AD-A194 723

DOCUMENTATION PAGE

DTIC FILE COPY

	1b. RESTRICTIVE MARKINGS – –
2a. SECURITY CLASSIFICATION AUTHORITY – –	**3. DISTRIBUTION/AVAILABILITY OF REPORT** Approved for public release; distribution unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE – –	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) ARI Research Note 88-27	**5. MONITORING ORGANIZATION REPORT NUMBER(S)** – –

6a NAME OF PERFORMING ORGANIZATION ARI Field Unit at Fort Leavenworth, Kansas	6b. OFFICE SYMBOL (If applicable) PERI-SL	7a. NAME OF MONITORING ORGANIZATION – –
6c. ADDRESS (City, State, and ZIP Code) P.O. Box H (PERI-SL) Fort Leavenworth, Kansas 66027-0347		7b. ADDRESS (City, State, and ZIP Code) – –
8a. NAME OF FUNDING/SPONSORING ORGANIZATION U.S. Army Research Institute	8b. OFFICE SYMBOL (If applicable) PERI-SZ	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER – –

DTIC
ELECTE
MAY 1 6 1988
S
E

8c. ADDRESS (City, State, and ZIP Code) 5001 Eisenhower Avenue Alexandria, VA 22333-5600	10. SOURCE OF FUNDING NUMBERS			
	PROGRAM ELEMENT NO. 62722A	PROJECT NO. 2Q1627 17A790	TASK NO. 1.3.3	WORK UNIT ACCESSION NO. 1.3.3.H1

11. TITLE (Include Security Classification)
The Effectiveness of a Rater Training Booklet
in Increasing Accuracy of Performance Ratings

12. PERSONAL AUTHOR(S)
Delane K. Garlinger

13a. TYPE OF REPORT Interim Report	13b. TIME COVERED FROM _____ TO June 86	14. DATE OF REPORT (Year, Month, Day) April 1988	15. PAGE COUNT 55

16. SUPPLEMENTARY NOTATION
– –

17.	COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Rater Training Subjective Measurement Rating Scale
			Rater Accuracy Performance Appraisal Ratings
			Rater Error Performance Measurement

19. ABSTRACT (Continue on reverse if necessary and identify by block number)
This research note examines the possibility of increasing the accuracy of performance ratings through the use of an instruction booklet to train raters. The booklet focussed on the cognitive processes of raters which can lead to faulty conclusions, bias, and misperceptions, with the goal of increasing the objectivity and fairness of observations and ratings.

The research utilized a between-groups "post test only" design. Consensus of opinion among subject matter experts provided the "true score" against which the subjects' ratings were compared for accuracy. The dependent measure was the absolute deviation score of each individual's rating from the "true score". The results indicated no significant group differences in accuracy. This finding is discussed, and suggestions for further research are given.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a NAME OF RESPONSIBLE INDIVIDUAL Delane K. Garlinger	22b. TELEPHONE (Include Area Code) 913/684-4933	22c. OFFICE SYMBOL PERI-SL

DD Form 1473, JUN 86   *Previous editions are obsolete.*   SECURITY CLASSIFICATION OF THIS PAGE

i

UNCLASSIFIED

THE EFFECTIVENESS OF A RATER TRAINING BOOKLET
IN INCREASING ACCURACY OF PERFORMANCE RATINGS

Delane K. Garlinger
U.S. Army Research Institute

for

ARI Field Unit at Fort Leavenworth, Kansas
Stanley M. Halpin, Chief

SYSTEMS RESEARCH LABORATORY
Robin L. Keesee, Director

Accession For	
NTIS GRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

U. S. Army

Research Institute for the Behavioral and Social Sciences

April 1988

U. S. ARMY RESEARCH INSTITUTE

FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the

Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

WM. DARRYL HENDERSON
COL, IN
Commanding

Technical review by

Rex R. Michel
Earl C. Pence

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE		Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS - -

2a. SECURITY CLASSIFICATION AUTHORITY - -	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release;
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE - -	distribution unlimited

4. PERFORMING ORGANIZATION REPORT NUMBER(S) ARI Research Note 88-27	5. MONITORING ORGANIZATION REPORT NUMBER(S) - -

6a. NAME OF PERFORMING ORGANIZATION ARI Field Unit at Fort Leavenworth, Kansas	6b. OFFICE SYMBOL (If applicable) PERI-SL	7a. NAME OF MONITORING ORGANIZATION - -

6c. ADDRESS (City, State, and ZIP Code) P.O. Box H (PERI-SL) Fort Leavenworth, Kansas 66027-0347	7b. ADDRESS (City, State, and ZIP Code) - -

8a. NAME OF FUNDING/SPONSORING ORGANIZATION U.S. Army Research Institute	8b. OFFICE SYMBOL (If applicable) PERI-SZ	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER - -

8c. ADDRESS (City, State, and ZIP Code) 5001 Eisenhower Avenue Alexandria, VA 22333-5600	10 SOURCE OF FUNDING NUMBERS			
	PROGRAM ELEMENT NO. 62722A	PROJECT NO 2Q1627 17A790	TASK NO 1.3.3	WORK UNIT ACCESSION NO 1.3.3.H1

11 TITLE (Include Security Classification) The Effectiveness of a Rater Training Booklet in Increasing Accuracy of Performance Ratings

12. PERSONAL AUTHOR(S) Delane K. Garlinger

13a TYPE OF REPORT Interim Report	13b. TIME COVERED FROM _____ TO June 86	14. DATE OF REPORT (Year, Month, Day) April 1988	15. PAGE COUNT 55

16. SUPPLEMENTARY NOTATION - -

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Rater Training Subjective Measurement Rating Scale
			Rater Accuracy Performance Appraisal Ratings
			Rater Error Performance Measurement

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This research note examines the possibility of increasing the accuracy of performance ratings through the use of an instruction booklet to train raters. The booklet focussed on the cognitive processes of raters which can lead to faulty conclusions, bias, and misperceptions, with the goal of increasing the objectivity and fairness of observations and ratings.

The research utilized a between-groups "post test only" design. Consensus of opinion among subject matter experts provided the "true score" against which the subjects' ratings were compared for accuracy. The dependent measure was the absolute deviation score of each individual's rating from the "true score". The results indicated no significant group differences in accuracy. This finding is discussed, and suggestions for further research are given.

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION Unclassified
22a NAME OF RESPONSIBLE INDIVIDUAL Delane K. Garlinger	22b TELEPHONE (Include Area Code) 913/684-4933	22c. OFFICE SYMBOL PERI-SL

DD Form 1473, JUN 86   *Previous editions are obsolete.*   SECURITY CLASSIFICATION OF THIS PAGE

i

UNCLASSIFIED

THE EFFECTIVENESS OF A RATER TRAINING BOOKLET IN INCREASING ACCURACY
OF PERFORMANCE RATINGS

EXECUTIVE SUMMARY

Objective:

The U.S. Army Science Board summer study of 1985 concluded that performance
measurement is the key to training improvement in today's Army. To optimize
training effectiveness there must be a complete cycle of performance measure-
ment and feedback. Performance ratings by observers is a common method for
obtaining the needed measurement, but ratings have often been criticized as
being too subjective, non-standard, and inaccurate to be of value. The objec-
tive of this research was to develop and evaluate the effectiveness of a train-
ing booklet designed to improve the accuracy of performance ratings.

Procedures:

This study utilized a between-groups, post test only experimental design.

The participants were military volunteers. Three instructors at the Com-
bined Arms Services Staff School (CAS^3) volunteered to serve as subject matter
experts, and twenty-two students of the Command and General Staff College
(CGSC) volunteered to participate as subjects.

A stimulus videotape, rater training booklet, and two parallel rating forms
were developed for the study. The stimulus videotape depicted the S2 (Intelli-
gence) and S3 (Operations) portions of a battalion OPORD briefing, and the
rating scales were developed to measure aspects of performance considered im-
portant in such briefings. The rater training booklet was based upon available
published literature pertinent to rater accuracy, and presented information
and examples of common mistakes made by raters.

The participants were randomly assigned to either treatment or control
groups. While the treatment group received rater training the control group
completed an anagram task. Both groups then viewed the videotape and made
their ratings on the S2 and S3 rating forms.

Consensus of opinion of the subject matter experts provided the "true
score" against which subjects' ratings were compared for accuracy.

Findings:

The absolute deviation scores of each individual's ratings from the "true score" provided by subject matter experts were analyzed using multivariate analysis of variance procedures. The results indicated no significant group differences on these measures.

Both the treatment and the control groups performed well as raters, with generally high interrater agreement and substantial agreement with the subject matter experts. The possibility exists that the training was not strong enough for trainees who were already at a high level of expertise or that the stimuli for rating were "too easy" for this sample of raters resulting in a restricted range of performance.

Furthermore, analysis of participant demographic variables indicated that treatment and control groups were not equivalent in terms of prior battalion experience. All participants of the control group had such prior experience whereas only approximately half of the treatment group did so. Regression analysis indicated that this variable accounted for a significant portion of the variance on the S2 form. Therefore, it is felt that this variable could have been a confounding factor in this study.

The psychometric qualities of the rating scale were examined and results indicated that the scales may have merit as an applied measurement instrument.

Utilization of Findings:

While the training booklet did not improve the accuracy of the partici- pants, it may be beneficial for personnel who are less proficient than the CGSC students who were subjects in this study. Also, the implication that the accu- racy of ratings is related to the prior experience of the rater on the tasks to be rated can have significance in the selection of training evaluators.

