represent the factor in the overall analysis.

LEADERLESS GROUP DISCUSSION--IOAC OFFICERS

A separate analysis of the Leaderless Group Discussion (LGD) was carried out for IOAC Officers, since LGD was a different problem for this group. For all other groups except 197th Infantry Brigade company commanders, LGD involved a discussion to select a Soldier-of-the-Month, but for the IOAC and the 197th groups the task involved allocation of end-of-year funds among several agencies on a large Army post. The fund allocation problem was considerably more complex and allowed for a greater latitude of compromise in the final solution.

Table 27

FACTOR ANALYSIS OF LEADERLESS GROUP DISCUSSION EXERCISE FOR BIOCC CANDIDATES

Scales	Factor Loadings [a]				
	I	II	III	IV	V
Assessor Ratings					
Formal Oral Communication	.79	.36			
Oral Organization	.85				
Presentation Impact	.91[b]				
Participation		.76			
Leadership and Group Facilitation		.85[b]			
Persuasiveness		.80			
Social Concern			.77		
Avoiding Negative Impression			.80[b]		
Conveying Information	.37	.72			
Participant Ratings					
Oral Communication		.31		.77	
Leadership		.62		.50	.37
Persuasiveness				.74	
Idea Quality		.43		.67	.37
Sociability				.70	
Overall Effectiveness				.85[b]	

[a] The five factors rotated accounted for 84% of the common variance in the matrix. N = 143. Loadings less than .30 have been omitted for clarity.

[b] Scale selected to represent the factor in the overall analysis.

Table 28 shows that, in spite of the different content of the problem and the greater experience of the captains, the analysis results are remarkably similar to those for BIOCC candidates. Factor I is essentially the same with the addition of loadings from participant ratings on oral communication and overall effectiveness. Its interpretation as representing impact of the formal oral presentation is unchanged. Factor II is clearly a dimension of leadership and group facilitation. However, in the IOAC groups there are additional loadings on social concern and on participant rating of overall effectiveness. On the other hand, loadings on assessor rating of formal oral communication and on participant ratings of oral communication and idea quality are under .30. Factor III is very much the same for the two groups and appears to represent social concern. In the IOAC groups, however, this factor shows a moderate negative loading on participation.

The remaining factors are defined by participant ratings. Factor IV is essentially the same for both groups. It is a strong global performance rating by participants and has no substantial loadings from assessor rating scales. Factor V of the BIOCC group is not reproduced in the IOAC group. Instead a factor with loadings on participant ratings of overall effectiveness and oral communication emerges. Two other factors also emerged but each had only one substantial loading. The fact that Factor VI loads on participant rating of leadership and Factor VII on sociability suggests something of the separation between these two peer impressions that emerged in the analysis of the Assigned Leader Group Exercise for the BIOCC group.

ASSIGNED LEADER GROUP EXERCISE

The 13 scales of this exercise are divided into four groupings. The first four scales were designed for rating the assessee's behavior while he was acting as assigned leader. The next three scales, motivation, physical ability, and stress tolerance, were used for rating his behavior throughout the exercise. The last two assessor rating scales, emergent leadership and group facilitation, were designed for rating his behavior as a follower. The remaining four scales consisted of average rankings given by the other participants in his group.

The factors which emerged from the analysis, shown in Table 29, correspond only roughly with the scale groupings. Factor I is primarily an assigned leadership factor, although it does show substantial loadings on scales from the two other assessor rating scale groups. Factor III seems to be primarily an emergent leadership factor, but again there are substantial loadings on motivation and physical ability. Factor IV is a factor of stamina and physical ability, with moderate loadings on group facilitation and flexibility from the emergent and assigned leadership scale groupings respectively.

Table 28

FACTOR ANALYSIS OF LEADERLESS GROUP DISCUSSION EXERCISE FOR IOAC OFFICERS

Scales	Factor Loadings[a]						
	I	II	III	IV	V	VI	VII
Assessor Ratings							
Formal Oral Communication	.83						
Oral Organization	.84						
Presentation Impact	.94						
Participation		.68	-.39				
Leadership and Group Facilitation		.83					
Persuasiveness		.73					
Social Concern		.33	.84				
Avoiding Negative Impression			.86				
Conveying Information	.44	.65					
Participant Ratings							
Oral Communication	.41			.64	.35		
Leadership		.57		.54		.32	
Persuasiveness				.87			
Idea Quality				.86			
Sociability				.55			.42
Overall Effectiveness	.30	.34		.65	.47		

[a]The seven factors rotated accounted for 88% of the common variance in the matrix. N = 88. Loadings less than .30 have been omitted for clarity.

Table 29

FACTOR ANALYSIS OF ASSIGNED LEADER GROUP EXERCISE FOR BIOCC CANDIDATES

Scales	Factor Loadings [a]					
	I	II	III	IV	V	VI
Assessor Ratings						
Planning	.75					
Leadership	.84		.33			
Decisiveness	.85[b]					
Flexibility	.54			.36		
Motivation	.46		.61	.41		.40
Physical Ability			.43	.60		
Stress Tolerance	.40			.70[b]		
Emergent Leadership	.48	.39	.64			
Group Facilitation	.38	.31	.72[b]	.38		
Participant Ratings						
Social Association					.69[b]	
Leadership		.91[b]				
Support of Leader		.72				
General Esprit		.68			.31	

[a]The six factors rotated accounted for 89% of the common variance in the matrix. N = 131. Loadings of less than .30 have been omitted for clarity.

[b]Scale selected to represent the factor in the overall analysis.

Factor II clearly reflects the participants' view of task orientation and leadership skill. This interpretation is supported by moderate loadings on emergent leadership as rated by assessors. Factor V is the participants' view of the assessee as a potential friend. These two participant factors show a clear split along the lines of task orientation and social association.

Factor VI is a weak one showing some separation of motivation as a distinctive quality. Due to the loadings of motivation on other factors, this factor was not represented in the overall analysis.

LEADER WAR GAME--IOAC OFFICERS

Whereas all other groups participated in the Assigned Leader Group Exercise, the IOAC captains and the 197th Infantry Brigade company commanders did not. Instead, they participated in another assigned leader exercise, the War Game. As with ALGE, the War Game was conducted in six periods. During each period, one member of the group was assigned as leader, and his planning, organizing, and supervisory skills were observed.

Whereas the ALGE focused on physical obstacles and situational problems, the War Game highlighted fact finding, prediction, and intellectual problem solving. The War Game differed also in introducing competition between separate six-man teams.

Table 30 shows the six factors which emerged from the internal analysis of the War Game. Factor I seems to represent clearly active leadership in the assigned leader role. Factor II, on the other hand, represents emergent leadership and effectiveness in all phases of the exercise, whether as leader or follower. Although both these factors have their primary loadings on assessor ratings, consistent, if not strong, support is shown by loadings on participant ratings. Factor III represents the global impression made on other participants in the exercise. The loadings from assessor ratings suggest that participants were more impressed by problem comprehension and overall participation than by performance in the assigned leader role. It is significant that assessor ratings of participation and problem comprehension make substantial contributions to all three factors. Factor IV is a relatively weak factor of planning and flexibility, while Factor V splits off organization as a somewhat unique dimension in addition to its contribution to the assigned leadership factor. Factor VI is quite weak. The negative loading on problem comprehension is shown despite its being below .30, in order to indicate the hint of separation between task orientation or competency and general esprit. The latter is the closest participant rating to the area of friendship and social association. This result may indicate something of the contrast between task and social evaluation shown by the ALGE participant ratings for BIOCC candidates and the LGD analysis for IOAC officers.

Table 30

FACTOR ANALYSIS OF LEADER WAR GAME EXERCISE FOR IOAC OFFICERS

Scales	Factor Loadings[a]					
	I	II	III	IV	V	VI
Assessor Ratings						
Organization	.58				.64	
Leadership	.85					
Planning	.62	.33		.41		
Flexibility				.58		
Supervisory Skills	.86					
Participation	.61	.57	.30			
Problem Comprehension	.36	.64	.36			
Emergent Leadership		.81	.44			
Overall Effectiveness		.73	.46			
Participant Ratings						
Problem Comprehension		.37	.80			(-.29)
Leadership	.32		.71			
Support of Leader		.36	.77			
General Esprit			.67			.30
Overall Effectiveness		.34	.86			

[a]The six factors rotated accounted for 90% of the common variance in the matrix. N = 88. Loadings less than .30 have with one exception been omitted for clarity.

CONGLOMERATE EXERCISE

The 13 scales of this exercise are divided into only two groups, those completed by participants and those completed by assessors. The results of the analysis of the Conglomerate Exercise are shown in Table 31. Factor I seems to reflect concern with interpersonal relations, although the substantial loading on decision quality is troublesome. Perhaps this loading reflects the extent to which decisions were judged on their social acceptability. Factor II is more readily interpreted as task-oriented leadership. Aside from receptivity on Factor I and emergent leadership on Factor II there is substantial overlap of loadings. The two factors consequently seem far from independent. It is significant, however, that only Factor II loads substantially on assessor rating of overall effectiveness.

Factor III sems to reflect a global judgment of effectiveness by the exercise participants. Assessor rating of overall effectiveness also has its highest loading on this factor, suggesting that the assessor's evaluation of the assessee's overall performance is not adequately represented in the other seven scales. Another possibility is that this is an artifact stemming from the facts that the participant scales reflect average ranks assigned by the other participants to each assessee, and that the overall effectiveness score was also based on a ranking, this time made by assessors. Factor IV loads only on participants' estimates of the degree to which each member of the group caused conflicts within the group.

PAPER AND PENCIL TESTS

Although these tests were given at several sessions, they were analyzed as a group in order to reduce the number of scores put into the overall analysis. Table 32 shows the results of this analysis.

Factor I is clearly one of general intellectual ability. It is disappointing to discover that the Watson-Glaser Test of Critical Thinking and the Social Insight Test contributed nothing beyond that already being measured by the Nelson-Denney Reading Test and the Henmon-Nelson Test. Factor X is a relatively weak factor showing a relationship between the Watson-Glaser and reading comprehension. It was not represented in the overall analysis.

