RESEARCH ON CRITERIA OF OFFICER EFFECTIVENESS

Glenn E. McClure, et al

Air Force Personnel and Training Research Center
Lackland Air Force Base, Texas

May 1951

AIR TRAINING COMMAND

HUMAN RESOURCES RESEARCH CENTER

Research Bulletin 51-8

RESEARCH ON CRITERIA OF OFFICER EFFECTIVENESS

By

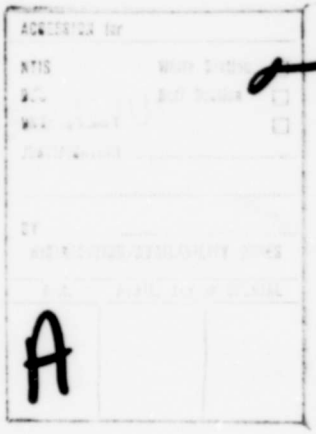
GLENN E. McCLURE
ERNEST C. TUPES
and
JOHN T. DAILEY

Lackland Air Force Base
San Antonio, Texas
May 1951

HEADQUARTERS
HUMAN RESOURCES RESEARCH CENTER
AIR TRAINING COMMAND
LACKLAND AIR FORCE BASE
San Antonio, Texas

COMMANDING OFFICER, COL. CLARK L. HOSMER

DIRECTOR OF OPERATIONS, DR. ARTHUR W. MELTON



RESEARCH ON CRITERIA OF OFFICER EFFECTIVENESS

Project No. 21-03-027

By

GLENN E. McCLURE

ERNEST C. TUPES

and

JOHN T. DAILEY

PERSONNEL RESEARCH LABORATORY

HUMAN RESOURCES RESEARCH CENTER

AIR TRAINING COMMAND

LACKLAND AIR FORCE BASE

San Antonio, Texas

RESEARCH BULLETIN 51-8

May 1951

SUBMITTED BY:

JOHN T. DAILEY
Associate Director of Research
Personnel Research Laboratory

ACKNOWLEDGMENTS

Major General Charles W. Lawrence, Major General Robert W. Douglass, Jr., and Brigadier General Charles F. Born, commandants at various times of USAF Officer Candidate School and the Indoctrination Division, made personnel and records available to research personnel for the conduct of this study.

Colonels Paul L. Barton, Samuel C. Gurney, Jr., Robert R. Little and Edward J. York, assistant commandants of USAF Officer Candidate School at various times during this study, cooperated and assisted in every possible way, as did Lt. Colonel Robert L. Hopsak, Majors Hugh D. Young and Raymond E. Hamlyn, Captain Edward R. Wiley and Lieutenant August V. Duda of the OCS staff.

Particular acknowledgment is made of the excellent cooperation and considerable work on the part of the Indoctrination Division Group and Squadron commanders and the members of their staffs who arranged for and made the evaluations which were essential to the criterion development portion of this study.

Dr. Robert L. Thorndike critically reviewed the entire manuscript.

Dr. Abraham S. Levine assisted in the final editing of the manuscript.

Miss Mabel P. Naas carried on the principal share of project coordination and field work associated with early phases of the study. In addition to the authors and contributors, others participating were Mr. John A. Cox, Mr. Nathaniel L. Dubberly, Mr. Harold D. Baker, Mr. Robert S. Coe, Miss Margaret Kirksey and Mr. Vernon O. Nelson.

TABLE OF CONTENTS

LIST OF TABLES

ABSTRACT

The research on measurement of officer qualities reported in this bulletin represents logical expansion of work begun earlier and reported in Air Training Command Human Resources Research Center Research Bulletin 47-1 (2). When the USAF decided to retain graduates of Officer Candidate School in the Indoctrination Division for a six-month on-the-job training tour immediately after commissioning, it became operationally possible to administer motivation and personality measures experimentally to samples that could not only be followed through Officer Candidate School but could also be evaluated for actual on-the-job officer proficiency.

This report is a resume of criterion research within the Officer Candidate School structure, and the development of an on-the-job measure of officer performance for use as a criterion of officer quality.

This study indicates that: (1) officer quality can be rated with sufficient reliability to warrant use of such ratings as a criterion for evaluation of predictive devices; (2) the reliability of on-the-job ratings of officers varies directly with the amount of prior briefing of raters as to what qualities to observe and with the amount of guidance each rater receives at the time of the rating; (3) the three most important proficiency traits of newly commissioned lieutenants seem to be executive ability, conscientiousness, and cooperativeness; and (4) ratings by peers (Buddy Ratings) during Officer Candidate School training have a higher positive relationship with later on-the-job performance ratings than do academic and military grades.

RESEARCH ON CRITERIA OF OFFICER EFFECTIVENESS

INTRODUCTION

This research bulletin represents a report on research done since 1947 by the Personnel Research Laboratory[1] on the development of criteria of officer effectiveness. It is in a sense a general progress report, to be followed by other more specific bulletins as soon as additional significant phases of research in this area are completed. No effort has been made to discuss wartime research in this area nor are any research studies reported in the area of aircrew selection. Work in this area has been reported elsewhere (4).

Recent studies in officer criterion development have been in two broad areas: (1) the development of criteria within the Officer Candidate School itself; and (2) the development of criteria of on-the-job proficiency of newly commissioned second lieutenants. The present report, in general, follows this delineation.

General background of the problem. The need for improving methods of selecting individuals for training and commissioning as officers in the military service has been a long, continuing one. Although great strides were made in psychological measurement of human aptitudes during World War II, these were necessarily in the areas in which the greatest immediate need existed, i.e., the selection of aircrew and technicians. Except as it fitted into or was a component part of such research, the measurement of the more intangible motivational and personality aspects of officer quality was not so thoroughly investigated.

It is the current thesis of most USAF research psychologists that any selection devices in the area of officer quality should be applicable to all individuals desiring commissioned officer status and that special aptitudes, such as those required for operation of aircraft, should be considered only after the general prerequisites of officer quality have been met. That is, only after it has been ascertained that an individual measures up to the general standards required of all officers should he be considered for aviation cadet or other specialized officer training.

---

[1]Formerly the Directorate of Personnel Research.

Until fairly recently, post World War II research in the development of predictive measures of officer quality in the United States Air Force was limited to the use of records of the Officer Candidate School for criterion data (2). It was well recognized that such in-school criteria, with the exception of practical applications courses, (2, p. 9) are not likely to be satisfactory predictors of later performance,[2] but their use was mandatory since better criteria were unavailable. Relatively little emphasis was placed on construction of predictive tests, since it was believed to be uneconomical to do so until an adequate criterion had been established. However, two devices, a biographical inventory and the Minnesota Personality Scale for Men, were investigated as possible predictors of graduation or elimination from the Officer Candidate School. The records of women candidates, not used thus far in this research, have been filed for study when enough cases have accumulated.

Rationale and scope of the present study. Early in 1947 a letter[3] from Air Force Training Command was addressed to the Psychological Research and Examining Unit,[4] directing the development of measures to evaluate and to predict officer quality, including such traits as leadership, emotional stability, decisiveness and dependability.