The rating scales developed for purposes of this study could be used as applied measurement instruments for training performance evaluation.

THE EFFECTIVENESS OF A RATER TRAINING BOOKLET IN INCREASING THE ACCURACY
OF PERFORMANCE RATINGS

CONTENTS

LIST OF TABLES

LIST OF FIGURES

THE EFFECTIVENESS OF A RATER TRAINING BOOKLET IN INCREASING ACCURACY
OF PERFORMANCE RATINGS

Introduction

The U.S. Army of today is striving for excellence. Inherent in this effort
towards excellence is the need to enhance the quality of training. The devel-
opment of performance measures is an essential step in the training process as
reliable and accurate measures of performance are necessary to diagnose defi-
ciencies, provide trainees with feedback, and to evaluate the effectiveness of
alternative training systems.

Ideally, performance measure methodologies/devices should be quantifiable
and objective. Automated performance measurement offers the greatest hope of
achieving quantifiable and objective evaluations, and are currently being
specified as a requirement for modern training simulators (Vreuls & Obermayer,
1985). However, there are many difficulties in the design and implementation
of automated measurement (for a comprehensive overview, see Vreuls & Obermayer,
1985) which necessitates a search for viable alternatives during the interim.

Furthermore, automated instrumentation and performance measurement systems
are not possible for all simulation exercises, or for all simulated tasks.
There are situations and tasks which require, and are likely to continue to
require, measurement by human observation.

Performance measurement based upon observations and ratings has been criti-
cized as being too subjective, non-standard, and inaccurate to be of value.

According to Walter Borman (1978), the causes of problems with performance
ratings can be classified into four general categories, as follows:

1. Raters often have little opportunity to observe the ratees performing
the tasks on which they are being evaluated.

2. The rating format is not appropriate for the rating task.

3. Error is introduced into ratings due to organizational constraints on
raters to provide performance evaluations which differ from their best esti-
mates of ratees' true performance. An example of this would be the general
demand perceived by raters to provide high ratings when they must subsequently
meet face-to-face with the ratee to discuss the ratings.

4. Raters have inadequate training in regards to evaluating and rating
performance.

Problem number one has been shown to be relevant to rating performance
measurement during various Army training exercises. Often, the number of
raters assigned to observe and rate performance is inadequate to provide an
opportunity for the relevant tasks to be fully and adequately observed. For

1

example, Barber & Solick (1980) report that in simulator exercises the controllers often do "double duty" as performance evaluators. In such a situation, the rater is usually completely occupied in his role as controller and has minimal time to devote to observation for purposes of evaluation. In the Barber & Solick study (1980), it was reported that controllers felt that they had little opportunity to observe performance and that the observation/rating requirements interfered with their controller duties, at least some of the time. On the other hand, evaluators whose sole purpose was observing and rating performance felt that they had extensive opportunity to observe, and placed more confidence in their ratings than did the controllers.

It would appear that an appropriate solution to problem number one would be to provide personnel for training exercises whose dedicated purpose is evaluation rather than using controllers for this "extra" duty.

Problem number two is one which must be given consideration when developing an evaluation plan for any training situation. The published literature regarding rating scale format indicates that no one format is best across all situations (Borman, 1979; Decoths, 1977). However, for any specific rating task some formats may prove better than others. This must be determined based upon the type of training being evaluated and should allow the raters to easily translate the behavior they observe to a specific level on the performance rating scale.

According to published research evidence, when raters are told in advance that their ratings will be used for training and development diagnostic purposes the ratings are significantly more accurate than when the ratings are to be used for promotion, raises, or retention purposes, which tends to inflate ratings (Zedeck & Cascio, 1982; McIntyre, Smith, & Hassett, 1984). Problem number three could be effectively reduced in training situations by informing raters that their evaluations are to be used for training diagnostic and feedback purposes only.

Problem number four is especially appropriate in Army training situations as little or no training is provided for personnel who will rate performance. Controllers are usually given adequate training on the aspects of their duties concerning computer terminal operations and other game-related functions. However, the observation/evaluation responsibilities have been generally overlooked from a training standpoint.

The present research addressed this problem directly by developing and evaluating the effectiveness of an instruction booklet for raters designed to strengthen the rating skills needed to improve the accuracy of performance ratings. The instructional material was based upon the cognitive processes involved in rating performance and emphasized skills to enhance objective and fair ratings.

Background

After reviewing the published research results relating to observer-rater training programs, Spool (1978) concluded that "accuracy in observation can be improved by training observers to minimize rating errors" (pp. 866-867). Spool was apparently led to make this optimistic statement based upon the assumption that decreasing various common errors made by raters would automatically lead to increased accuracy. This has been shown to be a questionable assumption.

The majority of the studies concerning rater training programs upon which Spool based his optimism utilized measures of the various psychometric errors such as halo, leniency/severity, central tendency, contrast (see Table 1) as the dependent variables, with no measures of accuracy included for comparison (Bernardin, 1978; Bernardin & Walter, 1977; Latham, Wexley & Pursell, 1975). Data from later studies, which included accuracy as a dependent variable, indicate that, at best, accuracy is unaffected by traditional psychometric error training, and there is some research evidence that such training may actually reduce accuracy (Bernardin & Pence, 1980; Borman, 1979; Pulakos, 1984).

These findings have been primarily attributed to the basic format of the traditional programs designed to reduce psychometric error (Bernardin & Pence, 1980; Pulakos, 1984). These programs were primarily concerned with changing the raters' response distributions, and have presented trainee participants with illustrations of rating distributions portraying the different types of error. A distribution with a negative skew would indicate leniency error; a positive skew would indicate severity error; a narrow bellshaped distribution with very low variance would illustrate the error of central tendency. Trainees would be urged to conform more closely to a normal distribution across raters. In addition, ratings at about the same level across the dimensions being rated for any one individual (resulting in intercorrelations of dimensions) would be illustrative of halo error. Therefore, trainees would be encouraged to spread out their ratings on the different dimensions relative to a single ratee. Anecdotal evidence obtained in a study by Pulakos (1984) indicates that one possible reason why this type of training may reduce accuracy (as well as statistical measures of error) is because concern for making errors may direct the raters attention away from observing relevant ratee behavior to monitoring their own rating behavior. Several rater trainees in the Pulakos study reported purposely spreading out their ratings to avoid halo error when they felt it may have been more appropriate to rate some ratees more uniformly across the dimensions.

Table 1.

Definition of Psychometric Terms

Accuracy - agreement with the consensus of expert judges.

Central tendency error - the tendency of a rater to evaluate all ratees
in the middle of the scale and to avoid extreme positions
(Anastasi, 1982). This error is referred to in the training
booklet as "middle of the road" error.

Contrast error - the tendency of a rater to evaluate a ratee relative to
other ratees and not relative to the specific job requirements
or the standard provided by the rating scale (DeGregorio & Foti,
1985). Contrast error is referred to in the training booklet as
comparison effects.

Halo effect - the tendency to be unduly influenced by a single favorable
or unfavorable trait, which colors the judgment of the ratee's
other traits, (Anastasi, 1982).

Leniency error - the tendency of raters to evaluate all ratees on the high
end of the scale, thereby avoiding assigning unfavorable ratings.
Referred to in the training booklet as generosity error.

Severity error - the tendency to rate all ratees at the low end of the
rating scale.

First-impression error - the tendency to make an initial favorable or un-
favorable judgment about the person to be rated, and then uncon-
sciously ignore or discard subsequent information so that the
initial impression is supported.

Similar-to-me effect - the tendency to judge individuals perceived as
similar to self more favorably than warranted based on
performance criteria.

The common ability and motivational factors involved in the various dimen-
sions of performance by a single individual makes dimension intercorrelation
inevitable, and efforts to eliminate the intercorrelations will result in dis-
torted ratings and decreased accuracy. Likewise, skewed rating distributions
may reflect reality resulting from underlying common characteristics of the
rated population due to selection procedures, similar training, background and
experience, etc. (Bernardin & Pence, 1980). Recognition of these factors, and
the lack of success of the psychometric error training programs in increasing
accuracy, led some researchers to attempt to apply a statistical approach to

controlling extraneous influences (primarily halo effect) on performance ratings. Generally, these methods have been condemned as being statistically and psychometrically unsound.

A more promising approach to increasing the accuracy of performance ratings has been evolving over the past few years. This approach, which explicitly seeks to improve accuracy (rather than reduce error) is based on the cognitive processes involved in the task of rating the performance of others. Such an approach concentrates on training raters to observe behavior more accurately and fairly rather than providing instruction on "how to or not to rate" with regard to response distributions (Bernardin & Pence, 1980). Raters receive training in such concepts as the independence of different aspects of job performance and the need to recognize this independence when rating performance on the various aspects, the importance of fair and critical evaluations, ways to strengthen observation skills, and criteria for determining effective and ineffective ratee performance on the dimensions being rated.

This approach has been compared to the more traditional psychometric error training (Bernardin & Pence, 1980; Pulakos, 1984). Although still inconclusive the results suggest that training in procedures which would facilitate the development and use of a common, job-relevant frame of reference for evaluating behavior results in ratings higher in accuracy than training procedures designed to explicitly change rating distributions.

Objective

The objective of this research was to develop and evaluate a training booklet for individuals who are assigned to serve as performance observers/raters.