Only three other factors had substantial loadings on more than one test. Factor II shows an interesting negative relationship between need for achievement as measured by the Edwards Personal Preference Schedule (EPPS) and consideration (in leadership) as measured by the Leadership Opinion Ouestionnaire (LOQ). Factor IX shows a relationship between need for aggression on EPPS and leadership potential as measured by the Leadership Q-Sort (LQS). Factor IX shows a relationship between the Work Environment Preference (WEPS) and the consistency score of EPPS. The WEPS score was selected for inclusion in the overall analysis because the consistency score is designed only as a check on seriousness in completing the EPPS--a form of lie scale.

Table 31

FACTOR ANALYSIS OF CONGLOMERATE EXERCISE
FOR BIOCC CANDIDATES

Scales	Factor Loadings[a]			
	I	II	III	IV
Assessor Ratings				
Receptivity to Ideas	.80[b]			
Group Facilitation	.37	.83	.30	
Leader Emergence		.84[b]	.35	
Sensitivity	.73	.33		
Oral Communication	.50	.74	.30	
Energy and Vigor	.33	.72	.30	
Decision Quality	.51	.68		
Overall Effectiveness		.43	.63	
Participant Ratings				
Popularity			.84[b]	
Quality of Ideas		.36	.83	
Energy			.80	
Conflict Creation				.57[b]
Working with Peers			.80	

[a]The four factors rotated accounted for 85% of the common variance in the matrix. N = 143. Loadings less than .30 have been omitted for clarity.

[b]Scale selected to represent the factor in the overall analysis.

Table 32

(Page 1 of 2)

FACTOR ANALYSIS OF PAPER AND PENCIL TESTS
FOR BIOCC CANDIDATES

Tests	Factor Loadings[a]										
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI
Henmon-Nelson											
Quantitative	.65										
Verbal	.80										
Nelson-Denney											
Vocabulary	.77										
Comprehension	.82[b]									.30	
Reading Rate											
Watson-Glaser	.68									.51	
Social Insight (WEPS)	.66										.30[b]
Leader Opinion Q (LOQ)											
Consideration		-.52									
Structure			.59[b]								
Leadership Q. Sort											
Leader Potential									.48[b]		
Personal Integrity				-.52	-.41	-.34	-.47	-.41			
Consideration								.83[b]			
Mental Health					.88[b]						
Tech. Information				.86[b]							
Decisionmaking						.76[b]					
Teaching and Communication							.79[b]				
Edwards											
Achievement	.82[b]										
Aggression									.38		
Consistency											.58

Table 32

(Page 2 of 2)

Tests	Factor Loadings[a]											
	XII	XIII	XIV	XV	XVI	XVII	XVIII	XIX	XX	XXI	XXII	XXIII
Edwards												
Achievement												
Deference	.84[b]											
Order		.79[b]										
Exhibitionism			.79[b]									
Autonomy				-.54						-.30		
Affiliation					.85[b]							
Introception						.86[b]						
Succorance							.80[b]					
Dominance								.83[b]				
Abasement									.83[b]			
Nurturance										.74[b]		
Change											.79[b]	
Endurance				.86[b]								
Heterosexuality												.88[b]
Aggression	-.30	-.37			-.50							
Consistency												

[a]The 23 factors rotated accounted for 92% of the common variance in the matrix. N = 143. Loadings less than .30 in absolute value have been omitted for clarity.

[b]Scale selected to represent the factor in the overall analysis.

The scale scores of the LOO, LQS, and EPPS are almost completely factorially independent. On LQS, the Personal Integrity score seems opposed to all the other scale scores except leadership potential. The remaining scale scores are independent of each other.

Among the EPPS scores aggression is negatively related to deference, order, and affiliation. Autonomy is negatively related to endurance and nurturance. Achievement relates negatively to consideration on the LOQ as mentioned above. The remaining EPPS scales form independent, unique factors.

In general it seems that, while not all the scales of the various exercises emerge as independent dimensions, each exercise appears to have produced measures of several potentially significant behavioral dimensions. The generality of these dimensions across exercises will be explored.

Chapter 10

DIMENSIONAL STRUCTURE OF OVERALL ASSESSMENT PERFORMANCE OF BIOCC CANDIDATES

As described in the previous chapter, a single score was selected to represent each substantial factor which emerged from the analyses of the separate exercises and the paper and pencil tests. In all, 62 scores were thus selected for inclusion in the overall analysis. Due to the size of this matrix, and resulting prohibitive computer time requirements, two modifications were made in the factoring procedure used. While the smaller matrices had been factored using an iterative solution, the overall matrix was factored using a non-iterative solution. The iterative procedure used multiple correlation coefficients as the communality estimates and reestimated these at each new step. The non-iterative procedure used one as the communality estimate and did not reestimate. These two procedures are called PA 2 and PA 1, respectively, in the Statistical Package for the Social Sciences. A second difference involved the decision on how many factors to rotate. With the smaller exercise matrices, all factors with eigenvalues of .4 or higher were rotated. In the overall analysis, only factors with eigenvalues of 1.0 or higher were rotated.

Emerging from the analysis were 21 factors. As mentioned in the previous chapter there was some concern that exercise specific factors would emerge from the overall analysis. It was hoped that by including only scales which loaded highly on the factor they represented, and which did not load substantially on other factors from the exercise, the emergence of exercise-specific factors could be avoided. This procedure was mostly successful. While a few factors emerged, on which all scales from a specific exercise loaded, in most cases scales from other exercises also loaded on the factor. Many of the factors were defined by scales from several exercises. Even in cases where all scales from an exercise loaded on a single factor, the same scales often loaded substantially on other multi-exercise factors. It was somewhat surprising, and satisfying, that on a number of factors paper-and-pencil tests as well as exercise scale scores loaded substantially.

Tables 33, 34, and 35 provide listings of scales which had loadings of .20 or higher in absolute value. The reader should not expect to find clear distinctions between the factors in the three tables. The grouping of factors is somewhat arbitrary and used simply as a way of organizing the following discussion.

FACTORS PRIMARILY DEFINED BY A SINGLE EXERCISE (Table 33)

Factor I. Factor I is clearly a global factor representing performance in the In-Basket exercise. All the IB scales load above .20 on factor. The only loading, on participant rating of social attrativeness in the Assigned Leader Group Exercise, does not add significantly to interpretation of the factor.

PRECEDING PAGE NOT FILMED
BLANK

Table 33
(Page 1 of 2)

FACTORS FROM OVERALL ANALYSIS DEFINED PRIMARILY
BY LOADINGS FROM A SINGLE EXERCISE

Factors and Scales	Loadings	Loadings on Other Factors
Factor I		
IB Decision Quality	.82	
IB Attention to Detail	.73	
IB Directing Skill	.72	VII .20, X -.21, XIII -.21
IB Initiative	.72	XVI .24
IB Planning and Organization	.61	VIII .29, X .21, XIII .29, XIV -.36, XV .21
IB Written Communication	.44	V -.20, XVII -.36
IB Self-Confidence	.36	XII -.49, XIII .25, XIX .20
IB Sensitivity	.24	III .27, XVI .58
ALGE Part., Social Association	.24	II .21, IV .20, VI .31, XII .44, XIX .49
IB Noting Racial Implications	.21	XV .57, XVI .36
Factor II		
SIM CAR Comm. Skills	.78	
SIM CAR Decision Making	.75	
SIM CAR Forcefulness	.64	VI .21, VII .35
SIM CAR Social Skills	.52	III .21, XI -.31, XXI .22
SIM PAR Supervisory Skills	.50	XIII .45
SIM CAR Admin. Skills	.47	VI .25, XV .38, XVII -.22
LGD Part., Overall Effectiveness	.30	IV .28, VII .45, XI -.22, XIX .22
ND Reading Comprehension	.30	VII .27, XIII .49, XVIII -.30
LGD Group Leadership & Facil.	.26	IV .24, VII .61
CONG Part., Popularity	.23	VI .48, VII .33, XIV -.27
ALGE Part., Social Assoc.	.21	I .24, IV .20, VI .31, XII .44, XIX .49
EPPS Need for Abasement	-.20	III .21, XIV -.24, XVIII .63
LOQ Structure	-.45	VI .22, IX .25, X .37, XIV .21
Factor IV		
ALGE Group Facilitation	.86	
ALGE Stress Tolerance	.80	
ALGE Decisiveness	.76	VII .22
ALGE Part., Leadership	.63	VIII .23, XIV -.20, XIX .30
AI Self-Confidence	.44	VII .43, XI .23, XIV .26
LGD Part., Overall Effectiveness	.28	II .30, VII .45, XI -.22, XIX .22
LGD Group Leadership & Facil.	.24	II .26, VII .61
ALGE Part., Social Association	.20	I .24, II .21, VI .31, XII .44, XIX .49

Table 33
(Page 2 of 2)

Factor and Scales		Loadings	Loadings on other Factors
Factor V			
EI	Goal Congruence	.81	
EI	Creativity	.64	III .23, XV .21
EI	Interest Range	.55	VI .48, VIII -.24, XVII .22
EI	Liability Evaluation	.51	III .24, XIII .25, XVII .20
EI	Sense of Humor	.49	VII .34, X .33, XI -.25
EI	Verbal Fluency	.44	VII .52, XII .34, XVII .23
LQS	Leadership Potential	.34	III .29, VI .26, VII -.33, XIII .37, XXI -.22
LQS	Technical Information	.23	VIII -.52, XVIII -.35
IB	Written Communication	-.20	I .44, XVII -.36
Factor IX			
AI	Planning	.80	
AI	Written Communication	.75	XXI -.20
LOQ	Structure	.25	II -.45, V .22, IX .25, X .37, XIV .21

Key to abbreviations:

AI	Appraisal Interview	LOQ	Leadership Opinion Questionnaire
ALGE	Assigned Leader Group Exercise	LQS	Leadership Q Sort
CONG	Conglomerate	ND	Nelson Denny Reading Test
EI	Entry Interview	SIM CAR	Simulate: Company Assessor Report
EPPS	Edwards Personnel Preference Schedule	SIM PAR	Simulate: Platoon Assessor Report
IB	In-Basket	WE	Written Expression
LGD	Leaderless Group Discussion	WEPS	Work Environment Preference Schedule

Several of the IB scales--notably self confidence, sensitivity, and noting racial implications--make greater contributions to other factors.