---

[2]Such criteria were especially unstable for the Officer Candidate School due to several factors inherent in the military situation, such as changing curricula and the numerous transfers of instructors, tactical officers, policy making commanders and executives. In addition, the use of OCS grades as criteria is based on the assumption that the candidates with the highest grades are those individuals who would later be judged the best officers by their associates. Inasmuch as grades are in large measure composed of intellectual factors, there is a strong possibility that rated officer proficiency (over and above a certain minimum of intelligence) is largely a function of personality and motivation factors only slightly correlated with intelligence. Thus another weakness of OCS grades as a criterion measure becomes apparent. This should not be construed as an argument against the use of graduation-elimination data as one part of a multiple criterion if some measure of later on-the-job proficiency is also included. Certainly Officer Candidate School is one hurdle which must be passed, and the ideal selection battery would consist of instruments which would first screen out those applicants who would be unlikely to succeed in Officer Candidate School, and then from the remainder select those who would be evaluated satisfactory as officers after graduation.

[3]Ltr, Hq AFTRC to CG, IDTRC, File 332 Misc., dtd 4 Dec 46, Subj: Psychological Research.

[4]Most of the functions of the Psychological Research and Examining Unit have now been taken over by the Personnel Research Laboratory, Human Resources Research Center.

The broad design of the research study evolving from this directive included two principal projects:

1. The administration, to students in the Officer Candidate School, of psychological measuring devices believed to be predictors of officer quality; these devices to consist of specifically constructed motivation and personality measures developed both empirically and on an a priori basis, as well as commercial tests in the area of motivation and personality measurement.

2. The development of a relevant and reliable measure of officer quality with which the individuals who entered and completed OCS training could later be evaluated on the job.

Selection research in this study has been almost entirely on the development of personality and interest tests (rather than new types of intellectual tests) for several reasons. Since psychological knowledge has advanced to the point where it is relatively easy to develop tests which will validly account for that part of the variance of any reliable criterion attributable to intellectual or ability factors, it was deemed advisable to postpone work on intellectual predictors and give priority to the development of non-intellectual measures. It was assumed on the basis of past experience that the academic grades in OCS could be predicted with an acceptable degree of accuracy by existing types of intellectual tests.[5]

In the meantime, it was considered desirable to proceed with the development of personality measures, which, at least on an a priori basis, appear to measure traits characteristic of effective or ineffective commissioned officers. Furthermore, it did not seem likely that any portion of the criterion variance accounted for by these personality tests would also be accounted for by the intellectual tests to be developed later, i.e., there seemed to be little danger of overlap. In a broad sense, the prerequisite of two years college or its equivalent sets a minimum intellectual level for entrance into Officer Candidate School.

While the criterion development and test development research studies were carried on simultaneously, the present report will present only the results of the criterion studies. The results of the test development and validation studies are to be presented in a forthcoming series of bulletins.

---

[5] A validation study is nearing completion on a test to measure intellectual functions (The Aviation Cadet-Officer Candidate Qualifying Test).

ANALYSIS OF EXISTING CRITERIA IN OCS

In order that predictive devices of good officer performance could be validated, a standard of effective, desired officer quality was a necessity. Since no usable official Air Force standards of officer quality existed at the time this study was initiated, steps were taken to develop such standards for use as criteria. The development of these standards is described later in this report. First, however, criteria obtainable within the OCS structure will be described and evaluated in detail. These criteria have included academic and military grades, ratings by fellow students, and graduation vs. elimination. Similar criteria have been used for many years in academic and industrial situations.

The OCS grading system. This grading system has been discussed at length in Research Bulletin 47-1 (2, p. 24) and was based on the conversion of raw scores into a 5-point scale. Candidates received three average grades for each of the three 2-month phases, one each for academic, military and the over-all average. If a student dropped below 2.5 on either the military or the academic grade or below 2.7 on the over-all average grade, he met a faculty board who determined whether or not he would be retained in the school.

The academic grade was the weighted average of all graded tests on such subjects as leadership, administration, supply, English, law, and small arms. The military grade was the weighted average of the subjective scores given by instructors in such classes as close order drill, physical training, and inspection plus a consideration of the candidate's standing on student rating forms, commonly referred to as Buddy Ratings. The over-all grade was an average of the military and academic grades.

In 1947, OCS officials and research psychologists jointly attempted to improve the validity of academic grading and evaluation systems within the OCS structure but with only partially satisfactory results as far as their use as criteria was concerned. One of the steps taken was the establishment of a "practical applications" course in which problem situations similar to those commonly confronting commissioned officers were given to small groups of candidates for study. During the 3-hour class period in which solutions to the problem were presented, at least two instructors rated each member of the reporting group. Unfortunately, this device soon became merely an investigating and reporting procedure with more emphasis on public-speaking ability during presentation than on evaluation of performance in a simulated officer situation.

The following correlations between ratings by supervisory officers[6] and grades for 336 graduates of OCS classes 48-A and 48-B illustrate the low predictive value of school marks for an on-the-job criterion.

---

[6]See p. 8 for the details of the development of the rating form used.

Variable	Correlation with on-the-job ratings
Leadership course grade	.14
Academic average	.14
Military average	.19
Over-all average	.16
Student rating (N = 265)	.34

The correlations suggest that these grades have very low validity as predictors of future job performance, and thus their use as a sole basis for elimination may not be justified. The relatively higher correlation between the student ratings and the relevant on-the-job ratings indicates that the student ratings appear to offer considerable promise as a criterion and should be investigated intensively. It must be remembered in interpreting these correlations, however, that they are based on only the graduating portion of the classes, and consequently are attenuated by restricted range on the grade variable and possibly on the evaluation variable as well. The obvious test of this thesis--letting one entire group of OCS entrants graduate--is unfortunately not feasible at this time.

The revised curriculum for 1950 classes includes a much more thorough and closely monitored practical applications course. OCS officials conferred with research psychologists to insure that data gathered in the new course can be used in the continuation of the general study of OCS criteria. A new grading system was inaugurated with the 50-A class which was designed to produce sharper delineation at the extremes of the grading curve. Those students whose relative standing is in the lower 15 per cent at the end of the eighth week or in the lower 10 per cent at the end of the sixteenth week are given a comprehensive essay-type examination. (In this way, candidates who, for one reason or another, are not adept at objective-type examinations are given a chance to show what they have learned in essay-type examinations.) Candidates who make less than 70 per cent on this test are recommended to appear before the Faculty Board to show cause why they should not be eliminated.

Pass-fail criterion. Status in terms of the pass-fail criterion is not solely a function of success in Officer Candidate School, or lack of it. Another group entered into the "fail" category, namely, the voluntary resignees. Whereas the reasons for academic failure were few and could be isolated, there were myriad reasons for resignation; and, although they fall in the general area of motivation, many are too intangible for isolation at the present time. For Class 49-B, all attrition amounted to 23 per cent. Resignations amounted to a little more than 9 per cent of the entering group.

Personnel of the Directorate of Personnel Research interviewed all entrants who did not graduate with Class 49-B in an attempt to determine the underlying reasons for withdrawal as contrasted with the stated

reasons given to OCS staff officers. Information derived from this survey indicated that the biggest factor precipitating resignation, and to a certain extent academic elimination, was the candidate's lack of correct information about OCS prior to his application for entry. Many candidates arriving at the school during this period were met by a situation entirely different from the one expected, a factor which led to an early resignation, especially if motivation was low or fairly lucrative jobs were available on the outside. This same situation was true for a few candidates who were eliminated academically. They apparently had a mental set for a type of academic training foreign to the actual situation.

From the foregoing, it would seem that OCS pass-fail criterion should treat resignation and elimination as separate entities. Not enough data have matured to date to make this separation possible in current validation studies although it will be done later as the combined samples become larger.

Buddy Ratings. The most promising OCS criterion found in this study was the Student Rating Form, commonly referred to as the Buddy Rating. In this procedure, each candidate was required to rank all other candidates in his flight, with the number-one ranking going to the man he considered the most outstanding of the 40 or 50 men in the flight. Each candidate's score was an average of all ranks assigned him. This process, carried out five times, once at the end of each month except the sixth, was used in conjunction with grades as a partial basis for elimination.