Methodology

Design:

This study utilized a between-groups, post test only experimental design.

The independent variable was rater training vs. no rater training conditions. The primary dependent variable of interest was accuracy of performance ratings. For purposes of this research, accuracy is defined as agreement with the evaluation of expert judges, which was considered to be the criterion or "true score" of performance.

Materials and Equipment:

This study was conducted in a laboratory setting and utilized the following materials and equipment.

5

a. <u>Rater training instruction booklet</u>. The instruction booklet was constructed to contain three principal elements, as follows:

(1) Motivation. The introductory material contained statements concerning the purpose of the booklet, the importance of accurate ratings, and assurance that ratings would be used for training and feedback purposes only and that the results would not be used to make decisions regarding individual promotions, awards, etc.

(2) Obstacles to Accuracy. This section contained an explanation of extraneous variables which can decrease the accuracy of performance ratings, together with examples of each. Explanations and examples of the following were included. Comparison effects, first-impression error, severity and generosity effects, halo effect, and similar-to-me effect.

(3) Practice Exercise. This section contained questions to check the understanding of the previous section. Four vignettes were provided with accompanying multiple choice questions. The subjects were provided an answer sheet so they could check the correctness of their answers. The answer sheet also provided the page number corresponding to each answer so that the subject could review the material related to an incorrect response.

A copy of the rater training booklet is attached as Appendix A.

b. <u>Stimulus Videotape</u>. The script used in producing the videotape was developed by transcribing the S2 (Intelligence) and S3 (Operations) portions of a battalion operations order briefing which was videotaped during a battalion training exercise. The transcribed script was used by two volunteer actors to recreate the briefings on videotape to be used as stimulus material for this study. The briefings were recreated in this manner due to the poor clarity and audibility of the original videotape. A copy of the transcribed script is attached as Appendix B.

c. <u>Rating Scale</u>. Two parallel forms (S2 Form and S3 Form) of a 7-point Likert-type scale were developed to be used in making the ratings for the purposes of this study. Each form was constructed to measure performance on four dimensions as follows:

1. <u>Clarity</u> - measured the extent to which the ratee used accepted military terminology and language which was clear, precise and unambiguous. (Questions 1 and 2).

2. <u>Information adequacy</u> - measured the degree to which the information provided in several key areas was adequate for successful execution of the mission. (Questions 3 through 9).

3. <u>Following commander's guidance</u> - measured how adequately the ratee followed the guidance provided by the commander in several key areas. (Questions 10 through 14 on S2 form; Questions 10 through 13 on S3 form).

6

4. <u>Delivery of briefing</u> - measured how well the ratee organized and paced his briefing and how effectively he used maps and other visual aids. (Questions 15 through 17 on S2 form; Questions 14 through 16 on S3 form).

The clarity factor was constructed as a frequency scale with the following choice points and descriptors, with the corresponding scale values (percentage of time) as reported by Dyer, Mathews, Stulac, Wright and Yudowitch (1976).

Choice Point	Descriptor	Scale Value (% of time)
1	never	2
2	seldom	18
3	occasionally	33
4	about as often as not	50
5	frequently	80
6	usually	90
7	always	100

The factors for adequacy of information, commander's guidance and briefing style were constructed to require judgment of adequacy with the following choice points and descriptors, with the corresponding scale values (means and standard deviations) as reported by Dyer, et al (1976).

Choice Point	Descriptor	Scale Value	
		(X	SD)
1	totally inadequate	-4.90	.412
2	largely inadequate	not reported	
3	somewhat inadequate	-1.88	.73
4	borderline	- .02	.31
5	somewhat adequate	1.32	.79
6	largely adequate	2.86	.99
7	totally adequate	4.62	.84

A copy of both the S2 and S3 forms of the rating scale are attached as Appendix C.

d. <u>Demographic and Opinion Questionnaire.</u> A questionnaire was constructed to collect demographic data which could be used to verify group equivalence on several variables which could potentially impact upon the treatment administered and/or the ability to accurately rate the performance of others. These variables included level of civilian and military education attained, prior training and experience in conducting evaluations, number of years in the military, and prior battalion staff experience.

The opinion portion of the questionnaire was designed to assess the perceived utility of the instruction booklet by those who received the rater training.

A copy of the questionnaire is attached as Appendix D.

e. **Anagram Task Sheet.** Due to scheduling constraints, experimental sessions were conducted simultaneously for both the treatment and control groups. This necessitated the creation of a task which the control group could attend to while the treatment group studied the instruction booklet. An anagram task was constructed for this purpose. A copy of the anagram task is attached as Appendix E.

f. **Videotape Player/Recorder.** This equipment was used to record and play the stimulus videotape material.

Participants:

The participants for this study, both subject matter experts and subjects, were military volunteers recruited from the Command and General Staff College, (CGSC), Fort Leavenworth, Kansas. Three instructors at the Combined Arms and Services Staff School (CAS³) volunteered to participate as subject matter experts whose ratings constituted "true scores" for purposes of the study. Twenty-two students of the Command and General Staff College (CGSC) volunteered to participate in the study as subjects.

Procedures:

a. **Expert Ratings:**

The criteria or "true score" ratings were obtained from the subject matter experts using a modified Delphi technique. The experts were shown the S2 portion of the videotape and then asked to make their individual ratings on each item of the S2 rating scale. The individual ratings were then discussed by the experts in an attempt to arrive at a consensus of opinion where discrepancies existed. The research plan was to use the average of the three judges' ratings on items where consensus could not be achieved in three trials. However, consensus was reached on all items with a maximum of two trials. Experts were provided the opportunity to review the videotape, or portions of the videotape, as they felt the need to do so.

Rating of the S3 portion of the briefing proceeded in the same manner as the S2 portion.

b. **Experimental Groups:**

Treatment and control group sessions were conducted simultaneously. As the subjects arrived in the room designated for the study, they were each presented with an envelope containing materials for the study which was marked

with either a one (1), which designated treatment group, or a two (2), which designated control group. The envelopes were shuffled before distribution to insure random assignment.

After all subjects had arrived and were seated, the researcher gave a brief overview of the study and asked all participants to sign an Informed Consent Form. The treatment group was then given 20 minutes to read and study the training booklet. The control group completed the anagram task during this time.

The participants were then told that they would be shown the S2 and S3 portions of an OPORD briefing and be asked to rate the briefers on the forms provided for that purpose. Before viewing the videotape, they were asked to read the instructions for using the rating form and to look over the forms to familiarize themselves with the contents. The entire videotape was then played (both S2 and S3 portion) for the participants who were allowed only one viewing. The participants were given 15 minutes to complete their ratings.

Before leaving the room, the participants were asked to complete the demographic and opinion questionnaire.

Results

Treatment Effect:

Accuracy: The measure of accuracy used for this study was the absolute deviation score of each individual's ratings from the true score ratings determined by subject matter experts on each of the four factors of each rating scale (S2 and S3). Thus, each individual participant had eight factor scores for analysis. These scores were analyzed using a separate multivariate analysis of variance procedure for the factors of each form. The results of the analyses indicate no significant group differences on these factor scores. The homogeneity of variance assumption was tested for all factors using the Bartlett's test, which indicated that there were no violations of this assumption $(p > .05)$.

Group medians, means, and standard deviations are presented in Table 2. The multivariate analysis of variance tables are presented in Appendix F.

An examination of the group means of the ratings on the individual items which composed the factors of each scale revealed some pattern commonalities for the treatment and control groups, but less commonality between the two experimental groups and the expert ratings. The means and standard deviations of the ratings on individual items are presented for the two groups in Appendix G and the patterns are graphically presented in Figure 1 and Figure 2 for the S2 and S3 forms respectively.

Table 2.