Factor II. This factor seems to be a global representation of performance in the Simulate exercise. Unlike Factor I however, this factor shows a number of other loadings which contribute substantially to understanding the nature of the dimension. Apparently the personal characteristics underlying this factor contribute to success in all three of the group exercises, at least to the extent of the impression made on other participants: participant ratings in the Leaderless Group Discussion (LGD), Conglomerate, and Assigned Leader Group Exercise (ALGE) all load on this factor. Assessor rating of group leadership in LGD also has a loading. Verbal intelligence is represented by the loading of Nelson-Denney Reading Comprehension. One thing which sets this factor apart from others is the relatively high negative loading on the structuring score from the Leadership Opinion Questionnaire (LOQ). Apparently this factor represents a competent but easy-going social interaction skill.

Factor VI. This factor is primarily defined by the scores from the Assigned Leader Group Exercise. All five of the ALGE scores entering the analysis load on this factor, four substantially and participant rating choice for social association only to the extent of .20. Performance in the Appraisal Interview, as represented by the self-confidence scale, also is fairly closely related to this dimension. Leaderless Group Discussion performance, as seen by both participants and assessors, is also related but somewhat less closely. This factor may represent a kind of performance similar to that represented by Factor VII, discussed in the next section. Both the LGD scores and the AI scores load substantially on both factors. Uniqueness of Factor VI must be found in the ALGE itself. Several unique characteristics of the ALGE should be considered in interpreting this factor: 1) both assigned and emergent leadership, 2) an actual physical task, 3) operation in an outdoor setting, 4) requirements for physical agility, and stamina. This factor seems to involve interpersonal skills which are verbal to some extent but go beyond this. Self-confidence may be the key attribute.

Factor V. This factor seems to represent the global impression created by the assessees in the Entry Interview. All the EI scales load substantially on this factor, though some have substantial loadings on other factors as well. The loading of the Leadership Q Sort's Leadership Potential score may be a little puzzling at first glance. If one speculates on the meaning of rated performance in the EI, however, an explanation suggests itself. The interviewers were very likely to be impressed by an assessee whose demeanor and verbal responses were congruent with the interviewer's expectations of an Army officer. In fact, the highest EI loading on this factor is for the scale which represents the interviewer's estimate of the congruity between the expressed goals of the assessee and the requirements of an Army officer career. The leadership potential score of the LQS indicates something similar between the value ratings assigned to statements about leadership by the assessee and those assigned by a selected group of leaders, including a substantial number of military officers. The loading on the LQS technical information scale sugests that assigning somewhat higher values to statements emphasizing technical know-how as valuable for a leader is congruent with

the judgment of Army officers. The negative loading on In-Basket written communication, in the presence of a positive loading on verbal fluency is puzzling but similar to the situation in Factor XVII. It will be discussed there. In summary, Factor V seems to represent the congruity of the assessees' expressed attitudes and values with experienced officers' estimates of Army career demands.

Factor IX. This factor is defined almost completely by loadings on scales rating the written plan and recommendations produced in the Appraisal Interview exercise. The rating of performance in the actual interviews, represented by the AI self confidence score, does not relate to this factor. The loading on LOQ structure suggests that better ratings on the written portion of AI are achieved by people who favor a structured approach to execution of assigned leadership responsibilities. The underlying skill represented is difficult to pinpoint. It is disappointing that none of the scales from the Writing Exercise or the In-Basket loaded on this factor.

FACTORS DEFINED BY SUBSTANTIAL LOADINGS ON TWO OR MORE EXERCISES (TABLE 34)

Factor VI. Factor IV, described in the previous section, was largely defined by loadings from ALGE; this factor largely by loadings on the Conglomerate exercise. The reason for assigning Factor IV to the single exercise category and Factor VI to the multi-exercise catgegory is that all the ALGE scores loaded on Factor IV and there were only three other substantial loadings. In the case of Factor VI, only three of the four conglomerate scales loaded on it and eight other exercise scales had substantial loadings.

It should be noted that the Conglomerate popularity rating represents the global appraisal of performance by participants. The participant ratings obtained in the Conglomerate exercise did not produce the separation between ratings of leadership performance and choice for social association that was found in ALGE. On the other hand, the ALGE participant rating on social association loads on this factor, reflecting the fact that the Conglomerate exercise involved performance that was social and interpersonal in nature. The loading on Entry Interview interest range indicates that persons who, in the interview, showed broad interests were more likely to be interested in the Conglomerate exercise and to take an active role in it. Of all the exercises, this exercise was least military in character. The participants seemed either to like it or dislike it with few in between.

The loadings on LQS, SIM, and LOQ suggest that in the somewhat ambiguous and unfamiliar Conglomerate exercise, successful performance required a forceful and structured approach. The negative loadings on EPPS need for Change and Achievement are puzzling. Change should indicate a desire to try something new and Achievement should have been engaged by the problems and challenge. Perhaps the change was too great and the problems so unfamiliar as to be frustrating.

Table 34
(Page 1 of 4)

FACTORS FROM OVERALL ANALYSIS DEFINED BY SUBSTANTIAL LOADINGS FROM TWO OR MORE EXERCISES

	Factors and Scales	Loadings	Loadings on Other Factors
Factor VI			
CONG	Receptivity to Ideas	.80	
CONG	Leader Emergence	.62	VII .23, XII -.37
EI	Interest Range	.48	V .55, VIII -.24, XVII .22
CONG	Part., Popularity	.48	II .23, VII .33, XIV -.27
ALGE	Part., Social Association	.31	I .24, II .21, IV .20, XII .44, XIX .49
LQS	Leadership Potential	.26	III -.29, V .34, VII -.33, XIII .37, XXI -.22
SIM	CAR Admin. Skills	.26	II .47, XV .38, XVII -.22
LOQ	Structure	.22	II -.45, IX .25, X .37, XIV .21
SIM	CAR Forcefulness	.21	II .64, VII .35
EPPS	Need for Change	-.21	III .79, XI -.22
EPPS	Need for Achievement	-.23	VIII -.23, XI -.24, XVI .21, XVIII -.24, XIX .51
FACTOR VII			
LGD	Presentation Impact	.78	
LGD	Group Leadership & Facil.	.61	II .26, IV .24
EI	Verbal Fluency	.52	V .44, XII .34
EPPS	Need for Dominance	.46	XIII -.37, XVII -.39, XVIII -.23
LGD	Part., Overall Effectiveness	.45	II .30, IV .28, XI -.22, XIX .22
AI	Self-confidence	.43	IV .44, XI .23, XIV .26
SIM	CAR Forcefulness	.35	II .64, VI .21
EI	Sense of Humor	.34	V .49, X .33, XI -.25
CONG	Part., Popularity	.33	II .23, VI .48, XIV -.27
ND	Reading Comprehension	.27	II .30, XIII .49, XVIII -.30
CONG	Leader Emergence	.23	VI .62, XII -.37
ALGE	Decisiveness	.22	IV .76
EPPS	Need for Exhibition	.21	XIX .68, XX -.20
IB	Directing Skill	.20	I .72, X -.21, XIII -.21
EPPS	Need for Heterosexuality	-.20	III .36, X .21, XI -.25, XVII .30, XVIII -.33, XX -.21
LQS	Leadership Potential	-.33	XXI -.30

Table 34
(Page 2 of 4)

Factors and Scales		Loadings	Loadings on Other Factors
Factor XI			
EPPS	Need to Show Deference	.76	
EPPS	Need for Order	.34	VIII .24, XV -.26, XX .62
AI	Self-confidence	.23	IV .44, VII .43, XIV .26
EPPS	Need for Succorance	.22	XIX -.22, XX -.44, XXI -.44
CONG	Part., Avoiding Conflict	.20	VI .76
LGD	Part., Overall Effectiveness	-.22	II .30, IV .28, VII .45
EPPS	Need for Change	-.22	III .79, VI -.21
EI	Sense of Humor	-.25	V .49, VII .34, X .33
EPPS	Need for Heterosexuality	-.25	III -.36, VII -.20, X .21, XVII .30, XVIII -.33, XX -.21, XXI -.30
SIM	CAR Social Skills	-.31	II .52, II -.21, XXI .22
WE	Accuracy	-.39	XIV .25, XIX -.49
FACTOR XII			
LGD	Avoiding Negative Impression	.75	
ALGE	Part., Social Association	.44	I .24, II .21, IV .20, VI .31, XIX .49
EI	Verbal Fluency	.34	V .44, VII .52, XVII .23
CONG	Leader Emergence	-.37	VI .62, VII .23
IB	Self-confidence	-.49	I .36, XIII .25, XIX .20
Factor XIII			
ND	Reading Comprehension	.49	II .30, VII .27, XVII -.30
SIM	PAR Supervisory Skills	.49	II .50
LQS	Leadership Potential	.37	II -.29, V .34, VI .26, VII -.33, XXI -.22
IB	Planning and Organization	.29	I .61, VIII -.29, X .21, XIV -.36, XV .21
IB	Self-confidence	.25	I .36, XII -.49, XIX .20
EI	Liability Evaluation	.25	III -.24, XVII .20
IB	Directing Skill	-.21	I .72, VII .20, X -.21
WEPS		-.25	III -.24, VIII .50, XVII -.27, XIX -.26
EPPS	Need for Dominance	-.37	VII .46, XVII -.39, XVIII -.23

Table 34
(Page 3 of 4)