These rankings appeared to be acceptably reliable. Correlation between the average rankings at the end of the third and fourth months was .91 and between those of the first and fourth months, .61.

TABLE 1

DISTRIBUTION STATISTICS AND INTERCORRELATIONS OF OCS CRITERIA

Sample: 336 graduates, Classes 48-A and 48-B.

Criteria	Mean	SD	Military Average	Over-all Average	Student Rating[a]
Academic Average	3.20	.39	.69	.87	.27
Military Average	3.24	.35		.88	.22
Over-all Average	3.26	.33			.29

[a] N for Student Ratings was 265.

Table 1 shows the intercorrelations of the student ratings and the three types of school grade averages. It can be seen that the military and academic grades are relatively highly intercorrelated and have little correlation with the student ratings. Of course both the military and academic grades correlate very highly with the over-all average because they are heavily weighted into it.

Beginning with 1950 classes, Buddy Ratings are being accomplished by use of selected items from AF Form 77A, the official Report of Officer Effectiveness. One of the reasons for this departure from the rank-order system is the desirability of measuring the performance of all candidates by a single standard rather than against the numbers of their own flight. A secondary benefit is the resulting thorough familiarization of all candidates with the official yardstick of Air Force officer performance.

Conclusions. Evidence derived from study of the 1948 and 1949 classes leads to the following conclusions:

(1) There is a higher relationship between ratings by peers in Officer Candidate School (Buddy Ratings) and later on-the-job proficiency ratings than between OCS academic or military grades and proficiency ratings. This suggests that the Buddy Ratings should be given considerable weight in the OCS final grade average, if the object of OCS is to produce officers who will later be judged satisfactory by supervisory officers.

(2) A pass-fail criterion should treat resignees separately from others in the fail category. (Future validation studies will do so as soon as enough cases have accumulated to permit such a treatment.)

(3) Low correlations between grades and on-the-job ratings indicate that some revision of the OCS grading system and possibly the curriculum seems necessary. (For 1950 classes, an entirely new OCS curriculum has been set up by the OCS staff, based on areas of performance contained in AF Form 77A, the official criterion of officer proficiency.)

ON-THE-JOB CRITERION DEVELOPMENT

One of the most difficult problems in connection with military selection and classification programs has been that of finding criteria other than training school grades for use in the validation of various experimental predictive devices. Even when the jobs were relatively clearly defined, such as aircrew jobs, a great deal of difficulty was experienced in the development of adequately reliable criteria of proficiency (3, pp. 42-47). In working with such an intangible factor as officer quality or leadership, the problems of criterion development become many times more difficult.

As mentioned earlier, a fundamental assumption made in this study is that the same basic officer qualities are necessary for both flying and non-flying officers and that aptitude for flying is a requisite to be determined after that of officer quality. When USAF procedures were changed so that graduates of OCS would remain at Lackland Air Force Base for a six-month on-the-job training tour of duty immediately after being commissioned, it became administratively feasible to secure an on-the-job proficiency measure of officer quality as a criterion for validation of experimental selective devices.

Development of Evaluation Scale for Graduates of USAF OCS, Form A

As a sub-project on the development of an on-the-job criterion, a study was conducted to develop a rating scale criterion. The population used was the 364 lieutenants graduating in the two 1948 OCS classes. A unique aspect of this study was the use of groups of raters, as well as ratees, who worked in a single Air Force Division where rating scales could be administered under standardized conditions by personnel trained in psychological research.

The Evaluation Scale, Form A developed for this study was based on data contained in a report, by the American Institute for Research, of research done under contract with the United States Air Force to determine those behavior traits believed by regular Air Force officers to be associated with effective and ineffective officer performance (6). In this study, 640 officers were interviewed. Distribution of ranks was: generals, 1 per cent; colonels, 19 per cent; lieutenant colonels, 28 per cent; majors, 36 per cent; captains, 15 per cent; and lieutenants, 10 per cent. Each officer was requested to describe several incidents of effective behavior and several incidents of ineffective behavior, which he had observed in officers with whom he had served in the recent past. These incidents were subsequently categorized around central traits or concepts of behavior. The final result was a 54-item check list[7] submitted to the Air Force as a recommended officer evaluation device to replace the previously used Army scale. The Evaluation Scale, Form A was essentially a brief version of the AF Form 77A especially designed for the Lackland on-the-job groups. Inasmuch as some of the items covered by the Form 77A recommended for use throughout the Air Force did not seem to be applicable to the situations the majority of newly-commissioned officers were likely to encounter during their six-month tour at Lackland, the Evaluation Scale, Form A was developed from it for use in the evaluation of graduates of OCS.

Selection of items. A board consisting of commissioned officers, non-commissioned officers, and civilian psychologists reviewed the original 54-item scale and agreed upon the deletion of 28 items on the basis of non-applicability to the apprentice officer situation. In order to eliminate the opportunity for a rater to mark a numerical "average", the

[7] Report of Officer Effectiveness, AF Form 77A, 15 Mar 49.

remaining items were rewritten so that four, instead of five, descriptive statements followed each trait. In addition, these choices were arranged randomly to minimize straight-line marking.

The same group of experts was asked to indicate whether they agreed or disagreed with an a priori key which assigned from 1 to 4 points to each alternative, depending upon the rank order of its desirability within the item. Based on suggested changes, a corrected key was constructed for later use in scoring of completed papers.

In order to resolve a few minor controversial points, the 24 items included in the final scale[8] were again reviewed and approved by a board of civilian and military psychologists.

The cover sheet for the three-page booklet contained blanks for the accumulation of identification data concerning both rater and ratees, and a statement as to the number of months the ratee had worked in his current assignment, the number of months he had worked under the immediate supervision of the rater, and the number of months he had been known to the rater.

Administration of the scale. OCS Class 48-A graduated at the end of June 1948 and served on-the-job training tours until the end of December 1948. Class 48-B graduated at the end of December 1948 and served on-the-job training tours until the end of June 1949. During the first month of duty for each class, a letter from the Commanding General was sent through channels to all commanding officers in the Indoctrination Division explaining the study and requesting certain information such as name, telephone and building numbers of the two supervisory or associate officers best situated and next-best situated to observe and evaluate OCS graduates in their organizations. These commanders were further requested to insure that the officers whose names were submitted make critical observations of the performance of the student officers in the following general areas:

a. Work habits.

b. Supervision of subordinate personnel.

c. Acceptance of team principle and organization discipline.

d. Personal habits and adjustments.

e. Interest in assignment.

f. Other characteristics believed to be indicators of effectiveness of USAF officers.

---

[8]Reproduced in Appendix A.

A closing paragraph of the letter gave notice of the approximate date and the method by which performance evaluation would be accomplished.

A brief single-page exposition entitled "Hints to Raters"[9] was written in laymen's language and distributed to raters just prior to actual marking of rating forms. This paper explained "halo" effect, "central tendency," "error of leniency," and "independence of dual ratings." In order to secure as high a degree of standardization as possible in the accomplishment of ratings, thoroughly briefed civilian psychologists called on supervisory officers at their place of work to confer with them and obtain the evaluations. After establishment of rapport, raters were given the long range over-all rationale for the study and a brief outline of the techniques to be used. Questions arising during the rating session were answered on the spot, insofar as possible in a predetermined manner, by the conferring psychologist.