Absolute Deviation Factor Scores

|			Group Means and Standard Deviations			
Form	Factor	Number of Items	Group	Median	Mean	SD
S2	Clarity	2	treatment	2.0	2.09	.94
			control	2.0	1.82	1.60
	Adequacy of Information	7	treatment	11.0	10.73	2.90
			control	10.0	10.00	2.90
	Following Commander's Guidance	5	treatment	14.0	13.73	5.08
			control	13.0	11.73	4.90
	Briefing Delivery	3	treatment	5.0	5.09	2.30
			control	5.0	4.36	1.29
S3	Clarity	2	treatment	2.0	1.45	1.13
			control	2.0	1.64	1.21
	Adequacy of Information	7	treatment	13.0	12.55	5.54
			control	11.0	12.64	5.18
	Following Commander's Guidance	4	treatment	8.0	7.91	2.34
			control	8.0	8.09	3.18
	Briefing Delivery	3	treatment	6.0	3.27	2.20
			control	5.0	3.82	1.83
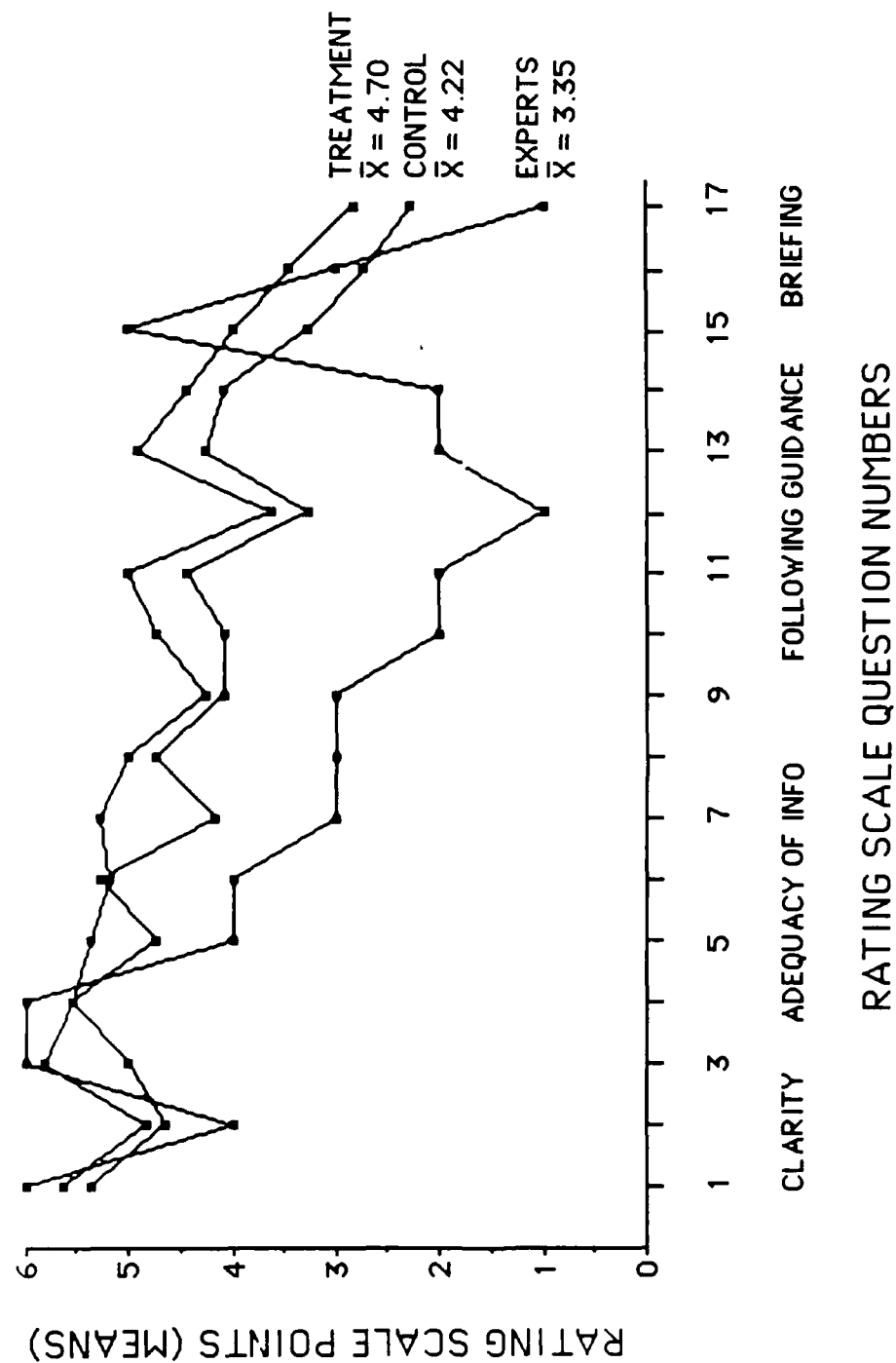
Figure 1. Individual Item Means – S2 Form
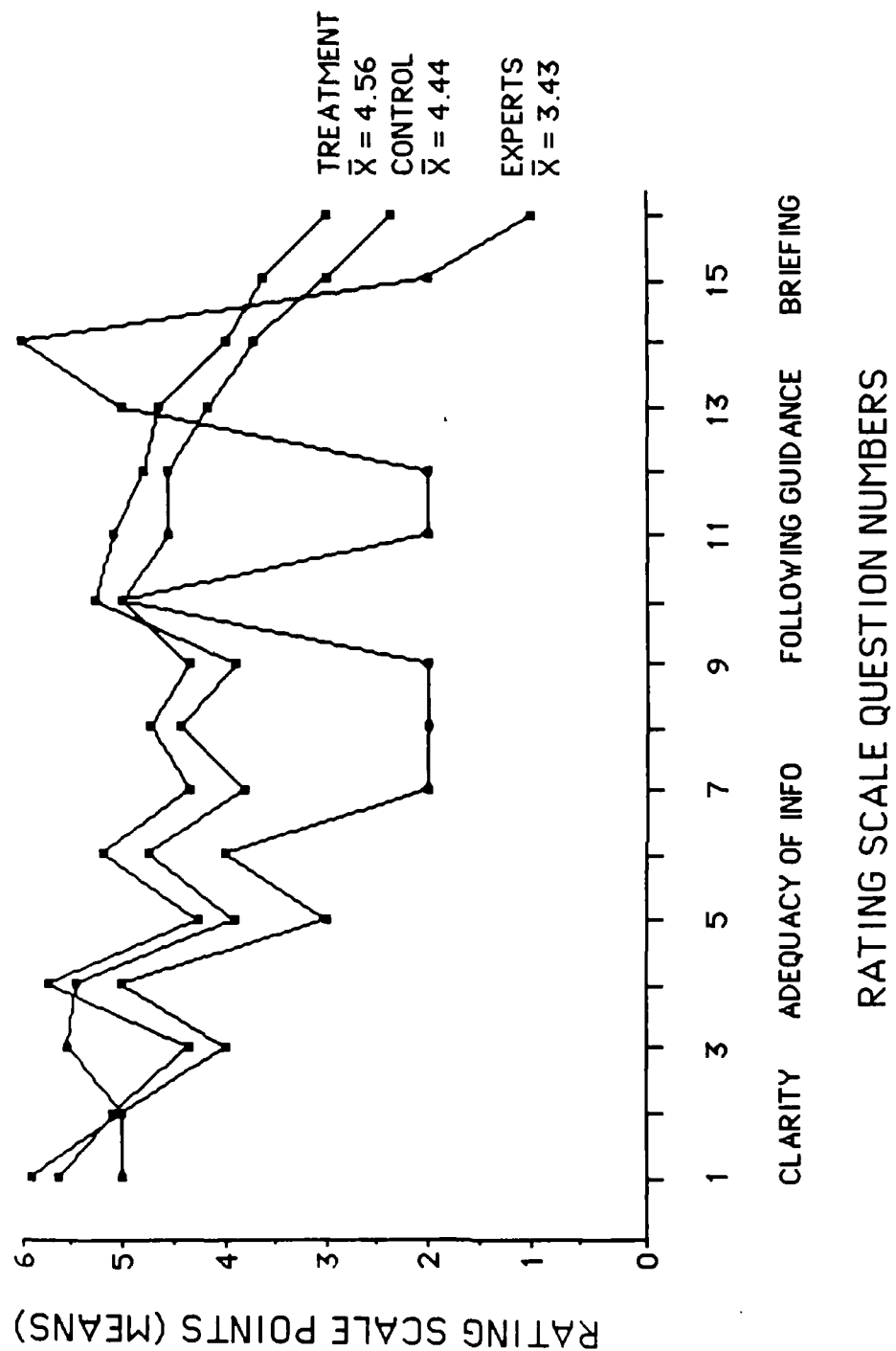
Figure 2. Individual Item Means – S3 Form

It would appear from an examination of the rating patterns on Figures 1 and 2 that the range of ratings for the experts is greater than for either of the experimental groups. However, this may be misleading since the data points for the experts represents one rating resulting from the consensus of opinion for three judges, whereas each data point for the experimental groups represents the arithmetic average of the eleven raters in each group, which would restrict the range of the data points.

Rater Error: Individual and group rating patterns were examined to investigate the effectiveness of the rater training booklet in reducing common rater errors.

Pearson correlation matrices were computed on the eight factors for each group separately. These matrices indicate significant cross-form correlation for the control group, but no significant cross-form correlation for the treatment group. Both groups showed significant correlation of the adequacy of information, commander's guidance, and briefing delivery factors within each form. These correlation matrices are presented in Table 3.

Since the performance of the two briefers was independent and the items included in the factors on the two forms were largely different (but parallel) due to the different content areas, significant cross-form correlation could indicate the presence of error. Therefore, these results indicate that the rater training booklet could have been effective in reducing global rating tendencies, such as central tendency, leniency or severity.

Demographic Data

Since the sample was divided into treatment and control conditions based upon random assignment, the participants were asked to complete a questionnaire concerning demographic variables which could potentially interact with the treatment. Participants' status in regards to civilian and military education, prior evaluation training and experience, number of years in the military, and prior battalion staff experience were examined. Chi-square analysis indicated that the treatment and control groups were statistically equivalent on all of the demographic variables except two: the treatment group had significantly more participants with prior evaluation experience ($p < .05$) and the control group contained significantly more participants who had prior battalion staff experience ($p < .05$).

A regression analysis was conducted to examine the influence of these two variables on the total absolute deviation score of each form. Prior battalion staff experience accounted for 39% of the variance in total deviation scores for the S2 form ($p < .05$), but was not significant for the S3 form. Prior evaluation experience did not explain a significant amount of variance of total absolute deviation scores for either the S2 or S3 forms.

Table 3.