Factor and Scales		Loadings	Loadings on Other Factors
Factor XIV			
WE	Spelling	.79	
LQS	Decisionmaking	.32	X -.21, XV -.27, XVI .27, XVIII -.32
AI	Self-confidence	.26	IV .44, VII .43, XI .23
WE	Accuracy	.25	XI -.39, XIX -.49
LOQ	Structure	.21	II -.45, VI .22, IX .25, X .37
ALGE	Part. Leadership	-.20	IV .63, VIII .23, XIX .30
EPPS	Need for Abasement	-.24	II -.20, III -.21, XVIII .63
CONG.	Part., Popularity	-.27	II .23, VI .48, VII .33
IB	Planning & Organization	-.36	I .61, VIII -.29, X .21, XIII .29, XV .21
Factor XV			
EPPS	Need for Affiliation	.81	III .21
IB	Noting Racial Implications	.57	I .21, XVI .37
SIM	CAR Admin. Skills	.38	II .47, VI .26, XVII -.22
EPPS	Need to Give Nurturance	.28	III -.55, XVIII .27, XIX -.37
EI	Creativity	.21	III .23, V .64
IB	Planning & Organization	.21	I .61, VIII -.29, X .21, XIII .29, XIV -.36
EPPS	Need for Order	-.26	VIII .24, XI .34, XX .62
LQS	Decisionmaking	-.27	X -.21, XIV .32, XVI .27, XVIII -.32
Factor XVI			
CONG	Part., Avoiding Conflict	.76	XI -.20
IB	Sensitivity	.58	I .24, III .27
IB	Noting Racial Implications	.36	I .21, XV .57
LQS	Decisionmaking	.27	X -.21, XIV .32, XV -.27, XVIII -.32
IB	Initiative	.24	I .72
EPPS	Need for Achievement	.21	VI -.23, VIII -.23, XI -.24, XVIII -.24, XIX .51
Factor VII			
WE	Completeness	.77	
EPPS	Need for Heterosexuality	.30	III .36, VII -.20, X .21, XI -.25, XVIII -.33, XX -.21, XXI -.30

Table 34
(Page 4 of 4)

Factors and Scales		Loadings	Loadings on other Factors
Factor VII			
EI	Fluency	.23	V .44, VII .52, XII .34
EI	Interest Range	.22	V .55, VI .48, VIII -.24
EI	Liability Evaluation	.20	III -.24, V .51, XIII .25
SIM	CAR Admin Skills	-.22	II .47, VI .26, XV .38
WEPS		-.27	III -.24, VIII .50, XIII -.25, XIX -.26
IB	Written Communication	-.36	I .44, V -.20
EPPS	Need for Dominance	-.39	VII .46, XIII -.37, XVIII -.23
Factor XIX			
EPPS	Need for Exhibitionism	.68	VII .21, XX -.20
EPPS	Need for Achievement	.51	VI -.23, VIII -.23, XI -.24, XVI .21, XVIII -.24
ALGE	Part., Social Association	.49	I .24, II .21, IV .20, VI .31, XII .44
ALGE	Part., Leadership	.30	IV .63, VIII .23, XIV .20
LGD	Part., Overall Effectivensss	.22	II .30, IV .28, VII .45, XI -.22
IB	Self-confidence	.20	I .36, XII -.49, XIII .25
EPPS	Need for Succorance	-.22	XI .22, XX -.44, XXI -.44
WEPS		-.26	III -.24, VIII .50, XIII -.25, XVII -.27
EPPS	Need to Give Nurturance	-.37	III -.55, XV .28, XVIII .27
WE	Accuracy	-.49	XI -.39, XIV .25

Key to Abbreviations:

AI	Appraisal Interviewer	LOQ	Leadership Opinion Questionnaire
ALGE	Assigned Leader Group Exercise	LQS	Leadership Q Sort
CONG	Conglomerate	ND	Nelson Denny Reading Test
EI	Entry Interview	SIM CAR	Simulate: Company Assessor Report
EPPS	Edwards Personnel Preference Schedule	SIM PAR	Simulate: Platoon Assessor Report
IB	In-Basket	WE	Written Expression
LGD	Leaderless Group Discussion	WEPS	Work Environment Preference Schedule

The dimension involved here seems to be one of interpersonal competence and sociability along with a breadth of interest and willingness to give somewhat forceful structured leadership in a relatively ambiguous situation.

Factor VII. Of all the factors, this one comes closest to representing general interpersonal leadership. While it has the largest number of loadings above .20 and has loadings from seven of the eight exercises as well as from paper and pencil tests, the loadings point rather clearly to personal impact as an emergent leader. Only the negative loadings on EPPS Heterosexuality and LQS Leadership Potential are puzzling. Certainly, the latter is not indicative of the strong personal leadership which this factor represents. A person who would score highly on a pure measure of this factor would seem to want to dominate, be verbally fluent, forceful, and self-confident. He would also tend to be decisive, have directing skill, and be a little exhibitionistic. It is interesting that he would be no more or less likely than a low scorer to be chosen as a friend, to be worried about making a negative impression, or to create conflict in group situations.

Factor XI. In contrast to the personal ascendance shown by Factor VII, this factor appears to be one of followership. The major loading is on EPPS need for deference. A person scoring high on this scale is described in the test manual as desiring, "to get suggestions from others,...to follow instructions and do what is expected,...to accept the leadership of others,...to let others make decisions." The EPPS need-for-order scale suggests an additional need to plan and organize and to avoid change, and the EPPS need-for-succorance scale suggests a desire for help, encouragement, and sympathy from others.

On the behavioral side, the person scoring high on this factor tends to avoid conflict in the Conglomerate but is judged as not very effective in the Leaderless Group Discussion. He also exhibits poor social skills in the Simulate and a lack of sense of humor in the Entry Interview. His lack of accuracy in the Writing Exercise and his self-confidence in the Appraisal Interview are puzzling in light of the previous discussion of his self-description on the EPPS scales. Perhaps these scores can be explained by the fact that in the AI he can perform effectively as an interviewer by showing deference to the interviewee. On the other hand, he has to work on his own in the WE, and also to explain and justify a negative personnel action.

Factor XII. This factor is described entirely by behavioral ratings rather than self-description on paper-and-pencil tests. Though there are few substantial loadings, they suggest a tendency to avoid interpersonal conflict. The behavior appears general since the five loadings come from five separate exercises. The high scorer on this factor goes out of his way to avoid making a negative impression on others in the LGD. He likewise avoids the potential conflicts involved with emergent leadership in the Conglomerate exercise. On the In-Basket he appears to lack self confidence, probably stemming from a desire to avoid conflict with those responsible for the problems he encounters there. On the positive side, he is verbally fluent and makes a positive impression on

the other participants in the ALGE, where he is chosen for social association but not for leadership. The only loadings which are missing in confirmation of the above interpretation of Factor XII are on avoiding conflict in the Conglomerate and sensitivity on the In-Basket. However, it may be that these scores represent positive social actions whereas this factor simply represents a tendency to avoid being involved in conflicts.

Factor XIII. This factor seems to represent basic intellectual or problem-solving ability. Such an interpretation is consistent with the intellectual jugdment required by the Leadership Q Sort, planning and organization in the In-Basket, or a rational evaluation of strengths and weaknesses in the Entry Interview. The high loading on SIM supervisory skills might suggest something more in the area of interpersonal skills. It must be remembered however, that this scale simply represents the global impression of the platoon assessor. It is very probable that the platoon assessor was impressed by calmness and clear thinking in handling the many problems the assessee faced in the Simulate. It is also probable that self confidence on the In-Basket arises from the assessees' ability in analyzing the problems encountered, reaching a solution, and acting on it.

It is also consistent with the above interpretation that a person who scored high on this dimension would be less comfortable in a highly bureaucratic organization, as is indicated by the negative WEPS loading. This and the additional negative loadings in directing skill and in dominance suggest that Factor XIII is a dimension with rational problem solving at one end and bureaucratic dominance at the other.

Factor XIV. A clear interpretation of this factor is not readily apparent. On the positive side, a high scorer on this dimension is accurate in his facts and his spelling on the writing exercise, he has self confidence as shown by AI, he emphasizes the decisionmaking function of leadership, and is structuring in his leadership approach. On the other hand, he is not seen as an effective leader by his peers in ALGE or in the Colglomerate exercise. He also fails to plan and organize adequately in the In-Basket. The negative loading on EPPS need for Abasement may suggest a tendency for a high scorer to fail to show needed humility in his leadership approach. The latter however, is a very tentative speculation. The overall impression, then, suggests a tendency to grab a problem, perhaps without looking at all the facts or other related problems, and to make a quick decision expecting others to take unquestionning action on that decision. Other data would be required to confirm that interpretation.

Factor XV. This factor appears to represent concern for interpersonal relations. The high loading on EPPS need for Affiliation indicates that a high scorer on this factor would desire close friendly relationships. He is also concerned about racial discrimination and somewhat nurturant in his interpersonal approach. He is seen as possessing administrative skills and creativity. He also is rated as somewhat above average in planning and organization. The planning and organizing rating from the In-Basket may stem largely from his concern

for the racial implications of the In-Basket problems. He also has a lower than average need for order and tends somewhat to downgrade decisionmaking as a leadership function.

Factor XVI. This factor represents maintenance of harmonious interpersonal relations, as did in different ways Factors XV, XII, and to some extent XI. There is personal need for friendship as shown in Factor XV, no indication of avoiding problem confrontation as shown in Factor XII, and no indication of deference or followership shown in Factor XI. This factor appears to be one of objective, positive, interpersonal problem solving. A high scorer on this factor avoids conflict in the Conglomerate exercise, is sensitive to human feelings in the In-Basket, and also is aware of possible racial implication in the problems of the In-Basket. Unlike the high scorer on Factor XV, he sees decisionmaking as an important leadership function. He takes initiative on the In-Basket problems and seeks achievement through successful problem resolution.

