12-1963

# The Application of the Psychophysical Law to the United States Army Officer Efficiency Reporting System

James M. Sturgeon
*University of Tennessee - Knoxville*

To the Graduate Council:

I am submitting herewith a thesis written by James M. Sturgeon entitled "The Application of the Psychophysical Law to the United States Army Officer Efficiency Reporting System." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Industrial and Organizational Psychology.

G. H. Whitlock, Major Professor

We have read this thesis and recommend its acceptance:

Charles P. White, James W. Burnett, A. J. Kelly

Accepted for the Council:
Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

December 11, 1963

To the Graduate Council:

I am submitting a thesis written by James M. Sturgeon entitled "The Application of the Psychophysical Law to the United States Army Officer Efficiency Reporting System." I recommend that it be accepted for nine quarter hours credit in partial fulfillment of the requirements for the degree of Master of Science, with a major in Industrial Management.

Major Professor

We have read this thesis and recommend its acceptance:

Accepted for the Council:

Dean of the Graduate Council

2 2
33

# THE APPLICATION OF THE PSYCHOPHYSICAL LAW TO THE UNITED STATES ARMY OFFICER EFFICIENCY REPORTING SYSTEM

---

A Thesis

Presented to

the Graduate Council of

The University of Tennessee

---

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

---

by

James M. Sturgeon

December 1963

ACKNOWLEDGEMENTS

The writer wishes to express his appreciation to Mrs. Georgia Troglen who typed and assisted in editing the final manuscript.

Also, my sincere appreciation to my wife, Mary Katherine, for her encouragement and assistance throughout the study.

The writer is particularly indebted to Dr. G. H. Whitlock of The University of Tennessee whose stimulating criticism and patient guidance made this study possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

# CHAPTER I

# INTRODUCTION

## I. GENERAL INTRODUCTION

When an employer has only a few employees and can observe their job performances personally, there is little need for a formal performance evaluation system. However, such is not the case in large organizations where decision-making with regard to personnel is necessarily centralized at a high echelon. In the latter instance the relationship between personnel manager and employee is likely to be impersonal.

Due to the size of the organization and other factors, the United States Army has employed some type of formal officer performance evaluation system as a basis for personnel decisions since the early 1900's.[1] The present evaluation plan was implemented in September, 1961. Like its predecessors, the current system is judged to be deficient in several respects.[2] Lack of an acceptable scheme for quantifying subjective judgement has been a major problem. One high ranking Army officer who has been closely associated with the development

---

[1]Headquarters, Department of the Army, The New Officer Efficiency Reporting System, Officer's Call, Department of the Army Pamphlet 355-25 (United States Government Printing Office, 1961), p. 2.

[2]Opinion expressed by Major General Rush B. Lincoln at an Army Officer's luncheon, Knoxville, Tennessee, April 17, 1963.

of the present system has gone so far as to state flatly that "the use of scales that produce a 'number' or a 'score' suggests a degree of precision beyond the reach of the social scientist."[3]

S. S. Stevens in stating his views with regard to the problem of quantifying human judgement has said that

> . . . when people try to describe a sensation in quantitative terms they face a difficult task, and the factors that affect the outcome are numerous and subtle. Patience and experimental skill can probably clean up part of the variance, but there will always remain irreducible dispersions to set a level below which we sink into uncertainty.[4]

Recent work in the field of psychophysics has helped to "clean up part of the variance" to which Stevens has referred. Experimental studies by Stevens of at least twenty prothetic continua such as loudness, brightness and electric shock have disclosed that a power function describes the relationship between many physical and psychological magnitudes. The Psychophysical Law is stated by Stevens as "equal stimulus ratios produce equal sensation ratios and is expressed as a power function."[5]

Stevens' findings prompted G. H. Whitlock to investigate the applicability of the Psychophysical Law to performance evaluation. Whitlock found that in all of four different situations studied, a

---

[3]Lieutenant Colonel Robert C. Storey, "The New Efficiency Report," Army, 12:30, August, 1961.

[4]S. S. Stevens, "On the Psychophysical Law," Psychological Review, 65:167, May, 1957.

[5]S. S. Stevens, "Problems and Methods of Psychophysics," Psychological Bulletin, 55:177, July, 1958.

power function described the relationship between observation of performance and evaluation of performance.[6]

This thesis is a further investigation of the applicability of the Psychophysical Law to the performance evaluation process, using the United States Army Officer Efficiency Reporting System as a basis for experimentation.

## II. STATEMENT OF THE PROBLEM

In rendering the United States Army Officer Efficiency Report, the rater is required to describe in narrative form the performance of the ratee. Too, the rater must express quantitatively his overall evaluation of the performance by placing the ratee in an adjectival, quantitative category. Army regulations declare it is imperative that the rater be consistent in these two portions of the report and that the categorical rating be justified by the narrative description.[7] Clearly some measure of "consistency" or "justification" is needed, yet none is prescribed or even suggested.

Apart from the above two evaluations, the Army rater is required to make a categorical evaluation of the ratee's personal qualities such as ambition, stamina, and dependability. The rater is instructed

---

[6]Gerald H. Whitlock, "The Application of the Psychophysical Law to Performance and Morale Evaluations" (Paper read to the American Psychological Association, St. Louis, Missouri, September, 1962).

[7]Headquarters, Department of the Army, Personnel Efficiency Ratings, Officer Efficiency Reports, Army Regulation 623-105 (United States Government Printing Office, 1963), p. 14.

to evaluate personal qualities and manner of performance separately, not allowing personal qualities to affect the numerical overall demonstrated performance score.[8]

The following hypotheses are to be tested by this thesis: (1) that a power function describes the relationship between observation of performance as narratively described and the overall demonstrated performance score; (2) that this function obtains regardless of whether the overall, quantitative evaluation is a response to observation of actual performance or a response to a written set of performance specimens which describe a hypothetical performance; and, (3) that the degree to which the ratee is deemed by the rater to possess or lack specified personal qualities has an effect on the quantitative overall demonstrated performance score.

## III. IMPORTANCE OF THE STUDY

The performance specimen checklist has received much attention of late and is often regarded as the most promising of several performance evaluation techniques. However, it appears that this method would have serious shortcomings when used by a large organization such as the Army. To develop one suitable checklist for use in rating performance of diverse duties by 100,000 Army officers would be difficult, if not impossible. Instead of a checklist, designers of the current

---

[8] Ibid., p. 17.

Army evaluation system chose to allow each rater personally to determine and report in narrative form relevant incidents of performance.

The Army's narrative description of performance technique also has definite limitations, one of which is that it is cumbersome for personnel managers to work with where large numbers of personnel are involved. For this reason, Army raters are required to quantify their evaluations.

If the results of this study indicate that the quantitative judgement of raters is lawfully related to the stimulus intensity, that is, the ratio of uncommonly effective to uncommonly ineffective performance specimens observed and narratively described, some grounds exist for the elimination of the controversial overall quantitative performance rating. Should top personnel managers be reluctant to eliminate the overall rating, perhaps the mathematical function, if discovered can be used as a measure of consistency between the narrative and quantitative ratings.

If it is shown that Army raters allow personal qualities ratings to influence the overall demonstrated performance score, revision of the system, additional instruction for raters, or adoption of a technique for making allowance for the bias will be in order.

A final aspect of this inquiry is of consequence. In addition to investigating further the applicability of the Psychophysical Law, a technique for determining stimulus intensity used recently by J. F. Chisholm in applying the Psychophysical Law to job evaluation is used

in this study.[9] Should Chisholm's methods prove equally useful in this realm of performance evaluation, another modest contribution to knowledge will have been made.

## IV. DEFINITION OF TERMS

Psychophysics--The discipline concerned with the responses that organisms make to the energies of their environment.

Psychophysical Law--Equal stimulus ratios produce equal sensation ratios. The mathematical expression is a power function.

Prothetic Continua[10]--"Continua having to do with how much---. Those continua on which discrimination is mediated by an additive or prothetic process at the psychological level."

Performance Specimen[11]--"An incident of relevant performance which is uncommonly effective or ineffective."

Form--The form used in this thesis is a modified version of Department of the Army Form 67-5 which is entitled U. S. Army Officer Efficiency Report. A copy of this form appears in Appendix I.

---

[9]Jack F. Chisholm, "The Function of Job Specimens in the Evaluation of Job Worth" (unpublished Master's thesis, College of Business Administration, The University of Tennessee, Knoxville, 1963).

[10]S. S. Stevens, "The Psychophysics of Sensory Function," American Scientist, 48:226-253, June, 1960.

[11]Whitlock, op. cit., p. 1.

## V. SOURCES AND METHODS

To obtain data to test the stated hypotheses, the cooperation of 101 Army officers enrolled in an advanced course of instruction at one of the U. S. Army service schools was obtained. Each officer was asked to recall and evaluate the personal qualities and job performance of the last officer whom he had actually rated or the officer with whom he had worked most closely on his last assignment. Ratings were made on a modified United States Army Officer Efficiency Report form and in accordance with current Army regulations.

The second method of securing data involved the presentation of six partially completed rating forms to the previously mentioned group of Army officers. The personal qualities evaluation and the narrative performance description were completed on each form by the writer. The ratings were constructed so as to possess systematically varied stimulus intensities. Each rater was asked to read both the personal qualities evaluation and the performance description on each form and respond by scoring the overall demonstrated performance on the basis of the information presented.

Methods of psychophysics were used in treating the data obtained. The ratio of effective performance specimens to ineffective performance specimens contained in each narrative description of performance was computed. Where only effective or ineffective specimens occurred in a description, the mere number of such specimens were determined. These data were then plotted against the categorical ratings on log-log paper

to determine if the Psychophysical Law obtained.

A similar procedure was followed with regard to the data obtained from the officers who scored categorically the overall demonstrated performance of hypothetical ratees on the basis of presented narrative performance descriptions and quantitative personal quality ratings. In this experiment, however, the E/I ratios were positively known since they were designed into the evaluations by the writer.

Standard statistical techniques were used to determine the relationships between the personal qualities evaluations and the overall demonstrated performance scores.

## VI. ORGANIZATION OF THE STUDY

Chapter I of this thesis serves to introduce the problem, state the purpose of the study, define certain terms, emphasize the importance of the undertaking and briefly describe the sources of data and methodology used.

Chapter II is a history of the problem and resume of related literature.

Chapter III describes how the data were obtained and analyzed.

Chapter IV presents the results of the investigation.

Chapter V is a discussion of the findings set forth in Chapter IV.

Chapter VI consists of a summary of the study and a statement of conclusions.

## CHAPTER II

## HISTORY OF THE PROBLEM

This chapter is divided into three parts. The first part presents a brief history of officer performance evaluation in the United States Army. Part two concerns the development of the field of psychophysics. The third portion is a review of the work of researchers who have investigated the usefulness of the Psychophysical Law in coping with personnel performance evaluation problems.

### I. THE HISTORY OF THE U. S. ARMY EFFICIENCY REPORT

Army officer performance evaluation in one form on another is almost as old as the Army itself. The early efficiency reports were not a part of an Army-wide system, but were of an informal nature. Shown below is an excerpt from a reported submitted in 1813.

Lower Seneca Town, August 15th, 1813

Sir:
I forward a list of the officers of the --th Regt. of Infty. arranged agreeable to rank. Annexed thereto you will find all the observations I deem necessary to make.

Respectfully, I am, Sir,
Yo. Obt. Sevt.,
Lewis Cass

--th Regt. Infantry
Alexander Brown--Lt. Col, Comdg.--A good natured man.
Clark Crowell-first Major--A good man, but no officer.
Jess B. Wordsworth--2nd Major--An excellent officer.

> Capt. Shaw--A man of whom all unite in speaking ill--A knave despised by all. . .[1]

In 1918 the War Department issued a formal directive requiring that each officer below the rank of brigadier general be rated quarterly in accordance with specific rating instructions. This early formal system required each rater to construct a rating scale for each of "five essential qualifications of an officer, namely; (1) Physical Qualities, (2) Intelligence, (3) Leadership, (4) Personal Qualities, and (5) General Value to the Service."[2]

To construct these scales the rater was instructed to write down the names of 12 to 25 officers he had known. These officers were to be of the rank immediately above the officer to be rated. From this list, the rater selected the Highest, High, Middle, Low, and Lowest officer with regard to each trait. The ratee was compared with regard to each trait to the officers who comprised the respective scales. Each of the five categories on each scale carried a quantitative value. The ratee was given the value of the officer to whom he was most nearly equal.[3]

---

[1]The Officers Guide (Harrisburg: The Military Press Publishing Company, 1956), p. 417, citing the Adjutant General's School Bulletin, April, 1942.