Pearson Correlations of Factors

Treatment Group

	S2 Form				S3 Form			
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 1	Factor 2	Factor 3	Factor 4
S2 Form:								
Clarity	--	.50	.35	.45	.43	.32	.24	.08
Adequacy of Info		--	.62*	.68*	.22	.28	.12	.42
Following Cdr's Guidance			--	.58*	-.004	.24	.07	.09
Briefing Adequacy				--	.22	.05	.13	.04
S3 Form:								
Clarity					--	.70**	.55*	.32
Adequacy of Info						--	.73**	.76**
Following Cdr's Guidance							--	.71**
Briefing Adequacy								--

Control Group

	S2 Form				S3 Form			
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 1	Factor 2	Factor 3	Factor 4
S2 Form:								
Clarity	--	.47	.10	.41	.45	.22	.26	-.06
Adequacy of Info		--	.79**	.63*	.41	.80**	.70**	.69*
Following Cdr's Guidance			--	.53*	.09	.70**	.48	.58*
Briefing Adequacy				--	.40	.61*	.55*	.44
S3 Form:								
Clarity					--	.50	.42	.32
Adequacy of Info						--	.79**	.80**
Following Cdr's Guidance							--	.46
Briefing Adequacy								--

NOTE: $* \ p < .05$ $** \ p < .01$

14

Opinion Questionnaire: All participants who received the rater training judged the instruction booklet to have been somewhat helpful (scale point 4 on a 5 point scale) to them in making ratings.

Rating Scale:

Reliability: Interrater reliability was estimated for each group separately within each of the eight factors. The coefficient alpha method was used which treats each rater as an item (Cronbach, 1951) and averages mean ratings over n raters. This method is equivalent to estimating rater reliability by interclass correlations as discussed by Guilford (1954).

Interrater reliability coefficients for the treatment and control groups for the eight factors are presented on Table 4.

Table 4

Interrater Reliability Coefficients*

S2 Form:	Treatment	Control
Clarity	.90	.98
Adequacy of Information	.63	.74
Following Commander's Guidance	.53	.40
Briefing Delivery	.79	.71
OVERALL (All Factors Combined)	.83	.86
S3 Form:		
Clarity	.96	.97
Adequacy of Information	.78	.73
Following Commander's Guidance	.12	.70
Briefing Delivery	.76	.80
OVERALL (All Factors Combined)	.84	.85

*Note: Interrater reliability coefficients computed by coefficient alpha method - Cronbach (1951).

The eight factors were examined for adequacy of content sampling with the Cronbach Alpha reliability measure. This measure reveals the extent to which the factors are homogeneous or internally consistent. These reliability coefficients are provided in Table 5.

Table 5.

Internal Consistency Reliability Coefficients

	Treatment	Control	Total Sample
S2 Form:			
Clarity	.82	.86	.84
Adequacy of Information	.85	.88	.86
Following Commander's Guidance	.73	.79	.79
Briefing Delivery	.90	.86	.89
OVERALL (All Factors Combined)	.91	.92	.92
S3 Form:			
Clarity	.44	.88	.78
Adequacy of Information	.90	.87	.88
Following Commander's Guidance	.75	.50	.69
Briefing Delivery	.95	.85	.90
OVERALL (All Factors Combined)	.94	.91	.92

Validity:

Borman (1978) suggested that correlation of group mean ratings with the "true" scores provided by subject matter experts would yield an estimate of validity. This procedure was used in computing the estimated validity coefficient for the S2 and S3 forms for the treatment and control groups separately, and with the groups combined as a total sample. These coefficients of validity are presented in Table 6.

Table 6.

Estimated Validity Coefficients

	Treatment	Control	Total Sample
S2 Form:			
Clarity	.99	.99	.99
Adequacy of Information	.76	.77	.84
Following Commander's Guidance	.92	.94	.94
Briefing Delivery	.99	.99	.99
S3 Form:			
Clarity	.99	.99	.99
Adequacy of Information	.84	.71	.87
Following Commander's Guidance	.01	.08	.05
Briefing Delivery	.88	.96	.93

Conclusions

Although the training booklet was perceived by the participants as benefi-
cial in the rating task, no firm conclusions can be drawn concerning the actual
utility of the booklet for several reasons. Results could have been con-
founded by non-equivalency of the treatment and control groups in regards to
prior battalion staff experience. Regression analysis indicated that this
variable accounted for a significant portion (39%, $p < .05$) of the variance of
deviation scores on the S2 form. Since all participants of the control group
had prior battalion staff experience, this could account for the lower devia-
tion scores for this group. Such prior experience could have provided
the basis for a frame of reference for the control group which more nearly ap-
proximated the frame of reference of the expert judges, all of whom had previ-
ous battalion command experience. This could have had considerable impact on
judgment, especially on questions were the performance criteria was less
concrete. Future research should control for relevant prior experience of the
research sample.

In addition, it is quite possible that the training provided by the instruction booklet was not strong enough to provide significant training effects. Both experimental groups (treatment and control) performed quite well as raters on the experimental task. Interrater agreement was high in both groups on most factors, as were the correlations of their ratings with expert ratings. Perhaps a more powerful treatment is required to produce observable changes in performance when the trainees are at an already high level of expertise. Better results may be achieved using the current training booklet with participants who are less proficient. Future research efforts should assess this possibility.

The low difficulty level of the rating task used for this research could also have contributed to the findings. The stimulus material to be rated could have been so easy that the range of rating performance was restricted. Follow-up research should strive for more "real-world" quality in the stimulus material to be rated.

The rating scales developed for this study have been shown to have potential merit as applied measurement instruments. Reliability and estimated validity were generally high, at least with this sample, and the factors of the scales could be expanded to enhance diagnostic measurement capability.

References

Anastasi, A. (1982). Psychological Testing (5th ed.). New York: MacMillan.

Barber, H.F. and Solick, R.E. (1980). MILES Training and Evaluation Test, USAREUR: Battalion Command Group Training. Army Research Institute Research Report 120.

Bernardin, H.J. (1978). Effects of rater training on leniency and halo errors in student ratings of instructors. Journal of Applied Psychology, 63, 301-308.

Bernardin, H.J. and Pence, E.C. (1980). Effects of rater training: creating new response sets and decreasing accuracy. Journal of Applied Psychology, 65, 60-65.

Bernardin, H.J. and Walters, C.S. (1977). Effects of rater training and diary-keeping on psychometric errors in ratings. Journal of Applied Psychology, 64, 410-421.

Borman, W.C. (1978). Exploring upper limits of reliability and validity in job performance ratings. Journal of Applied Psychology, 63, 135-144.

Borman, W.C. (1979). Format and training effects on rating accuracy and rater errors. Journal of Applied Psychology, 64, 410-421.

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. Psychometrica, 16, 297-334.

DeCoths, T.A. (1977). An analysis of the external validity and applied relevance of three rating formats. Organizational Behavior and Human Performance, 19, 247-266.

DeGregorio, M.B. and Foti, R.J. (1985). A Review of Rater Training Approaches and Effectiveness. Technical report for the Office of Naval Research: N00014-85-K-0289.

Dyer, R., Matthews, J.J., Stulac, J.F., Wright, C.E., and Yudowitch, K., (1976). Questionnaire Construction Manual, Annex, Literature Survey and Bibliography. (ARIBSS p-77-2).

Guilford, J.P. (1954). Psychometric Methods (2nd ed). New York: McGraw-Hill.

Latham, G.P, Wexley, K.N. and Purcell, E.D. (1975). Training managers to minimize rating errors in the observation of behavior. Journal of Applied Psychology, 60, 550-555.

19

McIntyre, R.M., Smith, D.E. and Hassett, C.E. (1984). Accuracy of performance ratings as affected by rater training and perceived purpose of rating. Journal of Applied Psychology, 69, 147-156.

Pulakos, E.D. (1984). A comparison of rater training programs: error training and accuracy training. Journal of Applied Psychology, 69, 581-588.

Spool, M.D. (1978). Training programs for observers of behavior: a review. Personnel Psychology, 31, 853-888.

Vruels, D. and Obermayer, R.W. (1985). Human-system performance measurement in training simulators. Human Factors, 27(3), 241-250.

Zedeck, S. and Cascio, W.F. (1982). Performance appraisal decisions as a function of rater training and purpose of the appraisal. Journal of Applied Psychology, 67, 752-758.

APPENDIX A

RATER TRAINING BOOKLET

Instructions

The following pages contain information which may be useful to you when rating the performance of others. Please read this information carefully and answer the practice questions provided.

You will have 20 minutes to study this material and answer the practice questions.

RATER TRAINING BOOKLET

U.S. Army Research Institute
Fort Leavenworth Field Unit

You are about to serve as an evaluator for a battalion staff training exercise. As an evaluator, your task will be to observe and rate the performance of the exercising staff. The information provided by your evaluation is enormously important to the learning process of the trainees. It is through the evaluation information that the staff will gain insight into its performance strengths, weaknesses and the degree to which current training objectives and goals are being achieved. The evaluation will also enable the staff to set appropriate training goals and objectives for the future. Quality training experiences are essential to the readiness of U.S. forces, and your contribution as evaluator is vital to the training mission.

Accurate performance evaluation is a complex task with several potential obstacles. The purpose of this training booklet is to alert you to the existence of these obstacles and to offer guidance on how to avoid them. This will help you to observe and evaluate more objectively and accurately so that the staff you are assisting will have quality evaluation support for their training efforts.

The results of your evaluation will be used for training and development purposes only, and you are assured that this information will in no way be used to make decisions regarding individual promotions, awards, punishments or penalties.