Since one of the objectives in the development of this scale was to make it as short and as easy to complete as was possible without loss of reliability, raters were asked to place a check in the space provided just adjacent to the alternative of their choice. The ease of this administration, as compared with the use of separate answer sheets, was a big factor in the ready acceptance of the form by busy officers who are often confronted with complicated and lengthy official papers to complete.

As was expected in the military organization, transfers of supervisory personnel frequently resulted in the subsitution of other raters deemed by their commanders to be "next best situated" to have observed ratees. Ratings from two supervisors were obtained on 169 graduates of Class 48-A and 158 graduates of Class 48-B. No ratings were made by any supervisory officer who had not observed the student officer for at least thirty days.

Reliability of Form A. Item reliabilities for the 169 paired ratings of Class 48-A are reported in Table 2.

On the basis of these statistics, items whose reliabilities were not significant at the 1 per cent level (1, 7, and 21) were omitted from the rating booklet revised for use with 1949 OCS graduates. Table 3 gives distribution statistics and total-scale reliabilities for both classes.

A search of the literature, for purposes of comparison, failed to produce directly comparable reliability studies and only a few reliability studies on officer-quality evaluation using an on-the-job performance rating scale.

In an evaluation of the Report on the Fitness of Naval Officers made by the Research Activity Office of the US Navy in July of 1945, it was

[9]Reproduced in Appendix B.

TABLE 2

RELIABILITY OF EACH ITEM ON EVALUATION SCALE, FORM A, FOR GRADUATES OF USAF OCS, AS ESTIMATED BY TETRACHORIC CORRELATION BETWEEN RATER I AND RATER II

Sample: 169 graduates, OCS Class 48-A

	Item	r_{tet}
1.	Understanding Instructions	.11*
2.	Planning and Scheduling Work	.29
3.	Presenting Finished Work	.21
4.	Securing Cooperation	.27
5.	Keeping Others Informed	.22
6.	Delegating Authority	.28
7.	Giving Orders and Instructions	.15*
8.	Setting A Good Example for Work Performance	.38
9.	Looking Out for Subordinates' Welfare	.21
10.	Problem Solving	.31
11.	Judgment	.38
12.	Assumption of Responsibility	.38
13.	Complying with Orders and Directives	.30
14.	Reaction to Organizational Procedure	.28
15.	Subordinating Personal Interest	.36
16.	Cooperating with Associates	.25
17.	Loyalty to Total Organization	.27
18.	Attending to Details	.33
19.	Keeping Appointments	.35
20.	Improving Effectiveness	.30
21.	Being Fair and Scrupulous	.18*
22.	Maintaining Military Appearance	.34
23.	Self Confidence	.31
24.	Temperament	.25

*Not significant at 1 per cent level. All other reliabilities are significant at or beyond the 1 per cent level.

TABLE 3

DISTRIBUTION STATISTICS AND RELIABILITIES OF EVALUATION SCALE, FORM A, RATINGS FOR GRADUATES OF OCS CLASSES OF 1948

	Class 48-A	Class 48-B
Number of cases	169	158
Mean	70.6	73.3
Standard Deviation	12.8	12.9
Rater I vs. Rater II (Pearson r)	.55	.51
Rater I vs. Rater II (Intra-class r)	.54	.51
Reliability of two ratings combined*	.71	.68

reported that correlations between ratings of successive commanders on 300 commissioned officers ranged from .22 to .38 for personal traits, and from .28 to .48 for job-proficiency items (6, Technical appendices and notes).

An officer rating scale, including such comparatively easily measured traits as proficiency in formation flying and proficiency in instrument flying, was used to evaluate 160 airplane commanders in the Second Air Force during World War II. A comparsion of ratings made independently by two supervisors showed a reliability coefficient of .55 (3, p. 104).

A study of another World War II officer effectiveness rating scale, developed for the evaluation of instructors at an advanced navigation school, showed it to have a reliability coefficient of .24 when used independently by supervisors and flight commanders (1, p. 159).

When a sample of 587 second ratings on the Army's Officer Evaluation Report was compared with those made by immediate superiors, the correlation was .57 (7). When stepped up by the Spearman-Brown prophecy formula, this became .73.

The reliability figures reported in Table 3 compare favorably with the highest found to date in the literature in this area of research on officer proficiency. This reliability estimate, though it may be inflated somewhat due to bias arising from local reputation, appears to be of such magnitude that a low correlation of a predictive measuring device with it would necessarily indicate an invalid predictive measure for this particular criterion; i.e., its failure to correlate significantly could not be attributed to the unreliability of the criterion used. It would appear that the Evaluation Scale, Form A would yield a criterion of fairly satisfactory reliability suitable for validation studies if the average of two or more such ratings is employed.

*Estimated by the Spearman-Brown prophecy formula.

Cluster Analysis of Evaluation Scale for Graduates of USAF OCS, Form A

Intercorrelations were computed between the items of the Form A ratings of graduates of OCS Class 48-A (N = 169). This was done by first combining the primary and secondary ratings for each item, yielding item scores whose distributions ranged from 2 to 8; then dichotomizing the distribution as near the midpoint as possible; and finally computing tetrachoric correlations between items. The median intercorrelation is .51.

The resultant correlation matrix was arranged in approximately hierarchical order (Table 4), according to size of median intercorrelations of each item, and examined for possible clusters. Constellations having intercorrelations within themselves of more than an arbitrary median of .60, and a low enough median intercorrelation with all other items to achieve a beta ratio[10] of 1.30 or better, were considered to be clusters.

The results of the analysis are given in Table 5. Three clusters and three doublets evolved. An examination of items making up Cluster I indicates that it is probably measuring executive and administrative abilities. These items have high intercorrelations with practically all other items, suggesting that these abilities have a halo effect on most other traits measured by this scale.

Cluster II, though not as clean-cut as Cluster I, appears psychologically significant. Items making up this cluster seem to indicate traits on the periphery of executive and administrative abilities. Persons characterized by the Cluster II configuration but not by the Cluster I configuration would seem to represent individuals who are trying eagerly to be proficient in the area measured by Cluster I, but who may not be successful in doing so.

An examination of Cluster III indicates a possible cooperative or "good Joe" trait. Persons possessing this trait may tend to go out of their way in attempting to cooperate with and to please everyone.

It should be noted that Cluster I contaminates the other clusters to some extent. It is debatable whether a cluster analysis on this type of matrix could ever yield clear-cut results, simply because of the nature of the data. That is, each item in the scale was used because it was believed to be highly correlated with officer quality, and thus a

(Text continues on page 16)

[10]The beta ratio is used to determine whether a group of variables correlate among themselves relatively more highly than they correlate with the other variables in the matrix. It is obtained by dividing the median intercorrelation within the constellation by the median correlation between the items in the constellation and all other items. The beta ratio is similar to Holzinger's B-coefficient, but is based on median rather than average intercorrelations and thus is much simpler to compute. The value of 1.30 as the lower limit of an acceptable beta ratio is purely arbitrary.

TABLE 4

INTERCORRELATIONS OF ITEMS FOR EVALUATION SCALE, FORM A, FOR GRADUATES OF USAF OCS

Sample: 169 graduates, OCS Class 48-A. (Decimal points are omitted.)