[2]Thomas L. Whisler and Shirley F. Harper, Performance Appraisal (New York: Holt Rinehart and Winston, 1962), pp. 216-218.

[3]Ibid.

Evidently the system just described was unsatisfactory since it was replaced by a new procedure in the early 1920's. The new system consisted of graphic scales and was known as Form 67. Later the form was revised, replacing the adjectival categories with numerical scores and providing the rater with space to comment on the ratee's qualifications.[4]

Form 67 was used during the entire period between World War I and World War II and was considered "quite typical and respectable" when compared to other rating systems of the day. However, over the years the "leniency" tendency on the part of raters eroded the value of the system. In other words, almost all officers received the same high ratings. Not until about 1940 when war was imminent did the seriousness of the situation come to the forefront.

To command the rapidly expanding Army, it was necessary to promote a large number of officers to high ranks. Though efficiency reports had been collected for such purposes, it was found that they provided no sound basis on which to make the requisite decisions. Instead of pointing out a few truly capable officers, the efficiency reports showed that of 4000 officers generally eligible, 2000 were superior and best. Facing an emergency situation, those charged with the task of choosing the men to lead the nation into war resorted to their per-

---

[4]Headquarters, Department of the Army, The New Officer Efficiency Reporting System, Officer's Call, Department of the Army Pamphlet 355-25 (United States Government Printing Office, 1961), p. 2.

sonal knowledge of the men under consideration as a basis for their decisions.[5]

In 1945 the entire professional staff of the Personnel Research Section of the Adjutant General's Office set out to fashion a truly valid efficiency reporting system. The result of this effort was an entirely new report form which employed the "forced choice" method of rating. Basically the new system required the rater to choose from several sets of four adjectives or phrases called tetrads, the adjective or phrase which best described the officer and the adjective or phrase which was least applicable. Pairs of adjectives or phrases were grouped so as to appear to be of equal value but actually to differ in their significance for success as an officer.[6]

The new system was put into effect in 1947 after preliminary tests showed that it produced a better distribution, was less biased by the rank of the ratee, and resulted in ratings more valid than those produced by the old system.[7] However, the new system was not popular with raters. Raters did not like not being able to know what kind of rating they were giving. That some of the tetrads contained no phrases which accurately described the ratee was objectionable.[8]

---

[5]E. Donald Sission, "Forced Choice--The New Army Rating," Personnel Psychology, 1:365-381, Autumn, 1948.

[6]Ibid. [7]Ibid.

[8]Headquarters, Department of the Army, The New Officer Efficiency Reporting System, Officer's Call, (Government Printing Office, 1961), p. 2.

Lastly, it appears that the fundamental obstacle was that officers simply did not relish being subjected to a system, the inner workings of which were veiled in secrecy.

In 1950 the forced choice system was scrapped in favor of a more conventional rating method. The 1950 version of the rating form consisted of several graphic and numerical scales and a narrative description of the ratee. Except for minor revisions, this rating method was used until the present system was adopted in 1961. The old system was discarded primarily because it had grown to lack validity and acceptability.[9]

The current Army Efficiency Report embodies graphic scales and several other features of previous report forms, but it also has some new aspects. Under the old system, numerical values entered by the rater were converted to appropriate index numbers on a standard scale at Department of the Army level. Values entered on the new form are not converted. The new system requires that each ratee be counseled about four months prior to the time that his Efficiency Report is prepared. It is mandatory that the ratee not be shown actual reports concerning him. To reduce unwarranted high and low ratings, raters must justify extremely high or low ratings with factual evidence.[10]

Perhaps the greatest innovation in the new system is the strong emphasis on performance of current duty. Whereas the previous system

---

[9] Ibid., pp. 2-3.

[10] Ibid., pp. 3-4.

required that raters describe the rated officer personally, the current report calls for a specific and factual narrative description of how the ratee performed his job.[11] In the scored section of the form, the rater must in essence quantify his response to the observed performance which he has already described narratively. In making this judgement, the rater is instructed to insure that the overall numerical evaluation is consistent with the narrative description and not unduly influenced by personal traits of the ratee. As spelled out in Chapter I, it is with this aspect of the system that this study is concerned.

## II. DEVELOPMENT OF PSYCHOPHYSICS

Psychophysics is the discipline concerned with the responses that organisms make to the energies of their environment. The field is not a new one. On the contrary, certain aspects of it are actually quite old. For centuries it has been recognized that sensory experience varies in intensity. In the mid-1800's Herbart submitted that there was a lower limit of sensation. During the same period Weber proposed the law which states that the just noticeable increment in a stimulus is proportional to the stimulus.[12]

Though Weber's law is not a psychophysical law since it relates two physical measurements, it provided the groundwork for Fechner's

---

[11]Lieutenant Colonel Robert C. Storey, "The New Efficiency Report," Army, 12:34, August, 1961.

[12]J. P. Guilford, Psychometric Methods (New York: McGraw-Hill Book Company, Inc., 1954), p. 3.

inquiry into the nature of the relationship between physical and psychological magnitudes. After extensive research, Fechner announced the law which bears his name and states that the strength of a response is proportional to the logarithm of the stimulus.[13]

Plateau disagreed with the law which Fechner proposed and suggested that the proper relationship between the two magnitudes was a power function.[14] Though Stevens has now presented conclusive evidence that Plateau was right and Fechner was wrong, it was Fechner's erroneous "law" which provided the basis for psychophysics for more than a century. Such an unfortunate beginning no doubt retarded the growth of the field.

Specifically, Stevens has shown that for more than twenty prothetic continua such as smell, brightness, heaviness and loudness that the magnitude of the sensation grows as a power function of the stimulus intensity.[15] The form of the equation is $Y = KS^b$, where Y = psychological magnitude; S = physical magnitude; and b = power of function.

Investigation of this sort is of genuine academic interest but of little practical value if treated as an end in itself. Such treatment of psychophysics by some in the field has sometimes caused the discipline to be considered inconsequential. Stevens comments on the value of psychophysics are as follows:

---

[13] Ibid.

[14] S. S. Stevens, "The Psychophysics of Sensory Function," American Scientist, 42:226-227, June, 1960.

[15] Ibid., p. 227.

> Seeking the laws that relate the responses of men and animals to the energetic configuration of their environment, it probes matters of deep human interest, and matters that often make a difference in the market place.[16]

## III. RECENT APPLICATION OF THE PSYCHOPHYSICAL LAW TO PERFORMANCE EVALUATION

Whitlock and student researchers at The University of Tennessee have devoted considerable effort over the past three years to determining the utility of psychophysical methods in the realm of personnel management. Experiments utilizing psychophysical concepts have been conducted with regard to the complex psychological processes involved in performance evaluation, job evaluation, and morale estimation. Performance evaluation has received by far the greatest amount of attention. The remainder of this chapter is a brief summary of a published article written by Whitlock which describes the research in this area.[17]

Experimentation concerning performance evaluation has been based on the assumption that performance evaluation is a response to observation of relevant incidents of performance which are uncommonly effective or uncommonly ineffective. Such incidents are termed performance specimens. The basic hypothesis has been that the rater's overall evaluation bears a lawful relation to the number of performance specimens observed.

---

[16]S. S. Stevens, "Problems and Methods of Psychophysics," Psychological Bulletin, 55:177, July, 1958.

[17]Gerald H. Whitlock, "Application of the Psychophysical Law to Performance Evaluation," Journal of Applied Psychology, 47:15-23, February, 1963.

The true relationship in question was first believed to be a logarithmic function. Research by Beard and Hedge disclosed that neither a logarithmic nor linear function was appropriate.[18]

Whitlock, prompted by an article written by Stevens and published in 1960, investigated the possibility that a power function was the proper description of the relationship. Data collected by Hedge were plotted by Whitlock on log-log paper. A good straight line fit resulted. Thus, for these data the Psychophysical Law obtained. This operation involved the plotting of the number of effective and ineffective specimens separately. Further investigation seemed to indicate that the proper way to combine effective and ineffective specimens was to use the logarithm of the quotient obtained by dividing the number of effective specimens by the number of ineffective specimens.[19]

To probe deeper into the matter, Whitlock and students under his direction have since conducted a number of experiments. Data previously gathered from a study concerning the Tennessee Valley Authority were reanalyzed. In this case, too, the power function properly described the relationship between observation and evaluation. Similar treatment of data collected in 1957 from executives at a large chemical plant again conformed to the Psychophysical Law, thus adding substance to early findings.

---

[18] Ibid.

[19] Ibid.

An experiment involving checklist ratings by supervisors in four plants in East Tennessee and a study concerning college students' responses to a set of specimens of teacher classroom behavior both resulted in the conclusion that observation and evaluation are related by the Psychophysical Law. A further inquiry revealed that the power function described the relationship in question even though the raters evaluated an actual performance and afterwards indicated on a checklist the performance specimens which they had observed.

The next experiment consisted of the presentation to student raters of forms on which only the number of performance specimens observed for hypothetical ratees was listed. Ratings made on the basis of such numbers also obeyed the Psychophysical Law. The most recent experiment was conducted in 1962 for the purpose of determining whether or not category scale performance evaluation values, when plotted against ratio scale values, would result in a curve concave downward. Such a curve, a universal characteristic of prothetic continua according to Stevens, did result.

Since the nature of the relationship between stimulus and response in the performance evaluation process seems to be established, Whitlock has submitted that it is now possible to set confidence limits for evaluations based on corresponding observational values. Whitlock further proposes that evaluations can even be dispensed with when observational values are known and equations are established.[20]

---

[20] Ibid.

This chapter has been devoted to supplying the reader with background information concerning the thesis problem. Specifically, a brief history of each of the following areas has been presented: (1) officer performance evaluation in the United States Army; (2) the field of psychophysics; and (3) application of the Psychophysical Law to performance evaluation. Chapter II which follows describes the design of the experiment used in testing the hypotheses stated in Chapter I.

# CHAPTER III

## METHODOLOGY

The primary purpose of this study was to test the hypotheses that, under the Army Officer Efficiency Report system, the Psychophysical Law describes the relationship between the narrative description of a performance and the overall categorical evaluation of the same performance. Secondly, the study was designed to measure the effect of Army officer personal qualities upon the quantitative performance rating.

### I. SOURCE OF DATA

In order to reach sound conclusions, it was felt that authentic Army raters should be used as the source of data. Hence, permission was sought and obtained from the Commandant of one of the Army service schools to use 101 student officers of a Career Officer Course in the experiment.

Of the officers participating, 1 was a major, 99 were captains, and 1 was a lieutenant. The average length of service time of the participants was approximately 8 years. All officers were familiar with the Army rating system prior to the survey.

The period of class time set aside by the school Commandant was shown on the student's schedule of classes as "Performance Evaluation." It was felt that integration of the survey into normal classroom

activity would result in better student cooperation than if conducted during the student's off-duty time.

The class surveyed consisted of two sections of 50 and 51 students respectively. Data were gathered from the two sections on the same day. Fifty minutes of class time of each section was allocated to the experiment.

## II. PRESENTATION OF FORMS

The survey was executed in two parts hereafter referred to as Phase I and II. After a brief introduction and general explanation of the study, Phase I was administered. This phase required the student raters to recall and evaluate the personal qualities and performance of the last officer whom they had rated or the officer (peer or subordinate) with whom they had worked most closely on their last job. The raters completed Part III-Manner of Performance, Part IV-Personal Qualities, and Part VI (1)-Overall Demonstrated Performance of one Army Efficiency Report form. A copy of the form used and instructions to the raters are shown in Appendix I. Notes on the instruction sheet concerning the Efficiency Report system are excerpts from current Army regulations.[1]

Phase II of the survey consisted of the presentation of six partially completed forms to each rater. Each of the six forms

---

[1]Headquarters, Department of the Army, Personnel Efficiency Ratings, Officer Efficiency Reports, Army Regulation 623-105 (United States Government Printing Office, 1963), pp. 12-20.

represented a hypothetical officer and were arranged in random order. Part III-Manner of Performance and Part IV-Personal Qualities were completed on each form by this writer prior to the survey. Stimulus intensity of the completed parts of the forms was systematically varied. An example of the instruments used and instructions to the raters are presented in Appendix II. Raters were instructed to score Part IV (1)-Overall Demonstrated Performance of each form based on information given in Parts III and IV.

Raters in the first section surveyed completed Phase I first, then Phase II. Raters in the second section completed Phase II first, then Phase I. This procedure was followed to insure that an adequate number of ratings would be received from both phases within the limited amount of time available. Excluding the time consumed by the introduction, raters had approximately 45 minutes in which to make the evaluations.

## III. DESIGN OF FORMS

Phase I of the survey, in which raters recalled and evaluated the personal qualities and performance of an officer whom they had actually observed, required the use of a blank Army Officer Efficiency Report form. The form as published by the Department of the Army was modified to eliminate parts non-essential to the experiment.