Factor XVII. The loadings on this factor do not lend themselves to a clear and unified interpretation. One tentative interpretation suggests that this factor represents verbosity. Given the problem of the Writing Exercise, the high scorer on this dimension produces a very complete (though no more accurate than average) document. On the In-Basket, which requires many memos, letters, etc. which are brief and to the point, his written communication is poorer than average. This interpretation is not incongruent with the positive EI scale loadings, nor with the negative loadings on the WEPS comfort with bureaucracy scale and the EPPS need for Dominance scale. It is also likely that a verbose person would tend not to have good administrative skills. As with some other factors, additional data would be required to confirm this tentative interpretation.

Factor XIX. This factor seems to represent a desire to be noticed and to hold center stage. Unlike Factor VII, it does not appear to represent a desire to lead or to solve problems. EPPS need for Exhibition involves desire to say witty and clever things, say things just to see the effect it will have on others, and be the center of attention. EPPS need for Achievement involves the desire to be a recognized authority, to accomplish something of great significance, to be able to do things better than others, to write a great novel or play. High scorers on this factor seemed to succeed in impressing their peers in ALGE and the LGD, but more strongly in terms of choice as a friend than for leadership effectiveness. They did not particularly succeed in impressing the assessors in these same exercises. They do tend to show self-confidence in dealing with the In-Basket. They have a lower than average need to give nurturance as well as to receive succorance, as shown by their EPPS scores. They also tend to be somewhat uncomfortable in bureaucratic organizations and to produce inaccurate written documents. The latter tendency may stem from a desire to produce a colorful or witty report.

FACTORS PRIMARILY DEFINED BY PAPER-AND- PENCIL TESTS (TABLE 35)

Factor III. The loadings on this factor are from a wide variety of paper-and-pencil test scales as well as a few behavioral ones. Only a tentative interpretation can be offered for the nature of the underlying dimension. On the positive end it appears to be one of desire for new and exciting experiences. Such an interpretation is supported by the high need for change and the interest in heterosexual activities. It is also congruent with the Entry Interview rating on creativity which probably represents a novel of looking at things and an expressed interest in novel things. The positive loading on sensitivity in the In-Basket is a little puzzling but may simply represent an approach to the In-Basket materials which went beyond the conventional surface problems to some of the subtler problems involving human feelings.

On the negative end, there seem to be two themes. One theme relates directly to the interpretation of the positive end of the scale and involves resistance to the constraints of rational conventionality. This interpretation is suggested by a negative loading on WEPS, and supported, though not strongly, by the negative loadings on LQS Leadership Potential and EI Liability Evaluation. The second theme is a disinterest in close interpersonal ties or in personal service to others. This is suggested by the negative loadings of need for Affiliation, need to give Nurturance, and the social skills scale from the Simulate. The negative loading on EPPS need for Abasement is not incongruent with the interpretations given.

Factor VIII. This factor's two highest loadings are from the Leadership Q Sort. A person scoring high on a pure measure of the factor would place high value on leadership statements dealing with mental health and downgrade the value of technical information. The positive loadings on WEPS and EPPS need for Order suggest that this may be a factor or authoritarianism or dogmatism. This interpretation is not incongruent with the negative loadings on EPPS need for Achievement. On the behavioral side, the authoritarian or dogmatic interpretation also receives some support from the positive loading on IB Planning and Organization and the negative loading on EI Interest Range. Also, a high scorer on this factor tends to be rated positively by his peers on leadership in the ALGE.

Factor X. The highest loading on this factor, LQS Teaching and Communication, indicates that it represents leadership as a helping or facilitating function. This interpretation is consistent with the positive loading on LOQ structure, a tendency to define jobs and relationships and to arrange and manage working assignments and conditions. This interpretation is further supported by the positive loading on LQS Consideration and the negative loading on LQS Decisionmaking. On the behavioral side, it is supported by the loading of IB Planning and Organization. The loading on EI Sense of Humor is not incongruent, but the negative loading on IB Directing Skill is a little puzzling. Perhaps a person with this orientation is not seen as forceful enough by the IB raters. The EPPS need for Heterosexuality loading adds little to the interpretation. For this factor, at least, LOQ Structure is consonant with

Table 35
(Page 1 of 2)

FACTORS FROM OVERALL ANALYSIS DEFINED PRIMARILY BY
LOADINGS FROM PAPER-AND-PENCIL TESTS

Factors and Scales		Loadings	Loadings on other Factors
Factor III			
EPPS	Need for change	.79	VI -.21, XI -.22
EPPS	Need for Heterosexuality	.36	VII -.20, X .21, XI -.25
IB	Sensitivity	.27	I .24, XVI .58
EI	Creativity	.23	V .64, XV .21
SIM	CAR Social Skills	-.21	II .52, XI -.31, XXI .22
EPPS	Need for Abasement	-.21	II -.20, XIV -.24, XVIII .63
EPPS	Need for Affiliation	-.21	XV .81
WEPS		-.24	VIII .50, XIII -.25, XVII -.27, XIX -.26
EI	Liability Evaluation	-.24	V .51, XIII .25, XVII .20
LQS	Leadership Potential	-.29	V .34, VI .26, VII -.33, XIII .37, XXI -.22
EPPS	Need to give Nurturance	-.55	XV .28, XVII .27, XIX -.37
Factor VIII			
LQS	Mental Health	.79	
WEPS		.50	III -.24, XIII -.25, XVII -.27, XIX -.26
IB	Planning & Organization	.29	I .61, X .21, XIII .29, XIV -.36, XV .21
EPPS	Need for Order	.24	XI .34, XV -.26, XX .62
ALGE	Part., Leadership	.23	IV .63, XIV -.20, XIX .30
EPPS	Need for Achievement	-.23	VI -.23, XI -.24, XVI .21, XVIII -.24, XIX .51
EI	Interest Range	-.24	V .55, VI .48, XVII .22
LQS	Technical Information	-.52	V .23, XVIII -.35
Factor X			
IQS	Teaching & Communication	.79	
LOQ	Structure	.37	II -.45, VI .22, IX .25, XIV .21
EI	Sense of Humor	.33	III -.23, V .49, VII .34, XI -.25
LQS	Consideration	.23	XVIII .76
IB	Planning & Organization	.21	I .61, VIII .29, XIII .29, XIV -.36, XV .21
EPPS	Need for Heterosexuality	.21	III -.36, VII -.20, XI -.25, XVII .30 XVIII -.33, XX -.21, X -.30
LQS	Decisionmaking	-.21	XIV .32, XV -.27, XVI .27, XVIII -.32
IB	Directing Skill	-.21	I .72, VII .20, XIII -.21

Table 35
(Page 2 of 2)

Factors and Scales		Loadings	Loadings on Other Factors
Factor XVIII			
LQS	Consideration	.76	X .23
EPPS	Need for Abasement	.63	II -.20, III .21, XIV -.24
EPPS	Need to Give Nurturance	.27	III -.55, XV .28, XIX -.37
EPPS	Need for Dominance	-.23	VII .46, XIII -.37, XVII -.39
EPPS	Need for Achievement	-.24	VI -.23, VIII -.23, XI -.24, XVI .21, XIX .51
ND	Reading Comprehension	-.30	II .30, VII .27, XIII .49
LQS	Decisonnmaking	-.32	X -.21, XIV .32, XV -.27, XVI .27
EPPS	Need for Heterosexuality	-.33	III -.36, VII -.20, X .21, XI -.25, XVII .30, XX -.21, XXI -.30
LQS	Technical Information	-.35	V .23, VIII -.52
Factor XX			
EPPS	Need for Endurance	.76	
EPPS	Need for Order	.61	VIII .24, XI .34, XV -.26
EPPS	Need for Exhibition	-.20	VIII .21, XIX .68
EPPS	Need for Heterosexuality	-.21	III .36, VII -.20, X .21, XI -.25, XVII .30, XVIII .33, XXI -.30
EPPS	Need for Succorance	-.44	XI .22, XIX -.22, XXI -.44
Factor XXI			
EPPS	Need for Intraception	.80	
SIM	CAR SOCIAL Skills	.22	II .52, III .21, XI -.31
AI	Written Communication	-.20	IX .75
LQS	Leadership Potential	-.22	III .29, V .34, VI .26, VII -.33, XIII .37
EPPS	Need for Heterosexuality	-.30	III .36, VII -.20, X .21, XI -.25, XVII .30 XVIII .33, XX -.21
EPPS	Need for Succorance	-.44	XI .22, XIX -.22, XX -.44

Key to abbreviations:

AI Appraisal Interview
ALGE Assigned Leader Group Exercise
CONG Conglomerate
EI Entry Interview
EPPS Edwards Personal Preference Schedule
IB In-Basket
LGD Leaderless Group Discussion
LOQ Leadership Opinion Questionnaire
LQS Leadership Q Sort
ND Nelson Denny Reading Test
SIM CAR Simulate: Company Assessor Report
SIM PAR Simulate: Platoon Assessor Report
WE Written Expression
WEPS Work Environment Preference Schedule

LQS Consideration and not with LQS Decisionmaking. Further exploration of these relationships might lead to a clarification of the meaning of structuring. It is disappointing that scores from the interpersonal exercises did not load on this factor.

Factor XVIII. All the loadings on this factor come from three paper-and-pencil tests. The factor seems to be one of placing other people above oneself in value and emphasizing self-effacing service. A high scorer on this factor highly values consideration as a leadership function and has a fairly high need for self-abasement along with a moderate need to give nurturance. He tends to downgrade the importance of decisionmaking and technical information to a leader and to be somewhat lacking in verbal intelligence and need for dominance, achievement, and heterosexuality. The whole picture suggests a relatively poor self-image, possibly stemming from a real disadvantage in ability. Since the factor has no behavioral exercise loadings, its value is questionable.

Factor XX. All the loadings on Factor XX come from the Edwards Personal Preference Schedule. The factor seems to indicate an orientation toward the dogged completion of work in an organized, planned and isolated fashion. The positive loadings, both of which are fairly high, are on EPPS need for Endurance, defined in the test manual as indicating desire to "work hard at a task,... put in long hours of work without distraction,... stick at a problem even though it may seem as if no progress is being made...," and EPPS need for Order, defined as desire to "to have written work neat and organized,... organize details of work,... have things arranged so that they run smoothly without change." The negative loadings, on need for Exhibition, Heterosexuality, and Succorance, suggest a lack of desire for attention, recognition or help from other people of either sex. It may be that the short time span of individual tasks in the Assessment Center may account for the lack of loadings from behavioral scales. It would be interesting to find out what behavioral measures outside the Assessment Center correlate with this factor, and with the preceding factor.