Item No.	Item	11	8	3	13	14	6	18	12	19	17	9	16	15	4	24	20	23	2	5	22	10	21	1	7
11.	Judgment		80	80	71	76	69	63	60	60	51	28	62	57	62	63	40	65	53	62	49	60	54	28	69
8.	Setting a Good Example	80		68	76	68	72	74	62	60	69	63	60	62	50	58	57	54	61	64	60	70	68	32	58
3.	Presenting Finished Work	80	68		74	73	69	70	55	62	54	42	32	64	58	54	52	66	62	52	28	60	46	11	60
13.	Complying with Orders and Directives	71	76	74		74	72	61	69	60	62	52	70	64	68	58	75	42	60	57	38	60	56	22	56
14.	Reaction to Organizational Procedure	76	68	73	74		74	66	70	70	60	35	62	55	60	54	55	42	57	55	40	52	49	34	66
6.	Delegating Authority	69	72	69	72	74		72	68	54	57	48	66	62	64	64	54	36	60	60	35	64	43	22	56
18.	Attending to Details	63	74	70	61	66	72		54	60	70	56	58	48	49	57	42	57	50	65	38	56	58	12	48
12.	Assumption of Responsibility	60	62	55	69	70	68	54		46	67	49	61	57	48	43	52	46	54	52	37	40	50	21	49
19.	Keeping Appointments	60	60	62	60	70	54	60	46		53	36	35	31	52	54	42	46	47	49	26	34	40	12	36
17.	Loyalty to Total Organization	51	69	54	62	60	57	70	67	53		70	51	42	47	54	60	55	44	54	39	48	61	19	37
9.	*Looking Out for Subordinates' Welfare	28	60	42	52	35	48	56	49	36	70		50	40	49	37	53	37	45	48	20	31	47	12	16
16.	Cooperating with Associates	62	60	32	70	62	66	58	61	35	51	50		74	52	51	42	51	49	60	31	38	40	25	43
15.	Subordinating Personal Interest	57	62	64	64	55	62	48	57	31	42	40	74		41	42	49	35	41	42	15	34	46	26	58
4.	Securing Cooperation	62	50	58	68	60	64	49	48	52	47	49	52	41		70	35	28	54	60	20	47	47	06	43
24.	Temperament	63	58	54	58	54	64	57	43	54	54	37	51	42	70		46	51	37	39	33	31	41	04	37
20.	Improving Effectiveness	40	57	52	55	55	64	42	53	42	60	53	42	49	35	46		46	47	54	40	53	40	19	30
23.	Self Confidence	65	54	66	42	42	66	57	46	46	55	37	51	35	28	51	46		35	42	42	48	35	06	27
2.	Planning and Scheduling Work	53	61	62	60	57	60	50	54	47	44	45	49	41	54	37	47	35		64	36	37	48	06	45
5.	Keeping Others Informed	62	64	52	57	55	60	65	52	49	54	48	60	42	60	39	54	42	64		35	42	46	20	42
22.	Maintaining Military Appearance	49	60	28	38	40	35	38	37	26	39	20	31	15	20	33	40	42	36	35		36	38	02	40
10.	*Problem Solving	60	70	60	60	52	64	56	40	34	48	31	38	34	47	31	53	48	37	42	36		42	35	41
21.	Being Fair and Scrupulous	54	68	46	56	49	43	58	50	40	61	47	40	46	47	41	40	35	48	46	38	42		14	25
1.	*Understanding Instructions	28	32	11	22	34	22	12	21	12	19	12	25	26	06	04	19	06	06	20	02	35	14		25
7.	Giving Orders and Instructions	69	58	60	56	66	56	48	49	36	37	16	43	58	43	37	30	27	45	42	40	41	25	25	
	Item Reliabilities																								
	For one rater	38	38	21	30	28	28	33	38	35	27	21	25	36	27	25	31	31	29	22	34	31	18	11	15
	For sum of two raters (Spearman-Brown prophecy formula)	55	55	35	46	44	44	50	55	52	43	35	40	53	43	40	47	47	45	36	51	47	31	20	26

*Unreliable. Not included in cluster analysis shown in Table 5.

TABLE 5

CLUSTERS OF ITEMS FOR EVALUATION SCALE, FORM A,
FOR GRADUATES OF USAF OCS

Sample: 169 graduates, OCS Class 48-A.

	Item No.	Item	Median r within Cluster	Median r with Variables Outside Cluster	Beta Ratio
	3.	Presenting Finished Work			
	5.	Keeping Others Informed			
	6.	Delegating Authority			
	8.	Setting a Good Example			
Cluster I	12.	Assumption of Responsibility	.68	.52	1.30
	13.	Complying with Orders and Directives			
	14.	Reaction To Organizational Procedures			
	15.	Subordinating Personal Interests			
	16.	Cooperating with Associates			
	17.	Loyalty to Total Organization			
	18.	Attending to Details			
Cluster II	5.	Keeping Others Informed	.70	.55	1.30
	14.	Reaction to Organizational Procedures			
	16.	Cooperating with Associates			
Cluster III	15.	Subordinating Personal Interests	.70	.56	1.30
	13.	Complying with Orders and Directives			
	4.	Securing Cooperation			
Doublet I	24.	Temperament	.70	.51	1.37
	20.	Improving Effectiveness			
Doublet II	13.	Complying with Orders and Directives	.75	.55	1.32
	23.	Self-confidence			
Doublet III	3.	Presenting Finished Work	.66	.52	1.25
Unique	22.	Maintaining Military Appearance			
and	2.	Planning and Scheduling Work } Possible Doublet			
Unclustered	5.	Keeping Others Informed } Possible Doublet	.64	.52	1.23
	10.	Self-confidence			

large general cluster must be expected, which would obscure any additional group clusters present. However, this cluster analysis suggests that the scale items are measuring more than one factor related to officer performance, and furnishes some insight into the possible factorial structure of the AF Form 77A.

More meaningful results would undoubtedly be obtained through centroid factor analysis, which was contra-indicated in the present instance by the relatively small number of cases (N = 169) and the low item reliabilities (median r_{ii} = .44). It is planned to use the centroid factor analysis procedure when ratings have been obtained on enough cases (at least 500) to yield relatively stable intercorrelations.

AF Form 77A, Report of Officer Effectiveness[11]

The result of the recommendations made by the American Institute for Research in a study discussed earlier (6), was adopted in 1949 as the official measure of officer proficiency throughout the Air Force. With the advent of this official criterion, a plan was devised to compare ratings made on it with those made on the HRRC Evaluation Scale for Graduates of USAF OCS, Form B.[12]

Reliability estimates of Form 77A. For Class 49-A, ratings were obtained on both the official Air Force Form 77A, and Form B of the Evaluation Scale. Approximately 182 graduates were rated on Form 77A by primary and secondary raters. One hundred sixty-six of these graduates had also been rated on Form 77A about two months earlier by Lackland Air Force Base authorities for other purposes. These earlier ratings were made available to HRRC. Reliability estimates are given in Table 6.

Examination of Table 6 indicates that on Form 77A the reliability of one rating varies from .38 to .52 for each of the areas and is equal to .55 for the total score, when the agreement between two independent ratings is used as the reliability estimate. If the ratings of two raters are combined, the reliability of the summed ratings varies from .55 to .68 for the areas and is equal to .71 for the total. As indicated earlier in this paper, reliabilities of this size are satisfactory for use in validation of predictive devices if the sums of the two ratings are employed as is ordinarily possible in experimental studies, although not always possible for operational use.

Rate-rerate reliability. A reliability estimate of a different type became available when HRRC-administered Form 77A ratings were

[11]Hereafter referred to as Form 77A. See Appendix C for list of areas covered by the form.

[12]Hereafter referred to as Form B.

TABLE 6

DISTRIBUTION STATISTICS AND RELIABILITIES OF AF FORM 77A

Sample: 166-122 graduates, OCS Class 49-A.