Evaluations on forms presented to raters in Phase II of the experiment were constructed so as to have varied stimulus intensities.

Whitlock has shown that stimulus intensity in the performance evaluation process is equivalent to the ratio of effective to ineffective performance specimens.[2] Such findings were accepted and used in this investigation.

Specimens used in the narrative descriptions of performance in Phase II were selected by the writer from personal recollections of actual Efficiency Reports read, from an official Army example of a "good" narrative description,[3] or were based on ideas obtained from various military publications on leadership, officer responsibilities, etc. An effort was made to use specimens of approximately equal weight. This is not to profess that weighting was done scientifically, but simply that use of performance specimens of obviously unequal weight was avoided.

The total number of specimens used in any one narrative description was either nine or ten. Based on the writer's experience, the example published by the Army, and the amount of space provided on the unmodified report form, it was estimated that an average Army Efficiency Report contains approximately ten performance specimens. The use of a nearly identical total number of specimens in each form

---

[2]Gerald H. Whitlock, "Application of the Psychophysical Law to Performance Evaluation," Journal of Applied Psychology, 47:15-23, February, 1963.

[3]Headquarters, Department of the Army, The New Officer Efficiency Reporting System, Officer's Call, Department of the Army Pamphlet 355-25 (United States Government Printing Office, 1961), p. 7.

was intended to reduce the possibility of rater bias due to the length of a description.

A random digit table was used to arrange effective and ineffective specimens within each description. Each hypothetical ratee was identified by a code name. Identification, total number of specimens, and E/I ratios were specifically as shown in Table I.

To measure the influence of Personal Qualities ratings on the Overall Demonstrated Performance scores, the sets of forms used in Phase II were divided into three categories designated A, B, C. Category A forms were designed to contain Personal Qualities scores of half the intensity of the narrative ratings. Category B forms contained Personal Quality ratings equal to the E/I values of the narrative ratings. Category C forms contained Personal Qualities ratings of twice the intensity of the narrative ratings.

Stimulus intensity of the Personal Qualities ratings was varied by use of the E/I ratio technique. Based on the writer's knowledge of rating tendencies of Army officers and at the suggestion of Dr. G. H. Whitlock, a rating .5 (Exemplary) was treated as an effective specimen. A rating of .4 (Above Average) was considered an indication of common behavior and attributed no value. Ratings of .3 (Average), .2 (Below Average), and .0 (Inadequate) were considered ineffective specimens. In designing the ratings, no rating below .3 (Average) was used. Effective, ineffective, and common behavior indicators were assigned on each form so as to be consistent with the accompanying

TABLE I

IDENTIFICATION, NUMBER OF SPECIMENS, AND E/I RATIOS OF THE NARRATIVE DESCRIPTIONS EMPLOYED IN PHASE I

| Name | Effective Specimens | Ineffective Specimens | Total Specimens | E/I Ratio |
|---|---|---|---|---|
| Lt. Green | 8 | 1 | 9 | 8 |
| Lt. Gray | 8 | 2 | 10 | 4 |
| Lt. Black | 6 | 3 | 9 | 2 |
| Lt. White | 6 | 4 | 10 | 1.5 |
| Lt. Brown | 5 | 5 | 10 | 1 |
| Lt. Blue | 3 | 6 | 9 | 0.5 |

narrative evaluation. Effective/ineffective ratios for each category were as shown in Table II.

## IV. ANALYSIS OF DATA

To determine the stimulus intensity of each of the narrative descriptions obtained in Phase I of the survey, the writer read carefully each of the descriptions and marked on each form a plus (+) for each valid statement which indicated effective performance and a minus (-) for each valid statement which indicated ineffective performance. Subjective judgement was used to differentiate between valid statements which concerned observed performance and unsupported generalization which described broad attitudes, qualifications, personal traits, or other factors not directly related to performance. Such unsupported statements were assigned no value. An example of a valid statement is as follows: On a recent combat mission, Captain Smith displayed ingenuity when he quickly chose a suitable alternate route of flight to his objective after the planned route was made impassable by enemy ground fire.' On the other hand, statements such as that which follows were considered to be unsupported and too general to be of value: Captain Green is an exceptional officer.

The assumption was made that all valid statements described performance specimens, that is, incidents of relevant performance which were uncommonly effective or uncommonly ineffective. Statements deemed valid were not necessarily complete sentences. In some cases one

TABLE II

IDENTIFICATION, NUMBER OF SPECIMENS, AND E/I RATIOS OF PERSONAL QUALITIES RATINGS EMPLOYED IN PHASE II

| Name | Narrative E/I Ratio | Category A E/I Ratio | | | Category B E/I Ratio | | | Category C E/I Ratio | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Lt. Green | 8 | 12 | 3 | 4 | 16 | 2 | 8 | 16 | 1 | 16 |
| Lt. Gray | 4 | 8 | 4 | 2 | 12 | 3 | 4 | 16 | 2 | 8 |
| Lt. Black | 2 | 6 | 6 | 1 | 10 | 5 | 2 | 12 | 3 | 4 |
| Lt. White | 1.5 | 6 | 8 | 0.75 | 6 | 4 | 1.5 | 9 | 3 | 3 |
| Lt. Brown | 1 | 4 | 8 | 0.5 | 7 | 7 | 1 | 10 | 5 | 2 |
| Lt. Blue | 0.5 | 3 | 12 | 0.25 | 4 | 8 | 0.5 | 7 | 7 | 1 |

sentence described two or more performance specimens, while in others, two or more sentences related only one performance specimen.

After the effective/ineffective scores were determined for all of the forms, the data were grouped and mean values plotted on log-log paper in order to make a preliminary determination as to the applicability of the Psychophysical Law. Data obtained in Phase II of the survey were treated similarly to that received in Phase I except that determination of stimulus intensity was unnecessary.

Data received in Phase I were subjected to multiple regression analysis. Multiple regression analysis was also used to define the association between the narrative descriptions and the overall evaluations made in Phase II of the survey. Analysis of variance was used to determine the effect of the personal qualities ratings upon the overall demonstrated performance evaluations made in Phase II.

In summary, this chapter has served to define the source of data, describe the design and presentation of the rating forms used, and set forth the analytical methods employed. Chapter IV which follows is a presentation of the results of the experiment.

# CHAPTER IV

## RESULTS

The results of this investigation are presented in this chapter in three parts. Part I is a report of the outcome of the portion of the experiment in which raters were asked to evaluate a performance which they had actually observed. Part II is a presentation of the results of the part of the experiment in which raters were asked to respond to written performance statements concerning hypothetical officers. Part III is the results of an effort to measure the effect of personal qualities ratings on the quantitative overall demonstrated performance scores.

### I. RESPONSE TO OBSERVED PERFORMANCE

The purpose of having each rater respond to an observed performance was to test the hypothesis that a power function describes the relationship between observation of performance specimens narratively described by the rater and evaluation of the same performance as indicated by a quantitative score.

This phase of the survey yielded 101 completed forms. Figure 1 illustrates the distribution of the overall demonstrated performance ratings received. On an average, each narrative description of performance contained approximately six performance specimens. Of the 101 forms, 40 contained both effective and ineffective specimens,

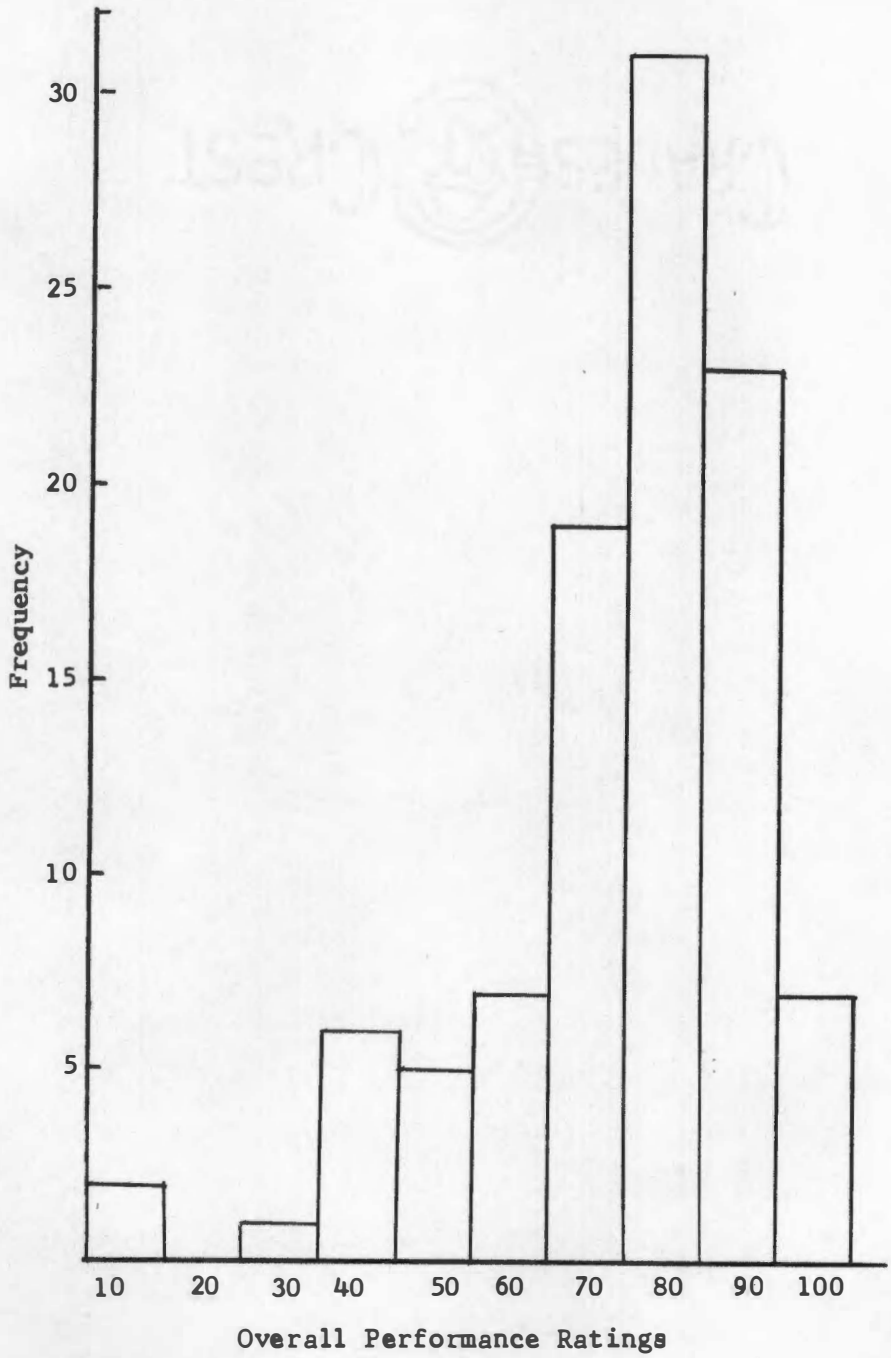


FIGURE 1

DISTRIBUTION OF 101 OVERALL DEMONSTRATED PERFORMANCE RATINGS

57 contained only effective specimens, and 4 contained only ineffective specimens. The forms containing only ineffective specimens were disregarded since they were considered too few in number to be of value in testing the first hypothesis. In testing the applicability of the Psychophysical Law, the 40 ratings which contained both effective and ineffective specimens and the 57 ratings which contained only effective specimens were treated separately.

Table III presents the distribution of the E/I values, the overall performance ratings and the mean scale values for the group of 40 forms containing both E's and I's. In order to plot mean scale values on log-log paper to get a preliminary estimate as to the applicability of the Psychophysical Law, it was deemed appropriate to first group E/I categories which contained a single rating with the next higher category. The grouped data and resulting values are also shown in Table III.

Figure 2 illustrates the distribution of the 40 overall ratings. Note that the distribution is skewed to the right and that 25 of the 40 scores are either 70's or 80's.

Figure 3 shows the log-log plots of the E/I values versus the mean performance ratings. The regression line shown is that obtained through regression analysis of the individual observations rather than mean values. A reasonably good straight line fit is apparent.