Obstacles to Rating Accuracy

When an individual observes and evaluates others, systematic rating errors often occur. Rating errors are mistakes in judgement resulting from misperceptions, predispositions, or other subjective, extraneous influence. Evaluators are usually unaware that they are making these judgement errors, but once they become familiar with the common obstacles, they can usually correct themselves and achieve a functional level of accuracy in rating performance.

The following paragraphs contain descriptions and examples of the most common types of bias which can adversely affect the rating process. As you read these paragraphs, try to think of other examples of each kind of judgement error from your own personal experiences. This will make the material more meaningful to you as an individual, and help you understand your personal thinking processes in making judgements of others.

Comparison Effects is an error in judgement which results from the tendency for a rater to evaluate a person relative to other individuals rather than on the requirements of a job or task. A rating should be given on the basis of the criteria established prior to the rating process and specified on the rating form, not on the basis of a comparison with others.

For example, think of the best looking man or woman you have ever met. Rate this individual on a 7-point scale, with 7 representing outstanding in terms of physical attractiveness. Now think of your favorite glamorous movie star. Rate the movie star on the same 7-point scale you used to rate the previous individual. Now, re-rate the first person. If you just lowered the rating of the

first person, you have made a comparison error. Each person should be rated on the degree to which they fulfill your predetermined criteria for physical attractiveness, not on how he or she compares with others.

Comparison error is especially troublesome in performance evaluation conducted for the purpose of training enhancement. The team or individuals you will be evaluating need to know how they perform in relation to a specific criteria, not how they compare to other teams or individuals. You may have observed someone in the past who performs a particular task extremely well, or extremely poorly. If you use this past experience to make comparisons with the individual presently being evaluated, you would be making a contrast error in judgement. If the present individual achieves an average level of performance on the task, rating him in comparison to the outstanding person would make his performance look bad and result in a lower rating than actually deserved. Conversely, contrasting him with the poorly performing individual would make average performance look superior in comparison, resulting in a higher rating than appropriate. Either way, the evaluation would result in misleading performance information upon which to base plans concerning future training needs.

First-impression error results from the tendency for a rater to make an initial favorable or unfavorable judgement about the person to be rated, and then unconsciously ignore or discard subsequent information, so that the initial impression is supported.

Everyone has a tendency to occasionally form first impressions when initially meeting someone new. For example, if a person you have just met seems reluctant to look at you while speaking or avoids eye contact, you may have a first impression that this is a dishonest person or perhaps has something to hide. During subsequent contact with this person you might erroneously view other behaviors in accordance with your concepts of dishonesty, thus confirming your original conclusion. Perhaps the person is actually scrupulously honest, but failed to look at you or make eye contact because of shyness rather than an untrustworthy nature.

The team or individuals which you evaluate may have trouble "getting off the ground" initially and then go on to better performance, or get off to a flying start only to deteriorate as the exercise proceeds. If you allow these first impressions to color the way you see the remainder of their task performance, the evaluation will be inaccurate.

"Middle of the road" error is committed by the rater who wants to play it safe. This error refers to evaluators who consistently give ratings at, or close to the midpoint of the scale, even when the performance of the person or team being evaluated clearly warrants a substantially higher or lower rating.

Suppose you were asked to rate different food items on a 5-point scale for taste. If you gave an overcooked burger from a fast food restaurant and a special dish from a 5-star restaurant both a rating of 3, you would probably be committing a "middle of the road" error. This is an extreme example, but serves to illustrate the concept. This is not to say that you should never give average ratings. In fact, the performance of many of the individuals that you

evaluate will be average and should be rated near the midpoint of the scale. However, you should stay alert to performance differences that warrant a higher or lower rating.

Severity and Generosity errors are committed by the evaluator who is consistently too hard or too easy in rating the performance of others. This type of judgement error is closely related to the "middle of the road" error discussed above. All three errors (middle of the road and severity and generosity) reflect the failure of the evaluator to recognize and record differences in levels of performance. When everyone is given approximately the same rating, whether that rating is high, low or average, the ratings are of no utility in determining training deficiencies and offer no insight as to how well current training objectives have been achieved. Rating according to the criteria specified on the rating scale rather than personal, subjective criteria will help you overcome any tendency to make this type of error.

Halo effect refers to the error of making inappropriate generalizations from one aspect of a person's performance to all aspects of the person's performance. For example, if you know that John Smith was a college football quarterback, currently stars on the post rugby team, swims, jogs and plays a superb game of golf, you may conceptualize him as a superior all-around athlete. If you later had to rate John's skill as a tennis player, you may be influenced by your other knowledge concerning his athletic ability, and rate him higher than justified. In reality, John may be a poor tennis player. The skills required to play tennis well may not actually be related to the skills required to play golf, swim, or play football quarterback even though they all fall under the general category of athletics. This error can also operate in reverse if you allow poor performance on a task to negatively influence your rating of the individual on an apparently similar or related task.

In your duties as evaluator, you may observe that a team or individual performs some category of tasks quite well, or poorly. This should in no way affect your evaluation on future tasks which appear similar or related. Different performance measures are not always as related as they may appear.

To further help you avoid this judgement error, do not listen to comments about the group you are evaluating, or about any of the individuals who comprise the group Also, if more than one evaluator is making ratings of the training exercise, be certain that all evaluators assign their ratings independently. Group discussion should be avoided until after all observations and ratings are completed.

Similar-to-me-effect is an error which occurs when raters judge more favorably those individuals whom they perceive as similar to themselves. That is, the more closely an individual resembles the rater in attitudes or background, the stronger the tendency of the rater to judge the individual favorably. This seems to occur because of the human tendency to like or think more highly of others who are perceived to be like us rather than unlike us because it is flattering and reinforcing. This effect may be acceptable and adaptive in social situations, but is a troublesome source of error when evaluating performance.

Practice Exercise

Now that you have learned about the potential obstacles in the path of accurate ratings, the following practice exercise will help reinforce your learning. Please read the following questions and circle the letter of the answer which you think is correct. After you have answered all of the questions, you may check the accuracy of your response by referring to the Answer Sheet. In addition to providing the correct response, the Answer Sheet will also refer to the page in the training material where information concerning each question can be located. It would be very helpful to review the material related to any question which you answer incorrectly.

1. CPT. Jones was asked to rate the performance of Lt. New. Lt. New had some difficulty initially with his duties as Co. B XO. He was having personal family related problems and just couldn't seem to keep his mind on learning the new job. After a few weeks, however, Lt New solved his problems and settled in with vigor and efficiency. Eventhough Lt. New became a very effective XO, CPT. Jones continued to rate his performance as "below average." Which of the following best describes the rating behavior of CPT. Jones in this instance?

      a. accurate ratings
      b. comparison error
      c. "middle of the road" error
      d. first-impression error
      e. generosity error

2. MAJ. Smith evaluated the performance of a unit during a training exercise. When the results of the evaluation were being calculated to provide the unit with information concerning their training strengths and weaknesses, it was noticed that the highest rating MAJ. Smith gave was 2 on a 5-point scale. If we assume that this unit was actually of average ability on at least some of the rating dimensions, what kind of error had MAJ. Smith made?

      a. comparison error
      b. first-impression error
      c. similar-to-me error
      d. severity error
      e. no error

3. LTC. Brass was asked to rate and interview several applicants for a secretarial position which was coming open in the office. The third applicant that LTC. Brass interviewed was very poorly qualified for the job. Not only were her typing and shorthand skills very poor, but she was sloppily dressed, used incorrect grammer, and was rude as well. The next applicant that LTC. Brass interviewed had average skills and would ordinarily make an average personal impression. However, after the previous interview, this applicant appeared extremely good to LTC Brass and he rated her well above average. What error did LTC. Brass make?

      a. comparison error
      b. halo error

c. generosity error
d. similar-to-me error
e. no error

4. Nurse Nancy Needle was required to rate the performance of the nurse's aides who worked on her floor of the hospital. One of the aid's, Susie Sorrow, tried very hard, but just could not seem to learn to do the procedures correctly. Shortly after Susie began working for Nurse Needle, they discovered that they grew up in the same neighborhood and that Susie was taking classes at the school where Nurse Needle received her nursing diploma. Nurse Needle, therefore, felt a certain kinship with Susie and gave her an excellent rating in spite of Susie's poor performance. What error did Nurse Needle make?

a. first-impression error
b. halo effect
c. contrast effect
d. similar-to-me effect
e. no error

Answer Sheet

1. d   (first impression error) page 3

2. d   (severity error) page 4

3. a   (comparison error) pages 2-3

4. d   (similar to me effect) page 5

APPENDIX B

TRANSCRIBED SCRIPT FOR STIMULUS VIDEOTAPE

INTRODUCTION

S3: We will now brief the OPORDER. Will start with the S2, who will brief
paragraph 1. Then I will brief paragraph 2 - Mission, and paragraph 3 -
Execution. Then I will be followed by the Fire Support Officer, the S1 and the
S4 in that order.

Time is now 1507. First thing we are going to do is the weather. The weather is projected from 10 August to 25 August. The weather will be clear with temperatures ranging from 80-105 degrees farenheit. Surface winds will be from the south, velocity from three to five knots. Minimal cloudiness with no precipitation predicted. Full moon scheduled for 24 August, however, right now we have a zero moon, as far as percentage of illumination.