	A	R	E	A[a]			
	I	II	III	IV	V	VI	Total Score
Means[b]	3.5	3.5	3.5	3.7	3.5	3.4	3.5
Standard deviations	.56	.50	.52	.55	.52	.51	.47
Agreement between two ratings (intra-class r)	.45	.52	.51	.38	.48	.42	.55
Reliability of sum of two ratings (Spearman-Brown prophecy formula)	.62	.68	.68	.55	.65	.59	.71
Rate-rerate reliability[c]							.77

[a]See Appendix D for list of areas covered by Form 77A.

[b]Each item in Form 77A is rated on a 1 through 5 basis or it may be marked "unknown." Area means are obtained by summing the item ratings in each area and dividing by the number of items rated. Total scores are obtained by summing the area scores and dividing by the number of areas rated.

[c]About two months between ratings.

correlated with ratings on the same form by Indoctrination Wing. For 166 cases either the primary or the secondary experimental rater was the officer who had made the Wing ratings about two months earlier. The "rate-rerate" reliability of the scale as estimated in this manner is .77.

Evaluation Scale, Form B. Form B of the Evaluation Scale differs from Form A, which was discussed in an earlier section,[13] in several respects. Three of the Form A items with exceptionally low reliability and negligible correlations with the total score were dropped. Definitions were written for all items in an effort to increase reliability. Two items were added but these were not counted in the total score. One of them asks the rater to mark "degree of personal liking" for the ratee. It was placed at the front of the scale in the hope that halo might be reduced.[14] The other item asks the rater to mark "over-all efficiency" of the ratee. It was placed at the end of the scale so as not to influence the marking of the other items. Another difference between Form A and Form B was the method of administration. Form A had been administered individually with a member of the HRRC staff conferring with each rating officer as the ratings were made. Form B was administered by assembling groups of rating officers and instructing them as a group, giving individual aid only if requested. Through an administrative error, Hints to Raters (Appendix B), handed to raters at the beginning of rating interview on Form A, was not used in group administration of Form B.

Reliability of Form B. The reliability estimates shown in Table 7 are of interest for several reasons. First, the reliability of Form B is significantly lower (at the 5 per cent level) than the reliability of Form A administered to earlier classes. The only differences between Forms A and B which appear related to this loss in reliability would seem to be the individual briefing as against group briefing of raters, and the fact that the "Hints to Raters" was not made available to Form B raters. Second, the reliability of Form B is significantly lower (beyond the 1 per cent level)[15] than the reliability of Form 77A when administered under identical conditions by the same group of raters on the same group of ratees. In part, this difference is probably due to

---

[13] See p. 8 above.

[14] Immediately following this item was a short paragraph describing the halo effect and its influence on ratings. The raters were advised that the "liking" ratings had been included to give them the opportunity to quantify their personal feeling toward the ratee, and they were asked to try to ignore those feelings in completing the remainder of the scale.

[15] Cochran's formula for estimating the significance of a difference between correlated coefficients of correlation was used (5, p. 218).

the difference in length of the two scales (Form 77A has 54 items while Form B has 20 which contribute to the total score). In fact, if the Form B were as long as the Form 77A, its estimated reliability (.79 if the Spearman-Brown prophecy formula is applied) would be higher than that of Form 77A.

TABLE 7

RELIABILITIES OF EVALUATION SCALE, FORM B, RATINGS OF 186 GRADUATES OF OCS CLASS 49-A

	Agreement between two raters[a]	Reliability of sum of two ratings[b]
Total score	.43	.60
Personal liking (Item 1)	.20	.33
Over-all efficiency (Item 22)	.42	.59

A third finding expressed in Table 7 is the estimated reliability (.59) of the average of two independent ratings on "Over-all Efficiency." This single item has a reliability as high as the Form B total score based on 20 other items. It is doubtful, however, if this single item could be used independently of the scale since much of its reliability is probably due to the stabilizing effect of the rater having first checked the 20 descriptive items.

Comparison of Form 77A and Form B. The correlation between total scores on Form 77A and total scores on Form B is equal to .82 (see Table 8). This indicates a very high relationship between the two criterion measures, especially when their level of reliability is considered. On the basis of this correlation and the higher reliability of Form 77A, no justification can be found for continuing to use the HRRC evaluation form. It is planned, therefore, to continue to obtain dual ratings of OCS graduates under standardized conditions, but to use only Form 77A for this purpose.

In Table 8 are shown the intercorrelations between the ratings on Personal Liking and Over-all Efficiency and total scores on Forms 77A and Form B.

---

[a]Estimated by computing the intra-class correlation between primary and secondary raters.

[b]Estimated by the Spearman-Brown prophecy formula.

TABLE 8

INTERCORRELATIONS OF ON-THE-JOB
OFFICER PERFORMANCE CRITERIA*

Sample: 187 graduates, OCS Class 49-A.

	Form B Over-all efficiency item	Form B total score	Form 77A total score
Form B Personal liking item	.46	.55	.45
Form B Over-all efficiency item		.76	.66
Form B total score			.82

It is apparent that not only is there quite a high relationship between the two rating scales, but further, that the single item rating on Over-all Efficiency correlates highly with both. The slightly higher correlation (.76 vs. .66) between Over-all Efficiency and Form B than between Over-all Efficiency and Form 77A is probably a function of the fact that ratings on Over-all Efficiency always occurred just after the other items on Form B were rated.

The item on degree of Personal Liking was included in Form B in an attempt to minimize the halo effect and also in an attempt to discover the part played by Personal Liking when judgments of efficiency are made. The hypothesis was that personal liking (or disliking) would account for much of the variance in the efficiency ratings. This hypothesis is not disproved by the data but neither is it completely sustained. Ratings on Personal Liking correlate .46 with Over-all Efficiency, .55 with Form B total score, and .45 with Form 77A total score. Personal Liking certainly is related to efficiency ratings but not to such an extent as to render the efficiency ratings suspect as criteria. Some correlation between Personal Liking and any criterion of officer efficiency based on ratings must be expected, insofar as the personality of the person being rated enters into the performance of his job.

In order to obtain some estimate of the stability of the 77A ratings over a period of time, total average scores on Forms B and later post-Lackland Form 77A ratings for OCS Classes 48-A and 48-B were correlated. The 77A scores were derived from ratings made in the field six months and

*In computing these correlations, the primary and secondary ratings were summed for each variable. All correlations in this table are significant at the 1 per cent level.

one year after Classes 48-A and 48-B were rated on Form B at Lackland. The obtained correlation of .34 is encouraging since it is free of local reputation bias, and is based on ratings considerably separated in time with opportunities for pertinent changes to take place in ratees. Moreover, this coefficient of .34 may be regarded as a conservative estimate of the relationship between ratings on Form 77A by different raters, at different stations and at different times, since Form B, despite its similarity and relationship to Form 77A, falls short of identity. It is planned to correlate successive ratings on Form 77A as soon as these become available.

Effect of Varied Order of Administration of Form 77A and Evaluation Scale, Form B. As stated above, evaluations of OCS Class 49-A were obtained on both Form 77A and Form B. It was believed that the rating obtained second might have higher reliability, regardless of actual merit, than the rating obtained first because of practice effect and clarification of raters' opinions. Therefore, it was decided to vary randomly the order of administration of the two forms so that any such effect might be canceled out.

The results were somewhat unexpected. Not only was Form B significantly lower in reliability, as discussed earlier, but, as indicated in Table 9, when Form B was administered first and Form 77A second, the reliabilities of both rating scales were significantly lower than when Form 77A was administered first, indicating that the differences probably did not arise by chance factors alone; i.e., chance differences in raters or ratees. The Form B ratings were not only less reliable than the Form 77A ratings, but, when Form B was administered first, the reliability of the Form 77A ratings was lowered. However, the reliability of Form B appears markedly increased by the immediately prior administration of Form 77A.