Table IV presents the results of regression analysis using various transformations of the raw data. Note that the regression

TABLE III

DISTRIBUTION OF 40 EVALUATIONS WHICH CONTAINED BOTH EFFECTIVE AND INEFFECTIVE SPECIMENS IN THE NARRATIVE DESCRIPTIONS

| Stimuli E/I Values | Overall Performance Ratings | | | | | | | | | Mean Rating | Grouped Data: Mean E/I Values | Grouped Data: Mean Rating |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | | | |
| 6.00 | | | | | | | 1 | 1 | | 75.00 | | |
| 5.00 | | | | | | 1 | 2 | 2 | | 72.00 | | |
| 4.00 | | | | | | | 3 | 7 | 1 | 78.18 | | |
| 3.00 | | | | | | 2 | 2 | | | 65.00 | | |
| 2.50 | | | | | 1 | | 2 | | | 63.33 | | |
| 2.00 | | | | | | 1 | 3 | | | 67.50 | | |
| 1.70 | | | | | | | | 1 | | 80.00 | 1.6 | 65.00 |
| 1.50 | | | | | 1 | | | | | 50.00 | | |
| 1.00 | | | | 2 | 1 | 1 | 1 | | | 52.00 | | |
| 0.75 | | | | 1 | | | | | | 40.00 | 0.58 | 40.00 |
| 0.40 | | | | 1 | | | | | | 40.00 | | |
| 0.33 | | | | | 1 | | | | | 50.00 | 0.24 | 30.00 |
| 0.14 | 1 | | | | | | | | | 10.00 | | |

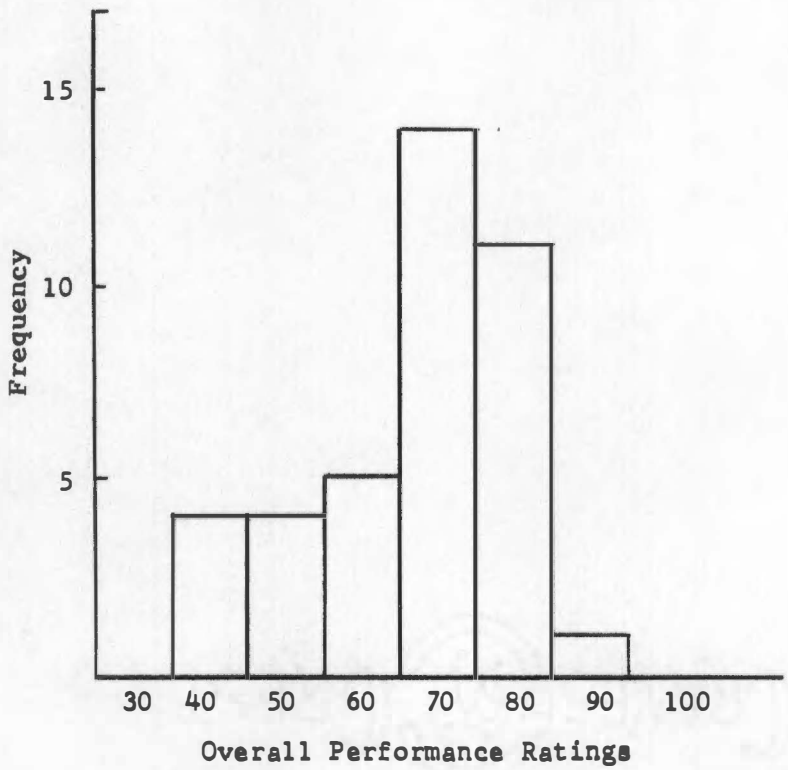


FIGURE 2

DISTRIBUTION OF 40 EVALUATIONS WHICH CONTAINED BOTH EFFECTIVE AND INEFFECTIVE SPECIMENS IN THE NARRATIVE DESCRIPTIONS

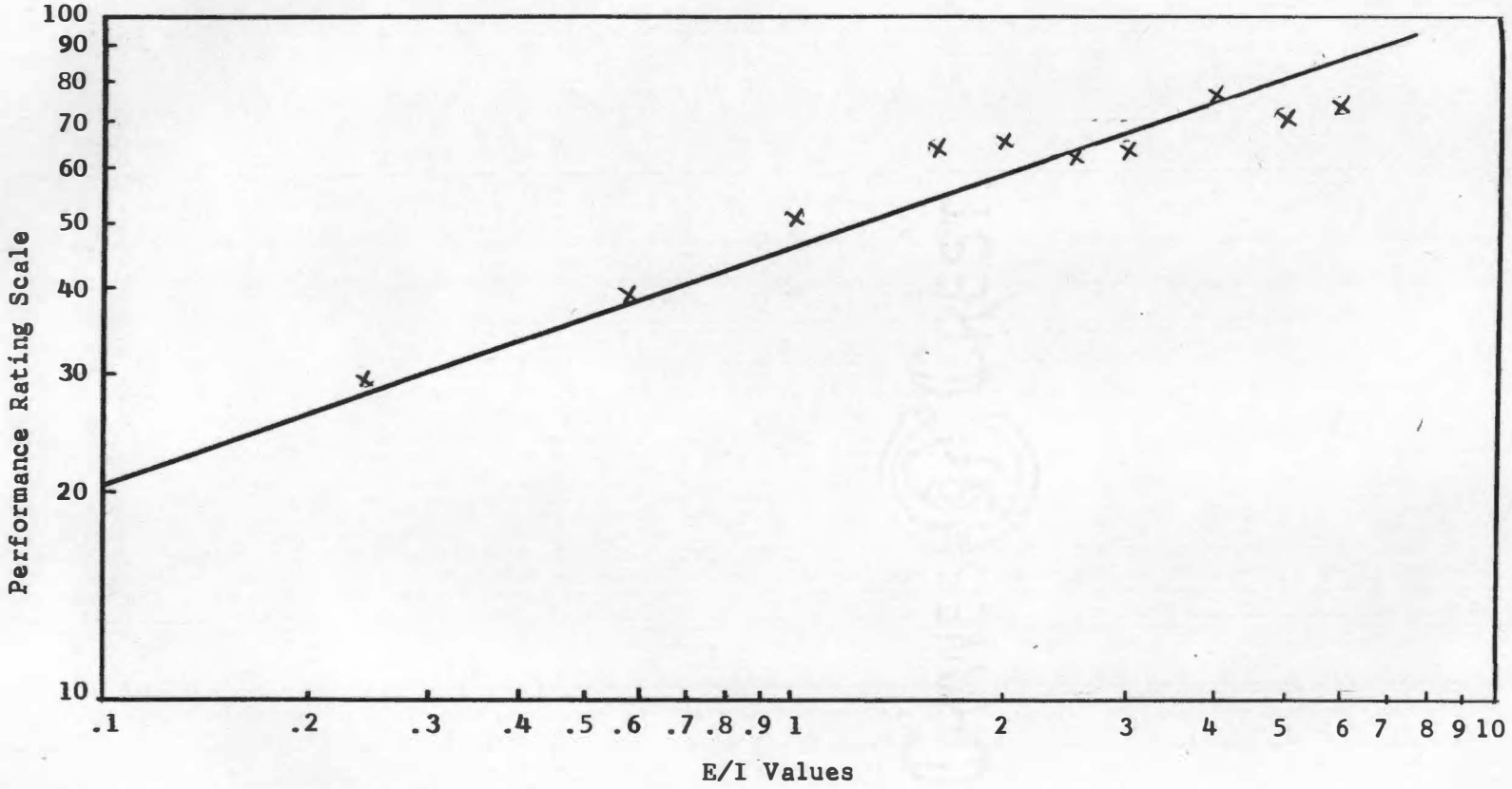


FIGURE 3

E/I RATIOS VERSUS MEAN OVERALL DEMONSTRATED PERFORMANCE RATINGS FOR 40 EVALUATIONS WHICH CONTAINED BOTH E AND I SPECIMENS

TABLE IV

REGRESSION ANALYSIS OF 40 EFFICIENCY REPORTS WHICH CONTAINED BOTH E AND I SPECIMENS IN THE NARRATIVE DESCRIPTIONS

| No. | Form | Regression Equation | Correlation Coefficient | Error of Estimate | Mean Values | Standard Deviation |
|---|---|---|---|---|---|---|
| 1. | $Y = KS_1^{b_1}$ where $S_1$ = E/I values | $Y = 46.49\ S_1^{.35}$ | .81 | 0.0960* | $S_1$ -00.3633* | 00.3688* |
| | | | | | Y 01.7954* | 00.1608* |
| 2. | $Y = KS_2^{b_2}S_3^{b_3}$ where $S_2$ and $S_3$ = number of E's and I's | $Y = 44.92\ S_2^{.37}S_3^{-.34}$ | Mult. r = .81<br>$S_2$ to Y = .62<br>$S_3$ to Y = -.73<br>$S_2$ to $S_3$ = -.40 | 0,0972* | $S_2$ 00.5686*<br>$S_3$ 00.2053*<br>Y -01.7954* | 00.1692*<br>00.2674*<br>00.1608* |
| 3. | $Y = K + b_4S_4 + b_5S_5$ where $S_4$ and $S_5$ = number of E's and I's | $Y = 66.84 + 3.14\ S_4 + (-6.87\ S_5)$ | Mult. r = .80<br>$S_4$ to Y = .43<br>$S_5$ to Y = -.76<br>$S_4$ to $S_5$ = -.25 | 9.7700 | $S_4$ 03.9500<br>$S_5$ 02.0000<br>Y 65.5000 | 01.3000<br>01.6300<br>16.0100 |

*Logarithms.

line plotted in Figure 3 represents Equation Number 1 in Table IV.

The group of 57 narrative descriptions which contained only effective specimens was treated in the same manner as just described. Table V presents the distribution of the stimuli, corresponding ratings, and mean ratings. Note that no grouping of data was necessary as in the previous group, since all stimulus categories contained two or more ratings.

Figure 4 illustrates the distribution of the performance ratings. As was the group of ratings which contained both E's and I's, this distribution is also skewed to the right and a great percentage of the ratings fall in two categories. Specifically, 42 of 57 ratings are either 80's or 90's.

Figure 5 shows the log-log plots of the number of effective specimens against the corresponding mean overall performance ratings. The regression line shown is that obtained through regression analysis of the individual observations rather than mean values. Again, a reasonably good straight line was obtained.

Table VI presents the results of regression analysis using both logarithms and nontransformed values. Note that the regression line plotted in Figure 5 represents Equation Number 1 in Table VI.

## II. RESPONSE TO WRITTEN PERFORMANCE DESCRIPTIONS

The purpose of soliciting responses to written descriptions of performance was to test the hypothesis that a power function describes

TABLE V

DISTRIBUTION OF 57 EVALUATIONS WHICH CONTAINED ONLY EFFECTIVE SPECIMENS IN THE NARRATIVE DESCRIPTIONS

| Stimuli Number of E's | Overall Performance Ratings | | | | | | Mean Ratings |
|---|---|---|---|---|---|---|---|
| | 50 | 60 | 70 | 80 | 90 | 100 | |
| 8 | | | | 3 | 2 | 2 | 87.57 |
| 7 | | | | 1 | 7 | 1 | 90.00 |
| 6 | | 1 | | 3 | 8 | 1 | 86.15 |
| 5 | | | 3 | 7 | 2 | 2 | 82.14 |
| 4 | 1 | 1 | 1 | 5 | 3 | 1 | 79.17 |
| 3 | | | 1 | 1 | | | 75.00 |