Factor XXI. This factor is represented by a high need for Intraception on the one hand and a low need for Succorance on the other. According to the EPPS test manual need for Intraception is characterized by desire "to analyze one's motives and feelings, to observe others,... to analyze the behavior of others,... to predict how others will act." Need for Succorance is defined by desire "to have others provide help when in trouble,... to have others be kindly,... to receive a great deal of affection from others,... to be helped by others when depressed,..." Apparently this factor represents a strong need to understand others, below average need to be helped by them, and somewhat below average heterosexual drive. The person scoring high on this dimension also tends to assign values to statements about leaders which are different from those assigned by successful leaders. He does tend to display somewhat better than average social skills in the Simulate exercise. His lower than average score on written communication in the Appraisal Interview may stem from his being more concerned with analyzing the feelings and motives of those he interviewed than with determining the extent to which their capabilities matched the demands of the position to be filled.

SUMMARY

The 21 factors extracted from the overall analysis seem to fall into four categories. These correspond roughly to the three groupings used in the foregoing discussion, with the second category divided into two groups--factors which represent task performance and those which represent interpersonal relations.

Six of the factors were largely determined by a single behavioral exercise. Concerning two factors, I--In-Basket and IX--Written Section of Appraisal Interview, litle could be said except that the factor represented performance in that exercise. For the other four factors, however, other loadings allowed richer interpretation of what was involved in performance on the key exercises. Factor II, performance in the Simulate was characterized by competent but easy-going social interaction skill. Factor IV, performance in the ALGE is characterized by self-confidence and personal impact. Factor V, performance in the Entry Interview represents congruence of expressed goals and values with those judged appropriate for a career as an Army officer. Factor VI, performance in the Conglomerate exercise, is characterized by social skill, breadth of interest, and providing structure in an ambiguous situation.

Four factors defined by loadings from several exercises, resented primarily a task orientation. Factor VII, the broadest and perhaps most significant, was a factor of personal impact including verbal intelligence and fluency, self-confidence, forcefulness, and desire for dominance. Factor XIII appeared to be a factor of calm, rational, competent problem solving. Factor XIV was tentatively interpreted as representing a tendency to reach a quick solution to a problem while failing to see its relationship to other problems. Factor XVII was tentatively identified as representing verbosity.

It is significant that no less than five factors defined by several exercises involved some kind of concern for interpersonal relations and had a substantial behavioral component. Factor XI seemed to represent deference and following behavior. Factor XII indicated an orientation towards avoiding interpersonal conflict and confrontation. Factor XV seemed to involve a personal desire for friendship and service to others. Factor XVI, in contrast to factor XV, seemed to indicate orientation toward objective interpersonal problem solving. Factor XIX involved colorful exhibitionism.

The final group of factors all represented orientations or values which emerged largely from self descriptions on paper-and-pencil tests. While these factors had little or no behavioral component from the Assessment Center, it is quite possible that they could be related to other behavioral measures such as performance in BIOCC or in duty assignments. Factor III was interpreted, tentatively, as representing a need for new and exciting experiences. Factor VIII seemed to represent a bureaucratic orientation to leadership. Factor X seemed to represent a belief that the main function of leadership is teaching, organizing, and arranging things for followers. Factor XVIII was interpreted as a self-effacing service orientation possibly stemming from a poor self concept.

Factor XX represented an orientation towards planning and organizing work and completing it in a dogged isolated fashion. Factor XXI appeared to represent a desire to understand oneself and others, but not to form close relationships.

All in all, the factors extracted demonstrated that the Assessment Center measured a broad spectrum of behavior.

Chapter 11

RELATIONSHIP OF MEASURED DIMENSIONS TO MEASUREMENT GOALS

It is appropriate to ask the extent to which the measurement goals were realized in the assessment process. During the development phase, 12 characteristics were selected as appropriate and measurable dimensions of leader behavior. After factor analysis of individual exercises and of representative scales from the entire asssessment process, dimensions appeared to have been measured. Each of the twelve dimensions listed in Chapter 3 will be considered relative to the factorial dimensions discussed in chapters nine and ten. Due to the nature of the analyses conducted, this discussion will consider only the BIOCC group.

Adaptability. Adaptability appeared as a scale in the Simulate. It did not define a factor for that exercise although it was the only company-assessor rating which loaded substantially on the global platoon-assessor rating factor. Flexibility was a scale title in the Assigned Leader Group exercise. Though it loaded on two factors, its loading was not high enough for it to be selected for the overall analysis. None of the dimensions which emerged from the overall analysis could be clearly labeled as an indicator of adaptability.

Administrative Skills. One of the exercises, the In-Basket, was designed specifically to measure administrative skills. The Appraisal Interview and the Simulate also required administrative ability. The analysis of the In-Basket revealed several dimensions of behavior crucial to administrative effectiveness. In the overall analysis, the In-Basket factor was undoubtedly administrative in character. In the analysis of the Simulate, the Administrative skills rating by the company assessor was selected to represent one of the exercise dimensions in the overall analysis. The planning and organizing factor and the written communication factor, from the Appraisal Interview, together with LOQ Structure, formed a second administrative dimension in the overall analysis. Several of the task-oriented factors from the overall analysis seem to have important implications for administrative success, particularly the one called rational problem solving (Factor XIII). Two others may represent deficiencies in the administrative area. They were interpreted tentatively as premature closure (Factor XIV), and verbosity (Factor XVII).

Communication Skills. This dimension in its written and oral forms was probably the one most thoroughly studied in the Assessment Center. Every exercise, with the exception of the Conglomerate, had at least one scale representing either written or oral communication. Many of these scales were selected for inclusion in the overall analysis. Among the standardized test, four (the Henmon-Nelson Test of Mental Ability, the Nelson-Denney Reading Test, the Social Insight Test, and the Watson-Glaser Critical Thinking Appraisal) clustered together as measures of verbal intelligence, which contributes to communication skill. In the overall analysis, the scales representing communication skills did not come together to form a single dimension. They did, however, make important contributions to several of the most significant behavioral dimensions.

Decisionmaking Skills. Four exercises used scales called variously decisionmaking (Simulate, both platoon and company assessors), decisiveness (In-Basket and Assigned Leader Group Exercise), and decision quality (In-Basket and Conglomerate Exercise). Three of these scales were selected to represent exercise dimensions in the overall analysis. These three scales loaded very highly on the exercise-specific factors for the In-Basket, Simulate, and Assigned Leader Group Exercise. The decisiveness scale from ALGE also had a moderate loading on Factor VII, the strongest cross-exercise factor, representing inter-personal leadership. Apparently, decisionmaking skills were important at the Assessment Center but tended to be tied to the situation in which they were elicited.

Forcefulness. It is perhaps in the area of forcefulness or interpersonal impact that the Assessment Center has its greatest advantage over more traditional measurement techniques. Among the individual exercises a number of scales appear to be closely related to forcefulness: (1) Forcefulness in the Simulate as evaluated by both platoon and company assessor, the latter's evaluation chosen for inclusion in the overall analysis; (2) Self Confidence in the Appraisal Interview, chosen for the overall analysis; (3) presentation Impact in the Leaderless Group Discussion, chosen for the overall analysis; (4) Energy and Vigor in the Conglomerate Exercise; (5) Need to Dominate from the Edwards Personal Preference Schedule, chosen for the overall analysis. All the measures that went into the overall analysis contributed flavor to the interpersonal leadership factor, Factor VII. This was the strongest cross-exercise factor. It appears that the goal of measuring forcefulness was fully realized.

Mental Ability. As stated above, under communication skills, no less than four of the standardized tests seemed to be measuring verbal intelligence. In addition, several scales in the administrative exercises assessed specific applications of mental ability. Participants were asked to rate idea quality in the Leaderless Group Discussion and in the Conglomerate Exercise. Neither of these participant ratings, however, was selected for the overall analysis. In the overall analysis the Nelson-Denney Reading Comprehension score, which represented the verbal intelligence factor, loaded moderately on Factor VII, the strong interpersonal leadership factor. It loaded more highly on Factor XIII termed rational problem-solving. It appears that the Assessment Center measured verbal intelligence as a personal attribute on the paper-and-pencil tests and also as it was applied in situational tasks.

Motivation. A number of attempts were made to assess motivation either directly or indirectly. Eight scales from six exercises could be considered more or less direct indicators of motivation. These were: (1) Task Orientation and Task Motivation in the Entry Interview; (2) Task Orientation in the In-Basket; (3) Motivation in the Simulate both platoon and company assessor ratings; (4) Participation in the Leaderless Group Discussion; (5) Motivation in the Assigned Leader Group Exercise; (6) Energy and Vigor in the Conglomerate Exercise. None of these scales were selected to represent individual exercise factors in the overall analysis. In the individual exercise analyses, it can be

seen that these scales usually made contributions to several dimensions. It appears that motivation was important but that it made its contribution through more readily observable behavior styles and task accomplishments.

Organizational Leadership. This dimension represents quality of performance as an assigned leader in a structured situation. Two exercises were specifically designed to measure this dimension, the In-Basket and the Simulate. While the Assigned Leader Group Exercise had, as its name implies, a strong assigned leadership component, it had little structure and few organizational constraints.

The Simulate for the BIOCC group lacked, by design, many organizational constraints. The assessors felt unable in a large number of cases to rate organizational leadership. Therefore, these scales were dropped from the analysis. Two paper-and-pencil tests, the Leadership Opinion Questionnaire and the Leadership Q Sort, were designed specifically to measure self described leadership styles and attitudes.