TABLE 9

RELIABILITIES OF PAIRED RATINGS ON FORM 77A AND PAIRED RATINGS ON FORM B FOR DIFFERENT ORDERS OF RATING

Reliability Estimate[a]	Order of Rating: Form B First N	Form B First r	Form 77A First N	Form 77A First r	Sign. Level of Diff.[b]
Dual-rating reliability, Form B	84	.45	88	.72	.02
Dual-rating reliability, Form 77A	84	.65	90	.83	.01

[a] For the sum of two ratings as estimated by application of the Spearman-Brown prophecy formula to correlation of the paired ratings.

[b] Estimates of the significance of these differences were made by converting the r's to Fisher's z function.

Conclusions

The relatively high agreement between independent raters on the traits of officer quality indicated that the Evaluation Scale for Graduates of USAF OCS, developed at Human Resources Research Center, is a fairly reliable as well as relevant criterion of proficiency in a population of newly commissioned USAF officers if the average of two independent ratings is employed. However, it was deemed advisable to replace the HRRC evaluation scale by the USAF Form 77A as the criterion of on-the-job officer proficiency against which predictive devices will be validated in the future. This decision has been influenced by the following factors: (1) the 77A has recently been adopted on an Air Force-wide basis as an officer evaluation device; (2) the 77A correlates very highly with the HRRC scale and is at least as reliable.

Cluster analysis of the HRRC evaluation scale suggest that proficiency of newly commissioned Air Force lieutenants may consist of at least three important traits: (1) administrative or executive ability, (2) conscientiousness, and (3) cooperativeness. These traits are not entirely independent but do represent three possible aspects of officer proficiency. Predictive devices should be developed to measure traits of this nature.

From the comparison of different methods of administration of the HRRC Form A, it seems probable that reliability of on-the-job ratings varies directly with the amount of prior briefing of raters as to what qualities to observe and the amount of individual attention each rater receives at the time the rating is made. An improved "Hints to Raters" sheet, incorporating illustrative material alongside of text, is planned for use just prior to procurement of future ratings.

The study of the influence of order of administration upon the reliabilities for the Form 77A and Form B indicates a possible tendency for the reliability of the scale administered second to be affected by the characteristics of the scale administered first. If this apparent tendency is verified by further research, it will have important implications for studies where an experimental rating scale is administered after an operational rating has been made.

BIBLIOGRAPHY

1. CARTER, L.F. (Ed.) Psychological research on navigator training, AAF Aviation Psychology Program Research Report No. 10, Washington, D.C.: Govt. Printing Office, 1947.

2. COWLES, J.T., DAILEY, J.T., and KELLER, R.J. Development of evaluative and predictive measures in the AF Officer Candidate School, Research Bulletin 47-1, Barksdale AFB: Headquarters, Air Training Command, Oct 47.

3. CRAWFORD, M.P. et al (Eds.) Psychological research on operational training in the Continental Air Forces, AAF Aviation Psychology Program Research Report No. 16, Washington, D.C.: Govt. Printing Office, 1947.

4. DAILEY, J.T., and GRAGG, D. B. Postwar research on the classification of aircrew, Research Bulletin 49-2, Lackland AFB: Air Training Command Human Resources Research Center, Nov 49.

5. LINDQUIST, E.S. Statistical analysis in educational research, Cambridge: Riverside Press, 1940.

6. PRESTON, HARLEY. The development of a procedure for evaluating officers in the United States Air Force, American Institute for Research, Pittsburgh: Jul 48.

7. WHERRY, R.J. Validation of a program for selection of officers for retention in the peacetime army, Study No. 704, PRS-SSU, PR-4011-19, Jul 45.

APPENDICES

APPENDIX A

EVALUATION SCALE FOR GRADUATES OF USAF OCS

Form A

Last Name _____ First Name or Initials _____ Group Number _____

Duty Assignment _____ Sqd Number _____ Wing Number _____

The person named above has worked in this assignment for ___ months. He has worked under my supervision for ___ months, and has been known to me for approximately ___ months.

Signature _____ Rank _____

_____ Sqd Group Wing

* * * * * * * * * * * * * * * * * * *

Instructions

You are asked to rate a graduate of the USAF OCS on the characteristics listed in this booklet. If there is any doubt as to the meaning of any item, secure clarification from your interviewer.

Following each item or trait, you will find four descriptive statements. It is realized that not all of the statements will EXACTLY describe any given officer. However, you are asked to use your best judgment in deciding which ONE of the four MOST NEARLY describes him. Avoid any predetermination of what proportions of officers fall at each level. Judge each officer as a single, unique individual; the statements intentionally are not relative to an "average."

Make a ✓ to the left of the statement you select. Circle your check mark if you lack sufficient evidence to make an honest rating on a trait.