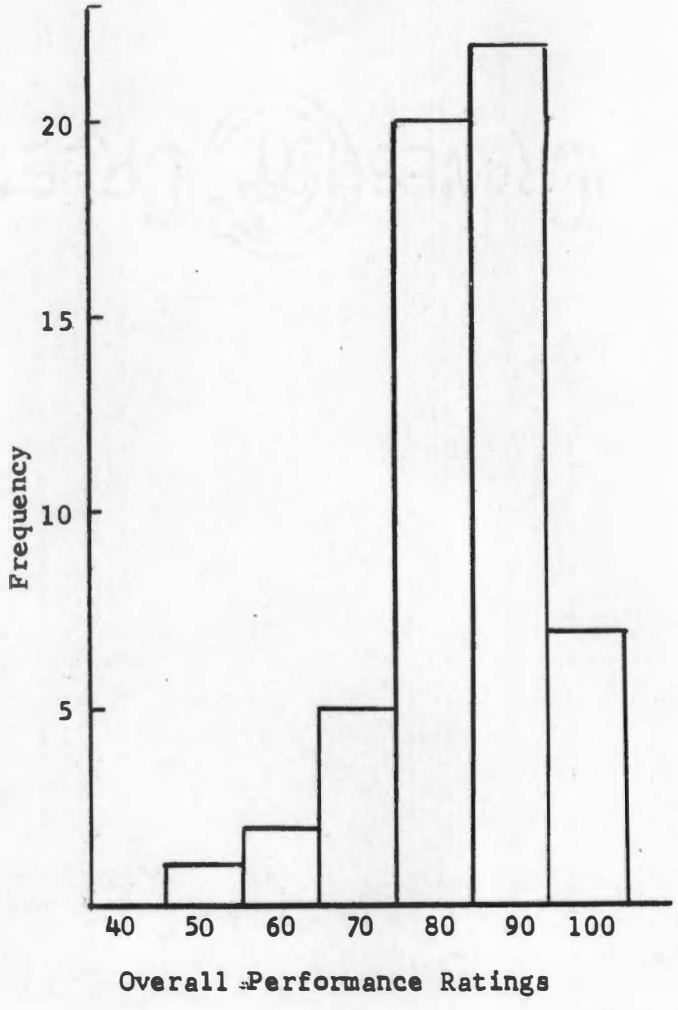


FIGURE 4

DISTRIBUTION OF 57 EVALUATIONS WHICH CONTAINED ONLY EFFECTIVE SPECIMENS IN THE NARRATIVE DESCRIPTIONS

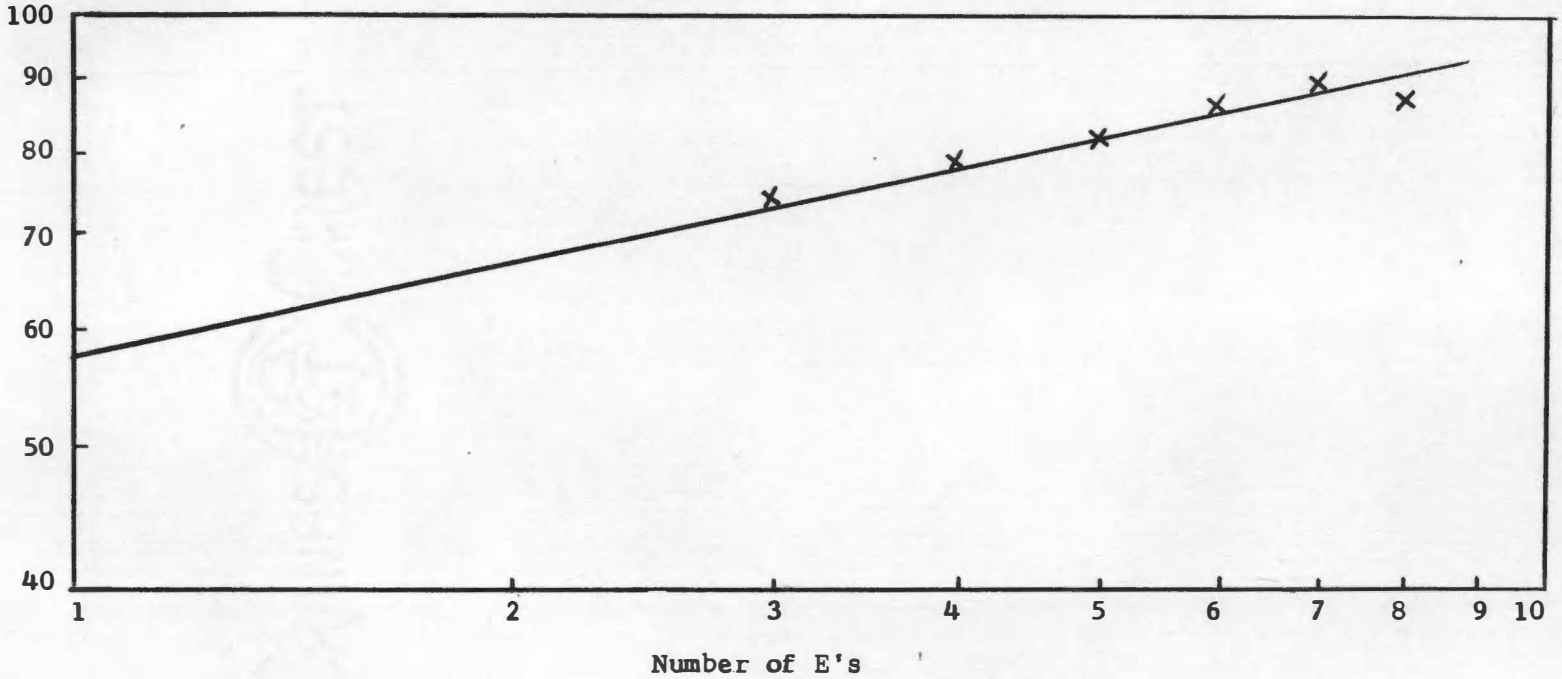


FIGURE 5

NUMBER OF EFFECTIVE PERFORMANCE SPECIMENS VERSUS MEAN OVERALL DEMONSTRATED PERFORMANCE RATINGS FOR 57 EVALUATIONS WHICH CONTAINED ONLY EFFECTIVE SPECIMENS

TABLE VI

REGRESSION ANALYSIS OF 57 EFFICIENCY REPORTS WHICH CONTAINED ONLY EFFECTIVE SPECIMENS IN THE NARRATIVE DESCRIPTIONS

| No. | Form | Regression Equation | Correlation Coefficient | Error of Estimate | Mean Values | Standard Deviation |
|---|---|---|---|---|---|---|
| 1. | $Y = KS_6^{b_6}$ where $S_6$ = number of E's | $Y = 58.86\ S_6^{.21}$ | .39 | 0.0546* | $S_6$ 00.7370* | 00.1113* |
| | | | | | Y 01.9217* | 00.0588* |
| 2. | $Y = K + b_7S_7$ where $S_7$ = number of E's | $Y = 67.80 + 2.91\ S_7$ | .39 | 9.78 | $S_7$ 05.63 | 01.40 |
| | | | | | Y 84.21 | 10.51 |

*Logarithms.

the relationship between observation and evaluation of performance when the evaluation is a response to a written set of performance specimens. The forms presented to the raters in this portion of the experiment also contained personal qualities ratings of varied magnitudes. As will be explained fully later, the personal qualities ratings had no significant effect on the overall performance scores. Therefore, the performance scores were considered to be solely a function of the sets of performance specimens.

Ninety-five raters completed 6 forms each for a grand total of 570 ratings. Table VII shows the distribution of the stimuli, the corresponding ratings, and the mean ratings.

Figure 6 illustrates the distribution of the overall performance ratings. Note that while this distribution is also slightly skewed to the right, it more closely approximates a normal distribution than does the distribution of the 101 ratings of actual performance shown in Figure 1.

Figure 7 shows the log-log plots of the E/I values versus the performance ratings. Each point plotted represents the mean of 95 observations. The regression line was obtained by analysis of individual observations. Equation Number 1 in Table VIII represents the regression line shown in Figure 7.

Table VIII presents the results of multiple regression analysis using the logarithms of all the variables and the results obtained when the nontransformed values of the variables were used.

TABLE VII

DISTRIBUTION OF 570 OVERALL DEMONSTRATED PERFORMANCE RATINGS OF 6 HYPOTHETICAL ARMY OFFICERS

| Stimuli E/I Values | Overall Performance Ratings | | | | | | | | | | Mean Ratings |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 | |
| 8 | | | | | 2 | 2 | 27 | 45 | 17 | 2 | 78.32 |
| 4 | | | | | 3 | 11 | 36 | 32 | 12 | 1 | 75.47 |
| 2 | | | | 2 | 10 | 31 | 36 | 15 | 1 | | 65.79 |
| 1.5 | | | | 12 | 29 | 38 | 13 | 3 | | | 56.42 |
| 1 | | | 2 | 11 | 30 | 39 | 11 | 2 | | | 55.26 |
| 0.5 | 1 | 8 | 14 | 35 | 28 | 7 | 2 | | | | 41.58 |