Light data - we are looking at a moon rise at 0916 in the morning and moon set at 2250. Sunrise at 0457, sunset 1846. EENT is 1944. Your BMNT is 0358. Your EECT 1912. Your BMCT is 0430 about an hour between twilight and sunrise - g·adual light build up.

The effects of weather on enemy course of action: Excellent visibility and weather will permit artillery resources to be used to the maximum, weather favors the use of TAC air by both sides.

Effect on friendly course of action: Clear weather favors the defense, enemy movement can easily be detected. TAC air and artillery can be used to the maximum. Continued dry weather will have no effect on already good traffic ability. Lack of sufficient rainfall and water sites will impact on CSS efforts to provide potable water.

TERRAIN - Cover and Concealment - Generally the overall area of operation we are looking at is at grid line 87. Grid line 87 to the Suez Canal (correction) terrain line slopes down to grid line 87 from the Suez Canal with varying formations of sand dunes. This entire area affords only a slight degree of cover and concealment, however, the high ground east of this north south grid line 87, and the sharply disecting and undulating sand dunes within this area, between this high ground, will provide cover and concealment for ground observations and direct fire weapons. And as you can see that is right in our task force area of operation.

OBSERVATION of Fires: Observation of fires is excellent from the high ground east of north/south grid line 87. The lack of relief west of grid line 87 generally permits flat trajectory fire from organic weapons at maximum range.

VEGETATION in Vicinity - Vegetation in this area is negligible.

OBSTACLES - Terrain will not present any obstacles to track or wheeled vehicle movement. Sand dunes and escarpments can be bypassed easily. The high ground we are occupying must be traversed at greatly reduced speeds if vehicles are to go up the terrain in this area. Vegetation in the area is not considered an obstacle.

Key terrain features: (I WILL NEED YOUR ATTENTION AT THIS MAP HERE). There are five. The MITLA pass which is key for division and brigade, and for our task force. We have four prominent pieces of high ground interspersed with waddies which stand out as key terrain. One is located here, second is a piece of high ground east of it going west/east, the third is located here. (POINTS) The fourth further on back in the rear area.

Avenues of Approach into our area: There are primarily three high speed avenues of approach - route 33, a dirt road from the central sector, and a paved route from the south. OPFOR is not obliged to stay on these roads.

Effect on enemy course of action, effect on enemy attack: Excellent weather and visability will allow use of artillery resources to the maximum.

Effect on enemy air: Weather favors use of TAC air by both forces.

Effect on enemy EW operations: Flat terrain favors his use of jamming to negate the effect of our communication and surveillance systems.

Effects on friendly courses of action: Best area for a friendly defense is in the high ground east of north south grid line 87 where we are projected to deploy to. Weather and terrain favor the defense in that any enemy movement can be observed. Continued dry weather will have no effect on already good traffic ability. TAC air and/or Army aviation can be exploited because of almost unlimited flying weather and visibility. There will be short periods of time when there will be blowing sand and dust storms in this area, but they are not to last for any extended length of time.

The enemy situation, composition: We are faced with elements of the 24th Combined Arms Army. They are broken down into the 46th motorized rifle division, the 49th motorized rifle division and the 26th tank division.

Committed forces within our area of operation is 9 motorized rifle battalions and seven tank battalions in our brigade sector. Judging from the OPFOR's doctrine, they will be deploying against our task force roughly in one regiment sized unit. This is based upon their doctrine of attacking on a 6-8 km front. (POINTS TO MAP) What you see here is the projection for expected OPFOR operations as it comes into our task force area. 4-8 km out a regimental sized unit will be breaking out into battalion columns. Coming in closer, 4-6 km out, in company columns, 2-3 km out, in platoon columns, and anywhere from a click to 300 meters they will be going into battle formation.

Reinforcements unknown at this time. Although it is possible an OPFOR regiment will be available to reinforce the regiment we are expecting in our AO.

RECENT AND PRESENT ACTIVITIES: There have been daily reconnaissance flights along the West Bank of the Suez. There has been an indication of troop movements, troop concentration, and equipment stockpiling within 50 km of the canal. Small groups of dissidents operating on the east side of the canal have been conducting sabotage operations. Few entities and weaknesses, the only descernable weaknesses of the enemy force his inability to move undetected from his staging area to the Suez Canal and across.

Electronic detection efforts will be capable of providing early warning of large scale movements.

Enemy capability: Numerations: attack with an estimated 9 motorized rifle battalions an 7 tank battalions in our brigade sector.

CONCLUSIONS: Enemy forces will conduct a coordinated attack within the next 12 to 48 hours by crossing the Suez Canal and moving into our area of operation. The enemy will probably commit elements of the 49th motorized rifle division in our task force sector/area.

VULNERABILITIES: Clear weather, flat terrain, will prevent undetected movement, also maximum utilization of TAC air and artillery. The enemy is vulnerable to friendly EW operations, to negate the effects of communication and surveillance.

RECONNAISSANCE and SURVEILLANCE: We will have four GSRs located in front. They will be able to detect OPFOR units moving into our AO. We will have 21 unattended ground sensors that we will be monitoring.

ANY QUESTIONS AT THIS TIME?

I'll cover our task organization.

Team Cold Steel - will be Co A, 2nd of the 120 minus 1 platoon. The platoon from A Co 1/2nd armor. Team 1 from Red-Eye, and 2 anti tank sections. 4 TOWS in addition to the 2 that you already have.

Co. B. You have team #2 from Red-Eye, you'll h ave 3 anti tank sections - 6 TOWS in addition to the 2 you already have.

Co. C. Minus your TOW section.

Team Tank - Your normal 2 tank platoons, platoon from Co A 2/120th team #3 of Redeye, 1 anti tank section from battalion anti-tank platoon, 1 anti tank section from Charlie Company.

Task Force Control will be the reconnaissance platoon. GSR - Redeye minus. Motor Platoon - 1st platoon, alpha battery 2/44th, a vulcan platoon and 2 platoons from A Co 52nd engineers.

Friendly Forces: 2nd Brigade defends in sector from Victor Uniform 900260 to Victor Uniform 865143. (SHOWS BRIGADE SECTOR ON MAP) not later than 14/2000 Aug and assists the rearward passage of the covering force of task force 8. Task force 1-2 armor is to our north and defends in this sector (POINTS TO MAP). 1/90th out of 4th division is to our south, which will be on our left. Task force 1-78 mech is to the north in 2nd Brigade's AO and they defend in this sector. I've already covered attachments and detachments.

Mission: Task Force 2-120 moves to and occupies defensive positions in sector from Victor Uniform 867143 to grid Victor Uniform 884203 not later than 14/2000 hours August, and assist rearward passage of task force 1-77 and task force 1-3. Again the mission. Task force 2-120 moves to and occupies defensive positions in sector from Victor Uniform 867143 to Victor Uniform 884203 not later than 14/2000 hours August and assist rearward passage of task force 1-77 and task force 1-3.

Execution: Concept of the operation.
Maneuver - task force 2-120 will desploy into the sector by air and land routes. 3 CH47s and 10 Hueys will be provided per company. Track and wheel vehicles will move along highway 33. Team Cold Steel personnel and Co B. personnel will move out of PZ1 which is at grid 080285 to LZ1 which is at grid 885170.

Charlie Co and Combat Support Company personnel will move from PZ 2 at 085275 to LZ 2 at 905170. For the track and wheeled vehicles the SP is at Whiskey Uniform 073266 up here on highway 33 (POINTS TO MAP). The RP is at Victor Uniform 913193. Be at the RP at 2000 hours. The task force will then desploy into battle positions with team Cold Steel to the north, initially in battle position X-ray. Co B in the center, battle (POINTS TO MAP) position Romeo. Team tank initially in battle position Yankee, and Co C in battle position

Delta. We will have 3 companys up. The tank heavy company is down to the south, infantry heavy team to the north, one company in reserve. Have all the TOWs forward as we are task organized. It works out to 8 TOWs - Bravo - 6 with Cold Steel, team tank you have 4. (All totaled will be 4).

Priority of Fires: Field artillery goes to Cold Steel correction - correction Team Tank. Field Artillery priority fires goes to Team Tank. Priority of battalion mortar fires goes to Company B.

Close Air Support: We are planning on 6 to 8 sorties per day.

Team Cold Steel: Established PZ control of PZ1. You are 2nd in order of movement from PZ1. Occupy blocking position X-ray. Be prepared to occupy positions, Oscar, Bravo, and Lima on order. I'll give you the center of mass grids for all these blocking positions back in the coordinting instructions, and also give you an overlay.

Company B - You will be LZ control for LZ1, you will have to provide an advance party that moves down there by wheeled vehicle. You are first in order of movement from PZ1. You will occupy blocking position Romeo and be prepared to occupy November, Sierra, and Pappa on order.

Company C - You are LZ control at LZ2. You will have to provide an advanced party of move down there in wheeled vehicles. You are first in order of movement from PZ2. Occupy blocking position Delta, and be prepared to occupy blocking position Hotel. Occupy Delta, be prepared to occupy Hotel.

Team Tank - You will occupy position Yankee. Be prepared to occupy positions Sierra, then Echo on order.

Combat Support Company - You'll be PZ control at PZ2. You are second in order of movement. You will furnish the alternate TOC.

Reconn Platoon - You will establish OPs forward of phase line yellow. Establish contact with the covering force. Link the covering force with guides from each of the companies, and as they pass through our friendly forward units they will report to the TOC - the TOC security for further missions. Recon again - establish OPs forward of phase line yellow, establish contact with the covering force which is task force 1-77 and elements of 1-3. Link the covering force with guides from the companies and upon passage of friendly forward units report to TOC for further missions.