1. Understanding Instructions.

___ Understands instructions when given in detail.

___ Is slow to grasp instructions.

___ Grasps instructions completely and accurately.

___ Grasps quickly the main points of instructions.

2. Planning and Scheduling Work.

___ Has relatively few delays in scheduling projects due to poor planning.

___ Progress on projects is smooth and final results indicate good planning and scheduling.

___ Progress on projects is uninterrupted, but final results indicate minor errors in planning or scheduling.

___ Has delays in completing projects due to faulty scheduling or planning.

3. Presenting Finished Work.

___ Presents fully documented findings so organized that his superior can make a decision quickly and with confidence.

___ Presents his findings in such form that his superior can locate the required information to make a decision.

___ Presents his findings in such form that they give an incomplete basis for a decision.

___ Presents his findings so that it is necessary for his superior to organize the material before making a decision.

4. Securing Cooperation.

___ Secures cooperation on important matters from most of those concerned with his project.

___ Secures cooperation in all phases of his project by dealing tactfully with those concerned.

___ Enlists the full and active support of all concerned with his project through his tactful and persuasive manner.

___ Makes no attempt to secure the cooperation of those concerned with his project.

5. Keeping Others Informed.

___ Usually informs people concerned of activities and developments.

___ Neglects to inform people concerned of some activities and developments.

___ Keeps affected persons thoroughly briefed on all activities and developments.

___ Keeps concerned persons well informed on essential activities and developments.

6. Delegating Authority.

___ Delegates authority properly and maintains good operating efficiency.

___ Makes overlapping or vague delegations of authority.

___ Delegates authority well enough to secure a fair degree of efficiency.

___ Makes clear cut delegations of authority which result in a high degree of operating efficiency.

7. Giving Orders and Instructions.

___ Is vague, indefinite or contradictory in giving orders or making assignments.

___ Issues orders in such a manner that very few questions are raised concerning what is to be done.

___ Orders are so clear cut and concise that no questions are raised concerning what is to be done.

___ Gives orders so that no more than the expected number of questions are asked.

8. Setting a Good Example for Work Performance.

___ Realizes the full importance of good habits and goes out of his way to set a good example for his subordinates.

___ Generally sets a good example for his subordinates to follow in their work habits.

___ Sets a poor example by his work habits.

___ Is occasionally erratic and unpredictable in his work habits.

9. Looking Out for Subordinates' Welfare.

___ Makes some extra effort to help subordinates and promotes their welfare.

___ Is aggressive in his efforts to assist subordinates and to improve their welfare.

___ Pays no particular attention to the needs and welfare of his subordinates.

___ Looks after the welfare of his subordinates adequately.

10. Problem Solving.

___ Attacks new problems in a haphazard or unorganized manner.

___ Shows great resourcefulness and ingenuity in the attack and quick solution of problem situations.

___ Attacks new problems methodically and gives an adequate solution in a reasonable amount of time.

___ Attacks new problems methodically but takes an excessive amount of time to solve the problem.

11. Judgment.

___ Shows unusually good judgment in rendering decisions.

___ Makes poor or hasty judgments.

___ Can be depended upon to use good judgment in almost every case.

___ Usually shows good judgment.

12. Assumption of Responsibility.

___ Assumes responsibility only when specifically requested.

___ Eagerly assumes additional responsibility.

___ Can be depended upon to assume responsibility when necessary.

___ Avoids assuming responsibility whenever possible.

13. Complying with Orders and Directives.

___ Carries out an order or follows a directive without undue delay.

___ Occasionally avoids compliance with an order or directive.

___ Occasionally delays in complying with an order or directive.

___ Carries out promptly and effectively the spirit and intent of an order.

14. Reaction to Organizational Procedure.

___ Carries through organizational procedures routinely.

___ Occasionally "slips up" on organizational procedures.

___ Accepts and promotes an understanding of organizational procedures.

___ At times, disregards organizational procedures.

15. Subordinating Personal Interest.

___ Disregards personal welfare and interests in the successful performance of duty.

___ Subordinates personal desires to the performance of duty.

___ Allows personal interest and welfare to interfere with performance of duty.

___ Accepts minor personal inconvenience in order to perform his duties.

16. Cooperating with Associates.

___ Gives some assistance to associates when requested to do so.

___ Cooperates grudgingly with associates and creates some dissension among them.

___ Voluntarily assists associates and establishes smooth and effective relationships with them.

___ Cooperates willingly with associates and maintains pleasant relationships with them.

17. Loyalty to Total Organization.

___ Exhibits a very active and inspirational interest in his organization.

___ Sometimes pessimistic or "gripes" about his organization.

___ Is interested in the total program of his organization.

___ Is relatively indifferent to the function of his organization.

18. Attending to Details.

___ Gives careful attention to most details of his job.

___ Handles satisfactorily the important details of his job.

___ Makes sure that all details of his job are completely taken care of.

___ Occasionally neglects routine details of his work.

19. Keeping Appointments.

___ Is sometimes late for appointments, but usually notifies the one concerned.

___ Is on time for appointments and always notifies other when he is unable to report.

___ Fails to keep or is often late for appointments.

___ Occasionally misses appointments, but notifies those concerned.

20. Improving Effectiveness.

___ Accepts most opportunities to improve his proficiency or potentiality.

___ Has not made use of opportunities offered him to improve his job effectiveness.

___ Is alert to an opportunity to improve his effectiveness.

___ Seeks out opportunities to improve his job proficiency and potentiality.

21. Being Fair and Scrupulous.

___ Uses direct and forthright methods in most of his contacts.

___ Occasionally uses questionable methods to gain an end.

___ Is satisfactorily accurate, frank, or fair in statement and practice.

___ Is scrupulously accurate, frank, and fair in statement and practice.

22. Maintaining Military Appearance.

___ Has a good appearance, is neat and well groomed.

___ Neglects dress and grooming.

___ Has an outstandingly smart military appearance.

___ Usually neat but is sometimes careless with dress and grooming.

23. Self-Confidence.

___ Usually confident of his ability to handle a situation but needs some assistance.

___ Has self-confidence required for most situations.

___ Consistently handles situations in a poised, confident manner.

___ Shows some lack of confidence in his own ability.

24. Temperament.

___ Invariably on an even keel, reasons calmly in the most trying situations.

___ Gets irritated or upset easily.

___ Even-tempered, rarely lets anything bother him.

___ Generally calm and even-tempered, but occasionally gets upset or rattled.

APPENDIX B

HINTS TO RATERS

Ratings are used where more objective tests or measures are not feasible. Ratings on carefully constructed rating scales can be quite objective and reliable, but only to the extent that the raters exercise certain necessary precautions. The most important of these are discussed below.

1. Avoid the "halo effect." This is the most common error made by untrained raters. They confuse their judgment of one trait by their judgment of another more obvious trait. For example, a secretary is judged to be an excellent typist merely because she is personable, good-looking, or intelligent. Items on a rating scale used by untrained raters frequently correlate highly, despite the actual independence of these traits, simply because the raters allow their judgments to be warped by certain more obvious traits possessed by the ratee. Rate only the trait in question; resist being influenced by irrelevant characteristics.

2. Avoid the error of leniency. It is a common tendency of untrained raters to assign more judgments above average than below average. This not only spuriously raises the "average" but also tends to reduce the range of the rating scale for the higher degrees of the trait being rated. This error is particularly apt to occur in the rating of desirable traits and in the rating of individuals we know quite well. However, some judges commit the opposite error when they displace all ratings toward the lower end of the scale. Keep in mind just the individual you are rating, and not some "ideal" group.

3. Avoid the error of central tendency. Raters frequently avoid giving extreme ratings, either high or low. They tend to rate too many individuals as near a middle point. This is a failure to distinguish between individuals and a failure to use the entire range of the scale. It is particularly difficult to distinguish between individuals who are not well known or not outstanding in a particular trait. The extremes of the scale have been intentionally designed to apply at least to some of the individuals you will rate.

4. Independence of raters. The reliability of a rating scale is determined by the consistency with which the same individual is rated alike by two or more raters. Many scales show spuriously high reliability, or their true reliability is indeterminate, simply because the judges are not making independent ratings. They discuss the ratees and their judgments consequently become similar. Conscientiously avoid utilizing someone else's judgment when making your own rating.

APPENDIX C

AREAS AND ITEMS COVERED BY AF FORM 77A, REPORT OF OFFICER EFFECTIVENESS

I. Proficiency in Handling Administrative Details

1. Understanding instructions
2. Scheduling work
3. Getting information from records
4. Getting ideas from others
5. Checking accuracy of work
6. Writing letters and reports
7. Getting cooperation
8. Presenting finished work
9. Keeping records
10. Keeping others informed
11. Rendering effectiveness reports

II. Proficiency in Supervising Personnel

12. Matching personnel and jobs
13. Delegating authority
14. Giving orders and instructions
15. Insuring comprehension
16. Giving reasons and explanations
17. Supporting authorized actions
18. Encouraging ideas
19. Developing teamwork
20. Setting a good example
21. Assisting subordinates in their work
22. Evaluating subordinates' work
23. Looking out for subordinates' welfare
24. Maintaining relations with subordinates

III. Proficiency in Planning and Directing Action

25. Taking responsibility
26. Solving problems
27. Making use of experience
28. Long-range planning
29. Taking prompt action
30. Suspending judgment
31. Making correct decisions
32. Making forceful efforts
33. Absorbing materials

IV. Acceptance of Organizational Responsibility

34. Complying with orders and directives
35. Accepting organizational procedure
36. Subordinating personal interest
37. Cooperating with associates
38. Showing loyalty
39. Taking responsibility for subordinates

V. Acceptance of Personal Responsibility

40. Attending to duty
41. Attending to details
42. Reporting for appointments
43. Meeting commitments
44. Being fair and scrupulous
45. Maintaining military appearance
46. Adapting to associates
47. Adapting to the job
48. Conforming to civil standards

VI. Proficiency in Duty Military Occupational Specialty

49. Possessing fundamental training
50. Improving effectiveness
51. Keeping well-informed
52. Applying training and information
53. Showing ingenuity in specialty
54. Handling related assignments