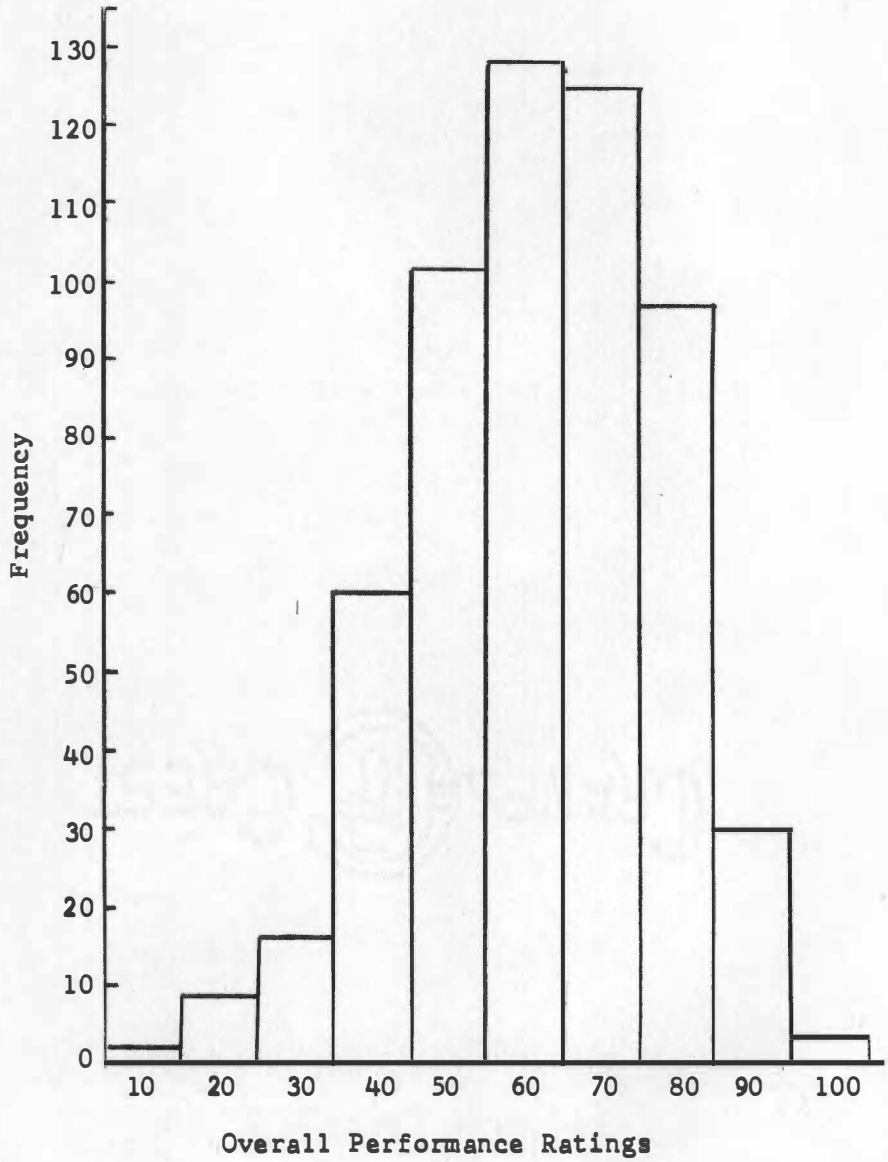


FIGURE 6

DISTRIBUTION OF 570 OVERALL DEMONSTRATED PERFORMANCE RATINGS OF SIX HYPOTHETICAL ARMY OFFICERS

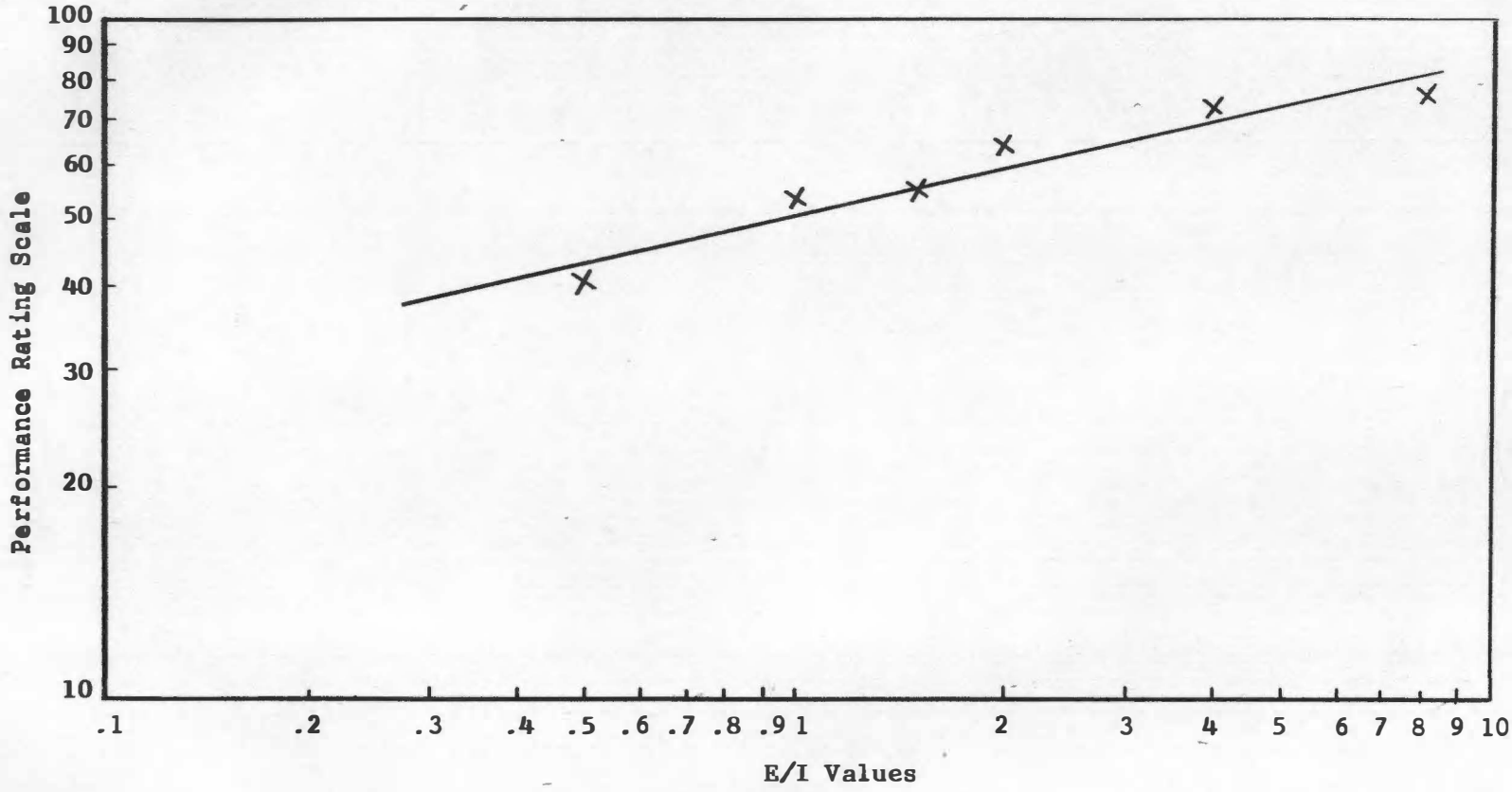


FIGURE 7

E/I RATIOS VERSUS MEAN OVERALL DEMONSTRATED PERFORMANCE RATINGS FOR SIX HYPOTHETICAL ARMY OFFICERS

TABLE VIII

REGRESSION ANALYSIS OF 570 EFFICIENCY REPORTS ON 6 HYPOTHETICAL OFFICERS

| No. | Form | Regression Equation | Correlation Coefficient | Error of Estimate | Mean Values | | Standard Deviation |
|---|---|---|---|---|---|---|---|
| 1. | $Y = KS_8^{b_8}$ where $S_8$ = narrative E/I values | $Y = 51.02\ S^{.23}$ | .73 | 00.0897* | $S_8$ | 00.2802* | 00.3918* |
| | | | | | Y | 01.7748* | 00.1296* |
| 2. | $Y = KS_9^{b_9}S_{10}^{b_{10}}$ where $S_9$ and $S_{10}$ = number of narrative E's and I's | $Y = 28.05\ S_9^{.50}S_{10}^{-.10}$ | Mult. r = .75<br>$S_9$ to Y = .73<br>$S_{10}$ to Y = -.67 | 00.0872* | $S_9$ | 00.7564* | 00.1446* |
| | | | | | $S_{10}$ | 00.4762* | 00.2629* |
| | | | | | Y | 01.7748* | 00.1296* |
| 3. | $Y = K + b_{11}S_{11} + b_{12}S_{12}$ where $S_{11}$ and $S_{12}$ = number of narrative E's and I's | $Y = 63.67 + 2.48\ S_{11} + (-4.75\ S_{12})$ | Mult. r = .77<br>$S_{11}$ to Y = .76<br>$S_{12}$ to Y = -.77 | 10.1000 | $S_{11}$ | 05.1000 | 01.7300 |
| | | | | | $S_{12}$ | 03.5000 | 01.7000 |
| | | | | | Y | 61.9300 | 15.9500 |

*Logarithms.

## III. EFFECT OF PERSONAL QUALITIES RATINGS ON THE OVERALL DEMONSTRATED PERFORMANCE RATINGS

It was hypothesized that the degree to which the ratee was deemed by the rater to possess or lack specified personal qualities has a significant effect on the overall demonstrated performance score. To test this hypothesis 95 raters were divided at random into three nearly equal groups. All raters were presented six partially completed Efficiency Reports. The narrative portions of the forms presented to the raters were the same regardless of group. However, the personal qualities ratings were systematically varied among the three groups as described in Chapter III.

The mean of the total of six overall performance ratings for the first group was 363.03, for the second group 374.67, and for the third group 377.19. Mean individual ratings were 60.51, 62.45, and 62.86 respectively. Analysis of variance, specifically an F test, was used to determine whether or not the differences in the group means were significant at the .05 confidence level. Computation yielded a F ratio of 1.16 which is not significant at the stated level, thus indicating that the three groups of raters did not consider the presented personal qualities ratings to an appreciable extent in arriving at overall demonstrated performance scores.

In an attempt to gain further insight concerning the relationship between personal qualities ratings, manner of performance descriptions, and overall performance scores, data received in Phase I of the

survey were subjected to additional regression analysis, the results of which are presented in Table IX.

Equations 5 and 6 in Table IX were selected for still further examination since they represented multiple regressions of both manner of performance values and personal qualities ratings on the overall demonstrated performance scores. Note that the multiple correlation coefficient in Equation 5 is .81 and in Equation 6 is .66 which indicates that 66 and 47 per cent of the total variance in the overall demonstrated performance scores of the two groups was due to the manner of performance values and the personal qualities scores. The question was, "How much did each independent variable contribute to the explained variation in the dependent variable?"

Results of computations designed to answer the above question were most interesting. With regard to Equation 5, it was found that manner of performance values contributed 31 per cent of the total variance, personal qualities scores accounted for 17 per cent, and the two combined explained the remaining 18 per cent. In Equation 6, manner of performance values contributed only 6 per cent, personal qualities explained 35 per cent and the combination of the two contributed the remaining 6 per cent. Although these results must be viewed with caution since the samples on which they are based consisted of only 30 and 46 observations respectively, it appears that the raters, when asked to respond to observed performance, gave a great deal of consideration to the personality traits of the ratees.

TABLE IX

REGRESSION ANALYSIS OF PERSONAL QUALITIES RATINGS

| No. | Form | Regression Equation | Correlation Coefficient | Error of Estimate | Mean Values | Standard Deviation | Number of Observations |
|---|---|---|---|---|---|---|---|
| 1. | $Y = KP_1^{b_1}$ where $P_1$ = E/I values | $Y = 60.09\ P_1^{.12}$ | .66 | 00.0803* | $P_1$ 00.2974*<br>Y 01.8155* | 00.5580*<br>00.1050* | 39 |
| 2. | $Y = K + b_2P_2 + b_3P_3$ where $P_2$ and $P_3$ = number of E's and I's | $Y = 68.78 + 2.88\ P_2 + (-5.45\ P_3)$ | Mult. r = .70<br>$P_2$ to Y = .39<br>$P_2$ to Y = -.65<br>$P_2$ to $P_3$ = -.21 | 11.0900 | $P_2$ 04.1500<br>$P_3$ 01.6200<br>Y 67.1800 | 01.3900<br>01.6300<br>15.0300 | 39 |
| 3. | $Y = KP_4^{b_4}$ where $P_4$ = number of E's when I = 0 | $Y = 44.99\ P_4^{.24}$ | .68 | 00.0336* | $P_4$ 01.1146*<br>Y 01.9248* | 00.1292*<br>00 0453* | 55 |
| 4. | $Y = K + b_5P_5$ where $P_5$ = number of E's when I = 0 | $Y = 68.78 + 2.83\ P_5$ | .45 | 07.9100 | $P_5$ 05.5600<br>Y 84.5500 | 01.4000<br>08.7800 | 55 |
| 5. | $Y = KS_6^{b_6}P_7^{b_7}$ where $S_6$ and $P_7$ = E/I narrative and E/I Personal Qualities | $Y = 53.11\ S_6^{.19}P_7^{.08}$ | Mult. r = .81<br>$S_6$ to Y = .72<br>$P_7$ to Y = .63<br>$S_6$ to $P_7$ = .39 | 00.0617* | $S_6$ 00.3339*<br>$P_7$ 00.1806*<br>Y 01.7954* | 00.3340*<br>00.5440*<br>00.1015* | 30 |
| 6. | $Y = KS_8^{b_8}P_9^{b_9}$ where $S_8$ = number of E's narrative when I = 0; $P_9$ = number of E's Personal Qualities when I = 0 | $Y = 34.79\ S_8^{.10}P_9^{.27}$ | Mult. r = .68<br>$S_8$ to Y = .36<br>$P_9$ to Y = .64<br>$S_8$ to $P_9$ = .20 | 00.03250* | $S_8$ 00.7516*<br>$P_9$ 01.1785*<br>Y 01.9328* | 00.1051*<br>00.0969*<br>00.0436* | 46 |
| 7. | $Y = K + b_9S_9 + b_{10}S_{10} + b_{10}P_{10} + b_{11}P_{11}$ where $S_9$ and $S_{10}$ = number of E's and I's narrative; $P_9$ and $P_{10}$ = number of E's and I's, Personal Qualities | $Y = 62.56 + 1.58\ S_9 + (-2.01\ S_{10}) + .97\ P_{10} + (-1.55\ P_{11})$ | Mult. r = .89<br>$S_9$ to Y = .69<br>$S_{10}$ to Y = -.79<br>$P_{10}$ to Y = .76<br>$P_{11}$ to Y = -.84<br>$S_9$ to $S_{10}$ = -.60<br>$S_9$ to $P_{10}$ = .56<br>$S_9$ to $P_{11}$ = -.66<br>$S_{10}$ to $P_{10}$ = -.62<br>$S_{10}$ to $P_{11}$ = .86<br>$P_{10}$ to $P_{11}$ = -.68 | 08.3200 | $S_9$ 04.7400<br>$S_{10}$ 01.0100<br>$P_{10}$ 10.8700<br>$P_{11}$ 02.5000<br>Y 74.6500 | 01.8300<br>01.7000<br>05.7700<br>04.3500<br>18.2500 | 101 |

*Logarithms.

This chapter has been devoted to a presentation of the results of the analysis of the data. Chapter V which follows is a discussion of these results.

# CHAPTER V

## DISCUSSION OF RESULTS

Results of the experiment are discussed in this chapter as they relate to each of the three original hypotheses. First, however, brief comment concerning the source of the data on which the results are based is warranted.

Few topics are of more interest to Army officers than efficiency ratings. Throughout the period in which the data were collected, it was clearly evident that the student raters as a group were most sincere and conscientious in fulfilling their part in the experiment. Examination of the data later proved this to be true. Also contributing to the quality of the ratings was the experience that most of the students had as raters and their familiarity with the rating system. Therefore, it is believed that the evaluations received closely approximate authentic Army Officer Efficiency Reports.

### I. DISCUSSION OF THE RESPONSE TO OBSERVED PERFORMANCE

The primary hypothesis tested using the data received in Phase I of the experiment was that a power function describes the relationship between narrative description of observed performance and quantitative evaluation of the same performance.

Of the 101 evaluations received in Phase I, more than half contained no ineffective specimens in the narrative descriptions, while

only 4 contained no effective specimens. These results are not at all surprising since they reveal the well known Army rater "leniency" tendency which has plagued Army evaluation systems over the years.

The 40 evaluations which contained both effective and ineffective specimens were considered first. These data, tabulated in Table III and plotted in Figure 3, yielded the regression equation $Y = 46.49\ S^{.35}$. Of the transformations of the raw data tried, none produced a better fit to a straight line than did the E/I values versus the performance scores plotted on log-log paper.

The correlation coefficient of .81 for these data is higher than the correlation coefficients yielded by previous studies of this nature. The exponent of the best-fit equation, that is, the slope of the regression line is .35 which is similar to findings in past inquiries in which the E/I values were regarded as the independent variable.[1]

For data which conform to the Psychophysical Law, an E/I value of 1 results in an overall categorical evaluation which falls near the midpoint of the rating scale. From Figure 3 it is seen that an E/I value of 1 for these data resulted in an overall performance score of 46.49 which is reasonably close to the rating scale midpoint of 55. Possible explanation for the deviation of this score from the midpoint is presented later in this chapter.

---

[1]Gerald H. Whitlock, "Application of the Psychophysical Law to Performance Evaluation," Journal of Applied Psychology, 47:15-23, February, 1963.