While the In-Basket produced an exercise specific factor in the overall analysis, thus precluding any separation of administrative skills from organizational leadership, several IB scales contributed to other factors as well. IB Directing Skills loaded to some extent on Factor VII, the interpersonal leadership factor. IB Self-Confidence formed the opposite pole to Factor XII, avoiding interpersonal conflict. IB Planning and Organization and Self-Confidence, along with Directing Skill negatively loaded, helped to define Factor XIII, called rational problem solving. IB Planning and Organization also formed the opposite pole to Factor XIV, tentatively called premature closure in decision-making. IB Sensitivity and Noting Racial Implications also made important contributions to Factors XV and XVI representing interpersonal styles.

Several of the dimensions from the overall analysis, defined primarily by paper-and-pencil tests also deal with orientations toward organizational leadership. External validation may show that some of these are important. It appears, therefore, that several factors associated with organizational leadership were measured at the Assessment Center.

Physical Fitness. The BIOCC group was given a physical conditioning exercise each morning. This was designed primarily to help simulate the rigor and physical demands of the BIOCC course. This, it was hoped, would improve the predictive validity of other measures obtained at the Center. The scores on these physical exercises were not included in the analysis.

None of the other asssessment groups participated in these exercises. The only other exercise with significant physical demands was the Assigned Leader Group Exercise. The physical Ability scale from the exercise was not chosen for the overall analysis although Stress Tolerance was included. The Stress Tolerance scale did not load on any factor other than the exercise specific ALGE factor. In general,

physical fitness did not receive specific measurement attention at the Assessment Center. It did provide background for other measures in that the entire assessment experience was intensive and rigorous. Opportunity for sleep was limited and little time was allowed for rest and relaxation.

Social Skills. Along with forcefulness, social skills is an area for which the Assessment Center had unique measurement strength. Several of the exercises had scales relating more or less directly to social skills. The In-Basket had two scales in the area of social skills--sensitivity to the needs of others and noting the racial implications of problems. Both these scales were selected for the overall analysis. In the Simulate, social skills was a global rating by both platoon and company assessors. The latter was selected for the overall analysis. The Leaderless Group Discussion had assessor ratings of social concern and avoiding a negative impression. These scales formed a factor, and the latter scale was selected to represent it in the overall analysis. The LGD contained a participant rating on sociability which was not distinguishable from the other participant rating scales. The Assigned Leader Group Exercise participant rating on desirability for social association was selected to represent an exercise dimension in the overall analysis. The Conglomerate had an assessor rating of sensitivity to the feelings of others, not chosen for the overall analysis. The Conglomerate had three participant ratings in the social skills area--popularity, avoiding conflict, and working with peers. The first of these was selected to represent the global participant rating factor, and the second represented a unique factor from the exercise in the overall analysis. In addition, most of the scales of the Edwards-Personal Preference Schedule represented needs for certain kinds of relationships with others. About nine such scales entered the overall analysis. The Social Insight Test, however, appeared to measure only verbal intelligence.

No, less than five of the cross-exercise factors and several of the factors defined primarily by paper-and-pencil test scores seemed to fall in the area of social skills or orientations. The goal of measuring social skills seems to have been well met.

Supervisory Skills. Adequate measurement of much of what is important in the area of supervisory skills would probably require a longer term relationship between supervisor and subordinate and more meaningful work than can be provided in a three-day assessment program. However, the In-Basket and the Simulate did provide for role-played supervision in carefully created situations. The Assigned Leader Group Exercise provided for less structured supervision, but with actual subordinates in the form of other assessees. From the In-Basket, the directing skill and supervision scales formed a dimension represented by directing skills scale in the overall analysis. In the Simulate the company assessor rating of supervisory skills was dropped from the analysis because the assessors did not feel they had observed enough supervisory behavior to give a realistic rating. The Simulate platoon assessor's rating of supervisory skills, on the other hand, had the highest loading on the platoon assessor global factor. It was selected to represent

this exercise factor in the overall analysis. In the Assigned Leader Group Exercise, supervision was not rated directly. The ratings of performance as assigned leader clustered together in a single dimension which was represented in the overall analysis by the rating of decisiveness. Supervision was probably not measured directly in the Assigned Leader Group Exercise.

None of the factors which emerged from the overall analysis could be labeled as strictly representing supervisory skills. Both the In-Basket and the Simulate produced exercise specific factors. Simulate rating of supervisory skills was quite closely related to Factor XIII, rational problem solving. Unfortunately, In-Basket directing skill loaded negatively on this factor, though not highly. In-Basket directing skill was related slightly to the interpersonal leadership factor, Factor VII. It was also slightly negatively related to Factor X which was primarily defined by paper-and-pencil test scores and represented a teaching or helping approach to leadership.

Technical and Tactical Competence. For the BIOCC group this dimension was not tapped, due to the wide range of Army training and experience found among BIOCC candidates. With the ANCOES and IOAC groups, the Simulate provided an opportunity for evaluation of this dimension. Such evaluation was not, however, a major goal of the assessment process, which was designed primarily for the other attributes listed above. It would probably be wise to leave the major burden for measuring this skill area to the Army schools and to more elaborate and specifically designed technical and tactical training and assessment exercises.

SUMMARY

The twelve dimensions selected for measurement at the Assessment Center can be separated roughly into four categories in terms of the success with which they were actually measured. These categories are: (1) not measured or not measured adequately, (2) not measured directly, but only by inference from more directly observable behavior, (3) measured, but in a manner needing some refinement or improvement in the instruments, and (4) those measured very well. The fourth category constitutes the real strengths of the assessment process.

Two of the 12 dimensions were not measured or not measured well by the Assessment Center: physical fitness and technical and tactical competence. Both could have been measured more accurately without a great deal of difficulty.

Physical fitness was felt to be a background factor rather than a subject for direct measurement since the assessed groups were all expected to meet fairly high minimum standards of fitness. Technical and tactical competence, as mentioned above, is probably better left to specific training and proficiency exercises in other settings.

An additional two dimensions, though the subject of various exercise rating scales, did not emerge as significant factors either in exercise analyses or in the overall analysis. These were adaptability and motivation. It seems reasonable to conclude that these characteristics tend not to be shown directly, but must be inferred from a combination of other traint and performance in specific situations. While exercises could probably be designed in which successful performance required either motivation or adaptability and little else, such exercises would probably be quite artificial. A more efficient approach would be to direct attention in assessment to characteristics with more directly observable behavior referents.

The next four dimensions to be discussed were measured quite well in the assesment process: administrative skills, decisionmaking, organizational leadership, and supervisory skills. In the case of each of these dimensions, however, it seems that some revisions in exercises or the creation of new exercises might give more precise measurement. Such revision might allow cleaner separation of these dimensions and reduce reliance on global exercise factors.

The success of measurement of the final four dimensions illustrates the power of the assessment process. None of these dimensions seems capable of really adequate measurement with alternative testing procedures. Although mental ability can be measured with standardized paper-and-pencil tests, the assessment process shows where measured intelligence counts and where it doesn't in actual performance. The success of the assessment process in delineating several social skill dimensions, in contrast to the failure of the Social Insight Test to measure social skills, again illustrates the strength of the assessment process. Both forcefulness and communication skills were measured in ways that seem impossible without the actual behavioral observation opportunities the assessment process provides.

Overall, the Assessment Center seems to have achieved the measurement goals outlined during its development. It remains to be seen whether this achievement can lead to the accomplishment of the longer range goal of predictive validity and leadership development.

APPENDIXES

Appendix			Page
A.	PRELIMINARY FACTOR ANALYSES ON ASSESSMENT GROUP EXERCISES		
Tables	A- 1.	Factor Analyses of Entry Interview Rating Scales	97
	A- 2.	Factor Analyses of In-Basket Rating Scales	98
	A- 3.	Factor Analysis of Writing Exercise Rating Scales	99
	A- 4.	Factor Analysis of Simulate, Global Rating Scales	100
	A- 5.	Factor Analyses of Appraisal Interview Rating Scales	101
	A- 6.	Factor Analysis of Leaderless Group Discussion Rating Scales	102
	A- 7.	Factor Analysis of Assigned Leader Group Exercise Rating Scales for IOBC, BIOCC, ANCOES, and ROTC Groups	103
	A- 8.	Factor Analysis of Leader War Game Rating Scales for IOAC and 197th Inf. Bde. Groups	104
	A- 9.	Factor Analysis of Conglomerate Exercise Rating Scales for all Assessment Groups Combined	105
	A-10.	Factor Analysis of Standardized Paper and Pencil Test for All Assessment Groups Combined	106

Table A-1

FACTOR ANALYSES OF ENTRY INTERVIEW
RATING SCALES

Rating Scales	IOAC					IOBC				ANOCES				
	I	II	III	IV	V	I	II	III	IV	I	II	III	IV	V
1. Conveys Information	.44			.56	.39	.36			.73				.74	.30
2. Sense of Humor	.83					.77				.77				
3. Verbal Fluency	.53			.38	.45				.80				.74	
4. Enthusiasm	.76					.69				.77				
5. Expresses Opinions					.71	.41			.49	.42			.45	.40
6. Interest Range				.61	.39				.69				.34	.50
7. Task Orientation			.64	.30		.39		.32	.51	.32		.47		.41
8. Asset Evaluation		.77					.88				.66			.37
9. Liability Evaluation		.87					.82				.73			
10. Goal Congruence			.64				.34	.63			.37	.65		
11. Self-Development		.40		.53				.35	.43					.64
12. Task Motivation			.63	.39	.37		.33	.80				.66	.32	.31
13. Creativity					.67			.65						.51
14. Overall Impression	.52		.32	.46	.47	.63		.34	.49	.60		.40		.44
	N = 53 Prop. of var. acct. for =.66					N = 60 Prop. of var. acct. for =.63				N = 87 Prop. of var. acct. for =.63				

Note: Factor loadings below .30 in absolute magnitude have been excluded to promote readability.