One final comment concerning these results is warranted. Note from Table IV that the correlation coefficients and errors of estimates in Equation 1 and Equation 2 are practically identical. In Equation 1, E/I values were used as the stimulus; in Equation 2 a multiple regression analysis was made using the logarithms of the number of E's and the number of I's. These results seem to substantiate Whitlock's theory that the proper way to combine effective and ineffective specimens is to use the logarithm of their quotient.[2]

The 57 forms which contained only effective specimens in the narrative descriptions were analyzed in a manner similar to that described above except that the number of effective specimens was used as the independent variable. The regression equation for the straight line which fits these data best is $Y = 58.86\ S^{.21}$. Although the correlation coefficient of .39 is considerably smaller than that obtained for the data which contained both E's and I's, the plot of the means as shown in Figure 5 reveals that the nature of the relationship is exponential.

One or all of several factors may have contributed to the relatively low correlation between the variables of this sample. It may have been simply that the raters did not have enough time or enough space on the rating form to describe all of the observed performance specimens which served as stimuli for the overall quantitative evaluation. On the other hand, the true explanation may be more complex. Army officers are expected to perform effectively. All training is

---

[2] Ibid.

aimed at producing more effective performance. It may be that Army raters have been exposed to effective behavior to such an extent that they have become relatively insensitive to the number of effective performance specimens when the performance consists of effective specimens alone. Whatever the reason, a related investigation by H. K. Stevenson produced similar results. Raters participating in the Stevenson experiment also lacked the ability to discriminate with any consistency among highly effective performances.[3]

If the overall performance scores were not primarily a function of observed performance, what other factors influence the evaluation of performances which contain no ineffective specimens? Results of an analysis of performance specimens and personality traits combined suggest an answer to this question and are presented in Part III of this chapter.

Another puzzling situation arose during the analysis of these data. Although a power function described the relationship very well, the fit was not decidedly better than the fit to a linear function. Graphic analysis explained this perplexity. When the exponential curve was plotted on ordinary graph paper, it was readily apparent that the farther the increments of the curve lay from the origin, the smaller the degree of curvature they possessed. The data in question fell on a portion of the regression line which was almost straight due to the distance of the segment from the origin. In addition, the range of the

---

[3]Harry K. Stevenson, "Factors Influencing the Transition from Observation of Performance to Evaluation of Performance." (Unpublished Master's thesis, College of Business Administration, The University of Tennessee, Knoxville, 1963.

values of the independent variable was small, thus making analysis even more difficult.

## II. DISCUSSION OF RESPONSE TO WRITTEN PERFORMANCE DESCRIPTIONS

It was hypothesized that a power function also describes the relationship between observation and evaluation of performance when the evaluation is a response to a set of written performance specimens.

From Figure 6 it is seen that the distribution of the performance ratings of hypothetical performances more nearly approximates a normal distribution than does either of the distributions of ratings of observed performance shown in Figures 2 and 4. Findings of a previous study concerning an Army rating system suggest an explanation for the difference. The study revealed that Army raters were more lenient when ratings of actual subordinates were "for keeps" than when the rating had only administrative implications.[4] In the present study the ratings based on written statements concerned hypothetical officers and had purely experimental implications, while the ratings of observed performance were more nearly analogous to "for keeps" evaluations.

Regression analysis of the E/I values versus the performance ratings yielded the equation $Y = 51.02\ S^{.23}$. The plot of the mean ratings against the E/I values as illustrated in Figure 7 reveals that

---

[4]Erwin K. Taylor and Robert J. Wherry, "A Study of Leniency in Two Systems," Personnel Psychology, 4:39-47, Spring, 1951.

the two are clearly related by the Psychophysical Law. It is important to note that the sample in this portion of the experiment consisted of 570 individual observations. Each of the points plotted in Figure 7 represents the mean of 95 performance scores.

The correlation coefficient for the data is .73 and the exponent of the best-fit regression line is .23. Both results are very similar to findings of previous studies in which the Psychophysical Law obtained.[5]

Further indication that the Psychophysical Law properly describes the relationship in this portion of the experiment is that the E/I value of 1 resulted in the near average performance score of 51.02. It is interesting to note that both this value and the corresponding value of 46.49 for the group of 40 responses to observed performance discussed earlier fall slightly below the scale midpoint.

The fact that an E/I value of 1 resulted in a slightly below average rating in both of the samples mentioned above might imply that Army raters are considerably more sensitive to ineffective performance specimens than they are to effective ones. This is not at all unlikely. As was mentioned previously, effective performance is more common than ineffective, therefore causing ineffective specimens to be more noticeable. Too, ineffective performance is perhaps more intolerable in the Army than in an organization with a less crucial purpose. Lastly,

---

[5]Whitlock, loc. cit.

it may be that, since raters are responsible to some extent for eliminating ineffective performance on the part of their subordinates, an ineffective specimen makes a deeper impression on the rater because it necessitates some form of corrective action on his part.

If the true explanation of the below average mean scores for E/I values of 1 for the data analyzed in study does not lie in the ultrasensitivity of Army raters to ineffective performance, another possibility exists. Perhaps, Army raters are simply intolerant of variability in performance. In other words, it may be that Army raters consider a performance which contains a few specimens of effective performance more desirable than a performance which contains a large number of effective specimens but also contains a few ineffective specimens.

With regard to the sample of 57 evaluations which contained only effective specimens, recall that 1 effective specimen resulted in a mean performance score of 58.86. A comparison of this figure with the mean performance scores of 46.49 and 51.02 for E/I values of 1 has implications concerning the fundamental methods of psychophysics used in experiments of this nature. The question has arisen as to whether or not an E/I value of 1 is equal as a stimulus to 1 effective specimen where the performance contains no ineffective specimens. The findings of this study suggest that the two are not equal.

Comparison of the correlation coefficients and the errors of estimates of Equation 1 and 2 of Table VIII indicates again that the proper way to combine effective and ineffective specimens is to use the

logarithm of their quotient. Such combination is most useful since it allows the use of simple regression procedures and graphic illustration.

Another observation concerning this portion of the experiment as well as the previous portions is in order. From Figures 3, 5, and 7 it is seen that the plot corresponding to highest stimulus value falls in each case below the line of best-fit. The explanation for this perhaps lies in the design of the rating system. The highest two and lowest two rating blocks on the rating form are designated with asterisks. The asterisks are to remind the rater that the governing Army regulation requires specific and extensive justification for ratings of 10, 20, 90, or 100.

It seems that most of the raters were reluctant to assign ratings higher than 80 regardless of the stimulus intensity. Since the raters were generally lenient, the special area at the lower end of the rating scale was of no consequence. The foregoing suggests that the data might have fit a straight line even better had the system employed a continuous scale instead of one on which special areas were designated.

The findings of this inquiry are remarkably similar to previous investigations of the psychophysics of performance evaluation. However, the previous studies by Whitlock and students under his direction and the present effort are dissimilar in one major aspect. Whereas Whitlock's work involved the checklist type rating form, this study has concerned the free-written or narrative rating system. The free-written

rating is widely criticized in the literature as being unrefined and undisciplined. Results of this study suggest a general methodology for quantitative analysis of narrative ratings where such a rating system is employed out of necessity.

## III. DISCUSSION OF THE EFFECT OF PERSONAL QUALITIES

Although the Army regulation which governs the Efficiency Report system prescribes that an overall demonstrated performance score be made strictly on the basis of performance, it was hypothesized that the degree to which the ratee is deemed by the rater to possess or lack certain personal qualities affects the rater's evaluation of performance.

Analysis of data collected in Phase II of the experiment revealed that the performance scores had not been influenced by the written personal qualities scores. Although the mean performance ratings for the three groups of raters which received different personal qualities scores did fall in the same order as the intensities of the personal qualities scores, the difference in the means was not significant at the .05 level of confidence. Thus, it appears that the raters adhered very closely to instructions when evaluating hypothetical officers.

Although the raters were not influenced by written personal qualities scores in Phase II of the experiment, such was not the case in Phase I. Analysis of two samples of responses to observed performance and personal qualities disclosed that the raters in these cases

had given a great deal of consideration to personality traits in arriving at performance scores.

These findings appear at first to conflict. However, this is not necessarily true. As was pointed out before, Army raters tend to be more objective in rating purely hypothetical officers than actual subordinates. The normal distribution of Phase II performance scores as compared to the severely skewed distributions of Phase I performance scores substantiates this to a degree. Perhaps also in separating performance and personality, the raters evaluated hypothetical officers much more objectively than they did actual ones. Since Army raters are normally sincere and conscientious in their efforts to evaluate subordinates, the above results seem to indicate that the amalgamation of performance and personality was more subconscious than deliberate.

Also, the degree to which the raters confused personality with performance in Phase I of the experiment is of consequence. Keeping in mind the relatively small size of the samples involved, it should be noted that the performance scores of the evaluations which contained some ineffective specimens were not affected by personal qualities nearly so much as the performance scores which corresponded to highly effective performance. Recall in connection with the former group that 17 per cent of the total explained variance of 66 per cent was due to personality scores while for the latter sample 35 per cent of the 47 per cent of explained variance was due to personality.

The reason for the difference in the degree to which personality affected the two groups is not clear. Perhaps, as was suggested previously, Army raters are insensitive to effective performance and as a result subconsciously select the ratee's personality as the factor on which to base the performance scores. Rather than highly effective performance causing contamination of the performance score, it may be that the rater only perceives performance to be highly effective as result of the ratee's exhibition of desirable personality traits. In other words, strong personality traits may have the effect of depressing the importance of performance.

Regardless of whether the confusion of personality and performance is explainable or unexplainable and regardless of whether it is done consciously or subconsciously, the results of this phase of the experiment have serious implications. In the Army system of evaluation, Efficiency Reports are viewed at various stages by individuals who do not have personal knowledge of the ratee. These individuals are in much the same position as the raters who evaluated hypothetical employees in this experiment. This suggests that users of Efficiency Reports are also likely to view written evaluations quite objectively, separating performance and personality completely.

The "indorser", who in the Army system is the rater's immediate superior, is usually personally acquainted with the ratee. However, the relationship is normally much less intimate than the ratee-rater relationship. This would seem to indicate that the indorser views the

rater's evaluation of the ratee more objectively than does the rater.

The relationship of the ratee with the "reviewer", who is the indorser's immediate superior, and with the personnel managers at Department of the Army level is as a rule impersonal. It is obvious that if these individuals view written performance evaluations objectively and with the assumption that the rater has not been influenced by the subordinates personality serious mistakes can result. As an example of the sort of error that could be made, it is conceivable that a Department of the Army personnel officer might select an officer for an important task largely on the basis of his overall performance scores, only to discover later through experience that the officer's performance scores were not an indication of his effectiveness but a result of his pleasing personality.

The apparent differences between evaluations made on a personal basis and evaluations made on the basis of impersonal, written descriptions also has implications with regard to experimental methodology used in studies of this nature. These findings suggest serious limitations with regard to the use of impersonal evaluations in investigating the actual performance evaluation process.

Though, perhaps, all the stimuli which are involved in the complex performance evaluation process can never be isolated and precisely weighted, it is nonetheless evident that such psychological processes do obey certain rules. This study has presented evidence that there is a definite relationship between observation and evaluation of performance

as accomplished under the Army rating system. Also of significance is the finding that performance scores are not solely a function of performance but are influenced by ratee personal qualities.

Probably the most important aspect of this study is that the results further substantiate the theory that quality of performance is a prothetic continuum for which equal ratios in stimulus magnitudes produce equal ratios in response magnitudes.

This chapter has been devoted to a discussion of the results of the experiment. Chapter VI which follows is a summarization of the study and a presentation of conclusions.

# CHAPTER VI

## SUMMARY AND CONCLUSIONS

The matter investigated during the course of this study was the applicability of the Psychophysical Law to the performance evaluation as accomplished under the United States Army Officer Efficiency Report system. More specifically, the primary purpose of the inquiry was to determine if a power function describes the relation between narratively described performance specimens and overall quantitative scores. A secondary purpose was to resolve whether or not ratee personal qualities influence the rater in his overall quantitative evaluation of the subordinate.

To obtain data with which to experiment, evaluations of actual ratees were solicited from 101 experienced Army officers. In addition the officers were asked to score six hypothetical officers on the basis of written descriptions of performance and personal qualities scores.

An analysis of the data revealed that the narrative descriptions of actual performance which contained both effective and ineffective performance specimens were related to the overall performance score by the power function, $Y = 46.49\ S^{.35}$. The relation between the quantitative performance scores and the narrative descriptions which contained only effective specimens was best described by the equation $Y = 58.86\ S^{.21}$. Correlation coefficients for the two groups were 0.81 and 0.39 respectively; the errors of the estimates were 0.0960 and 0.0546, expressed in log units.

For the hypothetical officers, the overall evaluations and the presented performance descriptions were related by the power function $Y = 51.02\ S^{.23}$. The correlation coefficient for these data was 0.73 and the error of estimate 0.0897 in log units. It was discerned that for this group the presented personal qualities score had no effect on the overall performance scores. Such was not the case when the raters evaluated actual subordinates. Here personal qualities had a decided effect on the overall performance scores.

The first two of the original three hypotheses were accepted, and the conclusions are: (1) that a power function describes the relationship between specimens of performance as narratively described and the overall demonstrated performance score, and (2) that this function obtains regardless of whether the overall quantitative evaluation is a response to observation of actual performance or a response to a written set of specimens which describe a hypothetical performance.

The third hypothesis was accepted with reservations. It was concluded that the degree to which an actual ratee is deemed by the rater to possess or lack specified personal qualities does have a significant effect on the quantitative overall demonstrated performance score. The effect with regard to hypothetical ratees was negligible. The difference between the evaluations made on a personal basis and those made on an impersonal basis appears to have serious implications with regard to the use of Efficiency Reports as a basis for personnel action. In addition, the difference in the two types of evaluations

suggests limitations with regard to experimental methodology used in investigations of the performance evaluation process.

In light of the findings of this investigation and previous work by Whitlock, it appears feasible that the complex psychophysics involved in the seemingly subjective performance evaluation process can be analyzed effectively through use of common, well developed quantitative techniques.

# BIBLIOGRAPHY

## BIBLIOGRAPHY


Chisholm, Jack F. "The Function of Job Specimens in the Evaluation of Job Worth." Unpublished Master's thesis, College of Business Administration, The University of Tennessee, Knoxville, 1963.

Guilford, J. P. Psychometric Methods. New York: McGraw-Hill Book Company, Inc., 1954.

Headquarters, Department of the Army. Personnel Efficiency Ratings, Officer Efficiency Reports. Army Regulation 623-105. United States Government Printing Office.

Headquarters, Department of the Army. The New Officer Efficiency Report System. Officer's Call, Department of the Army Pamphlet 355-25. United States Government Printing Office, 1961.

Sisson, E. Donald. "Forced Choice--The New Army Rating," Personnel Psychology, 1:365-381, Autumn, 1948.

Stevens, S. S., "On the Psychophysical Law," Psychological Review, 64: 153-181, May, 1957.

Stevens, S. S., "Problems and Methods of Psychophysics," Psychological Bulletin, 55:177-196, July, 1958.

Stevens, S. S. "The Psychophysics of Sensory Function," American Scientist, 48:226-253, June, 1960.

Stevenson, Harry K. "Factors Influencing the Transition from Observation of Performance to Evaluation of Performance." Unpublished Master's thesis, College of Business Administration, The University of Tennessee, Knoxville, 1963.

Storey, Lieutenant Colonel Robert C. "The New Efficiency Report," Army, 12:30-35, August, 1961.

Taylor, Erwin K. and Robert J. Wherry. "A Study of Leniency in Two Systems," Personnel Psychology, 4:39-47, Spring, 1951.

The Officers Guide. Harrisburg: The Military Press Publishing Company, 1956.

Whisler, Thomas L. and Shirley F. Harper. Performance Appraisal. New York: Holt, Rinehart and Winston, 1962.

Whitlock, Gerald H. "The Application of the Psychophysical Law to Performance Evaluation," Journal of Applied Psychology, 47:15-23, February, 1963.

Whitlock, Gerald H. "The Application of the Psychophysical Law to Performance and Morale Evaluations." Paper read to the American Psychological Association, St. Louis, Missouri, September, 1962.

# APPENDIXES

APPENDIX I

PERFORMANCE EVALUATION SURVEY-PHASE I

INSTRUCTIONS

1. Recall to mind the last officer whom you rated. If you never acted as a "rater," think of the officer (peer or subordinate) with whom you worked most closely on your last assignment.

2. Using the modified Efficiency Report form provided, rate the chosen officer by completing the following three (3) parts:

A. Part III-Manner of Performance
NOTE: Recall that this part is to be a narrative description of the ratee's performance. It should contain relevant and specific facts. Comments in this part are not to be an appraisal of "human worth" and must not be inconsistent with other ratings on the report.

B. Part IV-Personal Qualities
NOTE: Recall that this part is a subjective analysis of the degree to which the rated officer has displayed these qualities during the period covered by the report.

C. Part IV (1)-Overall Demonstrated Performance
NOTE: Recall that this numerical rating must be justified by factual incidents included in the narrative description of performance. It must be made strictly with respect to demonstrated performance. Care must be taken not to be unduly influenced by ratings given in Part III-Personal Qualities. Ratings denoted by an asterisk must be specifically justified in Part III.

## PART III - MANNER OF PERFORMANCE

## PART IV - PERSONAL QUALITIES

| LEGEND | DEGREE | INADEQUATE | MARGINAL | BELOW AVERAGE | AVERAGE | ABOVE AVERAGE | EXEMPLARY |
|---|---|---|---|---|---|---|---|
| | NUMBER | .0 | .1 | .2 | .3 | .4 | .5 |
| | RATER | | | | | | |
| | . | a. ADAPTABILITY *(Adjusts to new or changing situations & stresses; bears up under pressure)* | | | | | |
| | . | b. AMBITION *(Seeks and welcomes additional and more important responsibilities)* | | | | | |
| | . | c. APPEARANCE *(Possesses military bearing and is neat, smart, and well-groomed)* | | | | | |
| | . | d. COOPERATION *(Works in harmony with others as a team member)* | | | | | |
| | . | e. DEPENDABILITY *(Consistently accomplishes desired actions with minimum supervision)* | | | | | |
| | . | f. ENTHUSIASM *(Motivates others by his zeal)* | | | | | |
| | . | g. EXPRESSION *(Expresses himself clearly and concisely both orally and in writing)* | | | | | |
| | . | h. FORCE *(Executes actions vigorously)* | | | | | |
| | . | i. INGENUITY *(Finds solutions to problems regardless of obstacles)* | | | | | |
| | . | j. INITIATIVE *(Takes necessary and appropriate action on his own)* | | | | | |
| | . | k. INTELLIGENCE *(Acquires knowledge and grasps concepts readily)* | | | | | |
| | . | l. JUDGEMENT *(Thinks logically and makes practical decisions)* | | | | | |
| | . | m. LOYALTY *(Renders faithful and willing support to superiors and subordinates)* | | | | | |
| | . | n. MORAL COURAGE *(Intellectual honesty, willingness to stand up and be counted)* | | | | | |
| | . | o. SELF-DISCIPLINE *(Conducts himself in accordance with accepted standards)* | | | | | |
| | . | p. SELF-IMPROVEMENT *(Takes action to improve himself)* | | | | | |
| | . | q. SOCIABILITY *(Participates freely and easily in social and community activities)* | | | | | |
| | . | r. STAMINA *(Performs successfully under protracted physical and mental stress)* | | | | | |
| | . | s. TACT *(Says or does what is appropriate without giving unnecessary offense)* | | | | | |
| | . | t. UNDERSTANDING *(Appreciation of another person's viewpoint)* | | | | | |
| | | <— SCORE | | | | | |

## PART VI - OVERALL DEMONSTRATED PERFORMANCE

| RATING | EXPECTED DISTRIBUTION OF 100 OFFICERS RATED | OVERALL DEMONSTRATED PERFORMANCE (1) | |
|---|---|---|---|
| | | RATER | VALUE |
| a. OUTSTANDING | 1 | * | 100 |
| b. EXCEPTIONAL | 3 | * | 90 |
| c. SUPERIOR | 4 | | 80 |
| | 15 | | 70 |
| d. EXCELLENT | 34 | | 60 |
| | 34 | | 50 |
| e. EFFECTIVE | 15 | | 40 |
| | 4 | | 30 |
| f. MARGINAL | 3 | * | 20 |
| g. INADEQUATE | 1 | * | 10 |
| | | SCORE | |

APPENDIX II

PERFORMANCE EVALUATION SURVEY-PHASE II

INSTRUCTIONS

1. Attached are six (6) partially completed Efficiency Reports on six (6) hypothetical officers. Part III-Manner of Performance and Part IV-Personal Qualities have already been completed on each form.

2. Read carefully Part III which describes the hypothetical officer's performance. Note in Part IV each of the Personal Qualities ratings. Assume that the ratees are equal in all respects except those noted in the Efficiency Report.

3. Based on the information given, assume the role of rater and score Part VI (1)-Overall Demonstrated Performance. Recall that this evaluation should be made strictly with respect to performance.

## PART III - MANNER OF PERFORMANCE

Lt. Green has performed as rifle platoon leader during the rating period. Upon assuming command of the 2nd Platoon, which was understrength at the time, he effectively organized his personnel for the upcoming field training exercise. Lt. Green's platoon scored higher on field problems than the other platoons in the company. This officer in carrying out his additional duty as Company Training Officer has often benefited himself and the unit by his willingness to receive sound recommendations and advice from both superiors and subordinates. Lt. Green demonstrated an ability to gather facts and make good recommendations when on one occasion he was called upon for a solution to a complex administrative problem. Lt. Green has unselfishly devoted considerable personal time to special projects in order to insure their completion on schedule. One fault, however, which this officer has displayed is his tendency to pay too much attention to relatively insignificant details, while allowing the overall task to suffer. Lt. Green has often sought added responsibility as exemplified by his volunteering to serve on a newly formed educational advisory council. He has improved his own formal education by studying during off-duty hours at a local university. Lt. Green has also been diligent in improving his physical fitness.

## PART IV - PERSONAL QUALITIES

| LEGEND | DEGREE | INADEQUATE | MARGINAL | BELOW AVERAGE | AVERAGE | ABOVE AVERAGE | EXEMPLARY |
|---|---|---|---|---|---|---|---|
| | NUMBER | .0 | .1 | .2 | .3 | .4 | .5 |

| RATER | |
|---|---|
| .5 | a. ADAPTABILITY (*Adjusts to new or changing situations & stresses; bears up under pressure*) |
| .5 | b. AMBITION (*Seeks and welcomes additional and more important responsibilities*) |
| .5 | c. APPEARANCE (*Possesses military bearing and is neat, smart, and well-groomed*) |
| .5 | d. COOPERATION (*Works in harmony with others as a team member*) |
| .5 | e. DEPENDABILITY (*Consistently accomplishes desired actions with minimum supervision*) |
| .5 | f. ENTHUSIASM (*Motivates others by his zeal*) |
| .4 | g. EXPRESSION (*Expresses himself clearly and concisely both orally and in writing*) |
| .5 | h. FORCE (*Executes actions vigorously*) |
| .5 | i. INGENUITY (*Finds solutions to problems regardless of obstacles*) |
| .5 | j. INITIATIVE (*Takes necessary and appropriate action on his own*) |
| .3 | k. INTELLIGENCE (*Acquires knowledge and grasps concepts readily*) |
| .5 | l. JUDGEMENT (*Thinks logically and makes practical decisions*) |
| .5 | m. LOYALTY (*Renders faithful and willing support to superiors and subordinates*) |
| .4 | n. MORAL COURAGE (*Intellectual honesty, willingness to stand up and be counted*) |
| .5 | o. SELF-DISCIPLINE (*Conducts himself in accordance with accepted standards*) |
| .5 | p. SELF-IMPROVEMENT (*Takes action to improve himself*) |
| .5 | q. SOCIABILITY (*Participates freely and easily in social and community activities*) |
| .5 | r. STAMINA (*Performs successfully under protracted physical and mental stress*) |
| .3 | s. TACT (*Says or does what is appropriate without giving unnecessary offense*) |
| .5 | t. UNDERSTANDING (*Appreciation of another person's viewpoint*) |
| | ← SCORE |

## PART VI - OVERALL DEMONSTRATED PERFORMANCE

| RATING | EXPECTED DISTRIBUTION OF 100 OFFICERS RATED | OVERALL DEMONSTRATED PERFORMANCE (1) RATER | VALUE | INDORSER |
|---|---|---|---|---|
| a. OUTSTANDING | 1 | • | 100 | |
| b. EXCEPTIONAL | 11 | • | 90 | |
| c. SUPERIOR | 111 | | 80 | |
| | 1111111111 | | 70 | |
| d. EXCELLENT | 11111111111111111111 | | 60 | |
| | 11111111111111111111 | | 50 | |
| e. EFFECTIVE | 1111111 | | 40 | |
| | 1111 | | 30 | |
| f. MARGINAL | 11 | • | 20 | |
| g. INADEQUATE | 1 | • | 10 | |
| | | SCORE | | |