PRECEDING PAGE NOT FILMED
BLANK

Table A-2

FACTOR ANALYSES OF IN-BASKET
RATING SCALES

Rating Scales	Battalion Commander (IOAC & 197th Inf. Capt.) Factors				Company Commander (BIOCC, IOBC, ROTC) Factors			First Sargeant (ANCOES) Factors		
	I	II	III	IV	I	II	III	I	II	III
1. Written Communication		.39					.42		.30	
2. Planning & Organization	.49	.39		.37		.68			.72	
3. Problem Analysis	.45	.47	.36	.45	.40	.64		.40	.68	
4. Decision Quality	.65	.38	.44		.67	.37		.70	.38	.39
5. Directing Skills	.56	.47			.44	.40	.38	.65	.36	
6. Supervision	.67		.32		.53	.30	.45	.67		
7. Use of Information		.66	.38		.31	.64		.42	.62	
8. Attention to Detail	.38	.57	.37	.33	.31	.66	.37	.39	.75	
9. Sensitivity			.59				.33	.49		
10. Task Orientation	.67				.60	.41	.34	.66		
11. Self Confidence		.64				.54	.34	.41	.34	
12. Decisiveness	.67	.43			.66		.42	.79		
13. Working with Superiors	.37	.31	.47		.30	.31	.67	.55	.40	
14. Initiative	.45	.36	.59		.52		.64	.61	.38	.48
15. Seeing Racial Implications				.51			.40			.54
	N = 86 Prop. of Var. Acct. for =.57				N = 141 Prop. of Var. Acct. for =.53			N = 87 Prop of Var. Acct. for =.54		

Note: Factor loadings below .30 in absolute magnitude have been excluded to promote readability.

Table A-3

FACTOR ANALYSIS OF WRITING EXERCISE
RATING SCALES

Rating Scales	IOAC & 197th		IOBC & BIOCC	ANCOES
	I	II	I	I
1. Accuracy	.46	.41	.57	.40
2. Objectivity	.69		a	a
3. Grammar		.60	.72	.63
4. Spelling		.57	.68	.47
5. Completeness	.56		.61	.32
	N = 86 Prop. of var. Acct. for =.37		N = 96 Prop. of var. Acct. for =.42	N = 87 Prop. of Var. Act. for =.22

[a] Objectivity was not scored for the IOBC, BIOCC, or ANCOES groups.

Note: Factor loadings below .30 in absolute magnitude have been excluded to promote readability.

Table A-4

FACTOR ANALYSIS OF SIMULATE, GLOBAL RATING SCALES

Rating Scales	IOAC (N = 86)							IOBC, BIOCC, ROTC (N = 141)						ANCOES (N = 87)					
	1	II	III	IV	V	VI	VII	I	II	III	IV	V	VI	I	II	III	IV	V	VI
Subordinate Assessor:																			
Social Skills			.35					.65						.51					
Communication Skills	.56	.32	.45					.74	.34					.64					.34
Adaptability	.67							.68	.41					.68					
Motivation	.66							.79						.63	.49				
Forcefulness	.76							.66						.47	.63				
Decision Making	.78							.73						.59	.38				
Administrative Skills	.62							.69						.77					
Organizational Identification				.76				.66								.73			
Organizational Role Effectiveness	.73							.70						.70					.30
Supervisory Skills	.67	.31						.73						.55				.56	
Technical & Tactical Competence				.71				.69						.46		.64			
Superior Assessor:																			
Social Skills		.47			.35	.37				.32	.40							.55	
Communication Skills		.78						.31	.66					.37			.36	.31	.56
Adaptability		.65						.34	.69					.40	.42				.50
Motivation		.68							.35				.60		.51			.35	.34
Forcefulness		.66						.34	.31				.55		.75				
Decision Making		.35					.54		.59					.37	.52				
Administrative Skills						.50					.72							.62	
Organizational Ident.					.80					.84							.85		
Organizational Identification Effectiveness							.61		.31			.45						.60	
Supervisory Skills			-.45	.30	.41					.70							.52		
Techn. & Tact. Comp.					.83					.86							.82		
			(Prop. Var. = .63)						(Prop. Var. = .63)						(Prop. Var. = .64)				

Note: Factor loadings below .30 in absolute magnitude have been excluded to promote readability.

Table A-5

FACTOR ANALYSES OF APPRAISAL INTERVIEW
RATING SCALES

Rating Scales	IOAC & 197th		IOBC, BIOCC, ROTC			ANCOES		
	I	II	I	II	III	I	II	III
1. Planning		.77		.74			.62	.46
2. Topic Selection		.72		.73			.70	.37
3. Written Communication		.62			.73			.77
4. Written Organization		.50			.72			.71
5. Self Confidence	.80		.80			.77	.30	
6. Oral Comm.	.82		.81			.73		
7. Use of Information	.65		.71			.66		
8. Accommodation	.72		.63			.72		
	N = 86 Prop. of var. acct. for = .54		N = 141 Prop. of var. acct. for = .61			N = 87 Prop. of var. acct. for = .62		

Note: Factor loadings below .30 in absolute magnitude have been excluded to promote readability.

Table A-6

FACTOR ANALYSIS OF LEADERLESS GROUP DISCUSSION RATING SCALES

Rating Scales	IOAC (N = 71)				IOBC, BIOCC, ANCOES, ROTC (N=192)			
	I	II	III	IV	I	II	III	IV
Assessor Ratings								
Formal Oral Communication	.83				.84			
Oral Organization	.83				.84			
Presentation Impact	.86				.88			
Participation		.79				.80		
Leadership and Group Facilitation		.89				.82		
Persuasiveness		.72				.66		
Social Concern		.31	.72			.35	.58	
Avoiding Negative Impression			.73				.65	
Conveying Information	.38	.70			.62	.47	.40	
Participant Ratings								
Oral Communication	.50			.67				.79
Leadership		.66		.54		.43		.71
Persuasiveness				.79				.79
Idea Quality				.72				.76
Sociability				.63				.67
Overall Effectiveness	.30	.45		.66				.83
	(Prop. var. = .70)				(Prop. var. = .71)			

NOTE: Factor loadings below .30 in absolute magnitude have been excluded to promote readability.

Table A-7

FACTOR ANALYSIS OF ASSIGNED LEADER GROUP EXERCISE RATING SCALES FOR IOBC, BIOCC, ANCOES, AND ROTC GROUPS

Rating Scales	Factor Loadings		
	I	II	III
Assessor Ratings:			
Planning	.76		
Leadership	.77	.45	
Decisiveness	.79	.39	
Flexibility	.76		
Motivation	.50	.75	
Physical Ability		.64	
Stress Tolerance	.60	.53	
Emergent Leadership	.46	.72	
Group Facilitation	.39	.82	
Participant Ratings:			
Social Association			.73
Leadership			.76
Support of Leader			.82
General Esprit			.77

Note: Factor loadings below .30 in absolute magnitude have been excluded to promote readability. N = 132. The proportion of variance in the matrix accounted for by the factors extracted was .71.

Table A-8

FACTOR ANALYSIS OF LEADER WAR GAME RATING SCALES
FOR IOAC and 197TH INF. BDE. GROUPS

Rating Scales	Factor Loadings		
	I	II	III
Assessor Ratings:			
Organization	.71		
Leadership	.76	.38	.33
Planning	.76	.40	
Flexibility	.53		
Supervisory Skills	.88		
Participation	.56	.55	.49
Problem Comprehension	.45	.70	.34
Emergent Leadership	.30	.76	.47
Overall Effectiveness		.76	.47
Participant Ratings:			
Problem Comprehension		.44	.77
Leadership			.76
Support of Leader		.37	.83
General Esprit			.64
Overall Effectiveness		.34	.88

Note: Factor loadings below .30 in absolute magnitude have been excluded to promote readability. N = 71. The proportion of variance in the matrix accounted for by the factors extracted was .75.

Table A-9

FACTOR ANALYSIS OF CONGLOMERATE EXERCISE RATING SCALES FOR ALL ASSESSMENT GROUPS COMBINED

Rating Scales	Factor Loadings	
	I	II
Assessor Ratings:		
Receptivity to Ideas	.64	
Group Facilitation	.86	.41
Leader Emergence	.80	.49
Sensitivity	.67	
Oral Communication	.75	.46
Energy and Vigor	.79	.43
Decision Quality	.75	.37
Overall Effectiveness	.72	.55
Participant Ratings:		
Popularity		.85
Quality of Ideas	.31	.80
Energy		.75
Conflict Creation	-.45	
Working with Peers		.83

Note: Factor loadings below .30 in absolute magnitude have been excluded to promote readability. N = 209. The proportion of variance in the matrix accounted for by the factors extracted was .69.

Table A-10

FACTOR ANALYSIS OF STANDARDIZED PAPER AND PENCIL TEST FOR ALL ASSESSMENT GROUPS COMBINED

Tests	Factor Loadings								
	I	II	III	IV	V	VI	VII	VIII	IX
Henmon Nelson									
Quantitative	.72								
Verbal	.90								
Nelson-Denney									
Vocabulary	.90								
Comprehension	.89								
Reading Rate	.66								
Watson-Glaser									
Critical Thinking	.72								
Social Insight Test	.54	.31							
WEPS		.44		.39					
Leader Opinion Qr.									
Consideration		.59					.35		
Structure		.48		.34	.31		.30		
Leadership Q Sort									
Leader Potential					.52				
Personal Integrity	.39		.37	.33	.54				
Consideration		.39	.38				.33		
Mental Health		.56		.43					
Technical Information	.49				.44				
Decision Making	.43	.47		.31					
Teaching and Communication					-.32				
Edwards Personal Preference Schedule									
Achievement	.30	.31							
Order		.72							
Exhibitionism				.46					
Autonomy				.36			.41		
Affiliation			.65						
Intraception		.47							
Succorance									.58
Dominance		.34					.54		
Abasement		.30						.52	
Nurturance			.78						
Change				.60					
Endurance		.71							
Heterosexuality				.53					
Aggression							.64		
Consistency						.52			

Note: Factor loadings below .30 in absolute magnitude have been excluded to promote readability. N = 239. The proportion of variance in the matrix accounted for by the factors extracted was .58.