Battalion mortars are in general support. Initial location is at 900148.

Redeye has a team attached to each of the forward elements. One team attached to the TOC. One team attached to the Combat Trains.

An anti-tank platoon has been attached all-out to the forward elements. All TOW sections are along the line. The Vulcan platoon located in the vicinity of the TOC, by TOC security. The 2 platoons of the engineers will establish 3 minefields in this priority. 1st will be down to the south here stretching from grid 868143 to 869148. 2nd priority will be a mine field here in the center, out in front of Bravo Company, 895187 to 903187. 3rd priority

will be up behind Cold Steel 868160 to 868164. Second mission will be to es-
tablish defensive positions to the forward units. The engineer platoon leader
will link up with the company commanders here with their earth moving vehicle.
It takes about 15 minutes per position using this equipment.

The GSR platoon - the ground surveillance platoon, will be working in
conjunction with reconn. Be out in front working for the S2.

Are there any questions at this time?

APPENDIX C

RATING SCALE

The OPORDER briefing presents the final phase in the military decision making process. During this time, individuals report their unique information to company commanders in a structured manner, and provide for the coordination of actions necessary to carry out the decision of a commander in the conduct of an operation.

The videotape you are about to see contains the S2 and S3 portions of a battalion level OPORDER briefing. The battalion is part of a light infantry division and is assigned the missions of defending in sector and assisting in passages of lines.

Commander's guidance to the S2 is as follows:

1. Carefully consider trafficability in the AO.
2. Identify key terrain.
3. Identify a minimum of three possible avenues of approach.
4. Develop the Order of Battle.
5. Plan for GSRs and ground sensors.

Commander's guidance to the S3 is as follows:

1. Task organize to maximize combat effectiveness include Red-eye, anti-tank and TOW.
2. Must be in defensive positions NLT 142000 Aug to assist passage of TF 1-77 and TF 1-3.
3. Consider movement by helicopter.
4. All TOWs will be forward.
5. Insure security and positive control for all PZ/LZ operations.
6. Make sure Recon PLT is used effectively.
7. Coordinate engineer support and insure integration.

You will be asked to rate the videotaped briefings on a number of dimensions which are generally considered to impact upon the quality of an OPORD briefing. Please read through the attached instructions and rating scales at this time. This will help you to be alert to the things you must look for while viewing the videotape so that you may make judgments asked for by the rating scales.

Instructions: <u>Please Circle</u> the number which best represents your opinion on each of the rating scale items as illustrated on the following example.

Example Item:

<u>Communication</u>: It is important that staff officers share information with each other as quickly as possible. Information passed on to other staff members should be accurate and disseminated as quickly as possible.

Please rate the extent to which the S2 disseminated information in a timely manner.

1	2	3	4	5	6	7
Never	Seldom	Occasionally	About as Often as Not	Frequently	Usually	Always

---

If your observations indicate that the S2 performed the desired behavior frequently and you wish to make a rating of 5, you would do so by <u>circling</u> the number 5 on the rating scale, as shown.

Briefer: S2

1. To what extent did the S2 use accepted military terminology and phraseology?

1	2	3	4	5	6	7
			About as			
Never	Seldom	Occasionally	Often as Not	Frequently	Usually	Always

2. To what extent did the S2 use language which was clear, precise, and unambiguous?

1	2	3	4	5	6	7
			About as			
Never	Seldom	Occasionally	Often as Not	Frequently	Usually	Always

3. For each of the following items, please rate how adequate the information provided was for successful mission execution.

    a. Light data.

1	2	3	4	5	6	7
Totally	Largely	Somewhat		Somewhat	Largely	Totally
Inadequate	Inadequate	Inadequate	Borderline	Adequate	Adequate	Adequate

    b. Weather conditions.

1	2	3	4	5	6	7
Totally	Largely	Somewhat		Somewhat	Largely	Totally
Inadequate	Inadequate	Inadequate	Borderline	Adequate	Adequate	Adequate

    c. Terrain information.

1	2	3	4	5	6	7
Totally	Largely	Somewhat		Somewhat	Largely	Totally
Inadequate	Inadequate	Inadequate	Borderline	Adequate	Adequate	Adequate

    d. Vegetation information.

1	2	3	4	5	6	7
Totally	Largely	Somewhat		Somewhat	Largely	Totally
Inadequate	Inadequate	Inadequate	Borderline	Adequate	Adequate	Adequate

    e. OPFOR composition.

1	2	3	4	5	6	7
Totally	Largely	Somewhat		Somewhat	Largely	Totally
Inadequate	Inadequate	Inadequate	Borderline	Adequate	Adequate	Adequate

4. Overall, how adequate was the level of detail of the S2 briefing?

1	2	3	4	5	6	7
Totally Inadequate	Largely Inadequate	Somewhat Inadequate	Borderline	Somewhat Adequate	Largely Adequate	Totally Adequate

5. Overall, how did the S2 relate the facts presented to the mission objectives where appropriate?

1	2	3	4	5	6	7
Totally Inadequate	Largely Inadequate	Somewhat Inadequate	Borderline	Somewhat Adequate	Largely Adequate	Totally Adequate

6. How adequately did the S2 follow the commander's guidance in the following areas?

    a. Carefully consider trafficability in the area of operation?

1	2	3	4	5	6	7
Totally Inadequate	Largely Inadequate	Somewhat Inadequate	Borderline	Somewhat Adequate	Largely Adequate	Totally Adequate

    b. Identify key terrain features?

1	2	3	4	5	6	7
Totally Inadequate	Largely Inadequate	Somewhat Inadequate	Borderline	Somewhat Adequate	Largely Adequate	Totally Adequate

    c. Identify a minimum of three possible avenues of approach?

1	2	3	4	5	6	7
Totally Inadequate	Largely Inadequate	Somewhat Inadequate	Borderline	Somewhat Adequate	Largely Adequate	Totally Adequate

    d. Develop the Order of Battle?

1	2	3	4	5	6	7
Totally Inadequate	Largely Inadequate	Somewhat Inadequate	Borderline	Adequate	Largely Adequate	Totally Adequate

    e. Plan for GSRs and ground sensors?

1	2	3	4	5	6	7
Totally Inadequate	Largely Inadequate	Somewhat Inadequate	Borderline	Somewhat Adequate	Largely Adequate	Totally Adequate

7. How adequately did the S2 organize the briefing so that listeners could follow the change of topics?

1	2	3	4	5	6	7
Totally Inadequate	Largely Inadequate	Somewhat Inadequate	Borderline	Somewhat Adequate	Largely Adequate	Totally Adequate

8. How adequately did the S2 pace the tempo of the briefing to allow listeners time to take notes if desired?

1	2	3	4	5	6	7
Totally Inadequate	Largely Inadequate	Somewhat Inadequate	Borderline	Somewhat Adequate	Largely Adequate	Totally Adequate

9. How adequately did the S2 use maps and/or visual aids to insure that the content of the briefing was understood?

1	2	3	4	5	6	7
Totally Inadequate	Largely Inadequate	Somewhat Inadequate	Borderline	Somewhat Adequate	Largely Adequate	Totally Adequate

Briefer: S3

1. To what extent did the S3 use accepted military terminology and phraseology?

1	2	3	4	5	6	7
			About as			
Never	Seldom	Occasionally	Often as Not	Frequently	Usually	Always

2. To what extent did the S3 use language which was clear, precise, and unambiguous?

1	2	3	4	5	6	7
			About as			
Never	Seldom	Occasionally	Often as Not	Frequently	Usually	Always

3. For each of the following items, please rate how adequate the information provided was for successful mission execution.

    a. Task organization.

1	2	3	4	5	6	7
Totally	Largely	Somewhat		Somewhat	Largely	Totally
Inadequate	Inadequate	Inadequate	Borderline	Adequate	Adequate	Adequate

    b. Statement of the mission.

1	2	3	4	5	6	7
Totally	Largely	Somewhat		Somewhat	Largely	Totally
Inadequate	Inadequate	Inadequate	Borderline	Adequate	Adequate	Adequate

    c. Overall concept of the operation.

1	2	3	4	5	6	7
Totally	Largely	Somewhat		Somewhat	Largely	Totally
Inadequate	Inadequate	Inadequate	Borderline	Adequate	Adequate	Adequate

    d. Priority of fires.

1	2	3	4	5	6	7
Totally	Largely	Somewhat		Somewhat	Largely	Totally
Inadequate	Inadequate	Inadequate	Borderline	Adequate	Adequate	Adequate

    e. Close air support.

1	2	3	4	5	6	7
Totally	Largely	Somewhat		Somewhat	Largely	Totally
Inadequate	Inadequate	Inadequate	Borderline	Adequate	Adequate	Adequate

4. Overall, how adequate was the level of detail of the S3 briefing?

1	2	3	4	5	6	7
Totally Inadequate	Largely Inadequate	Somewhat Inadequate	Borderline	Somewhat Adequate	Largely Adequate	Totally Adequate

5. Overall, how adequate did the S3 relate the facts presented to the mission objectives where appropriate?

1	2	3	4	5	6	7
Totally Inadequate	Largely Inadequate	Somewhat Inadequate	Borderline	Somewhat Adequate	Largely Adequate	Totally Adequate

6. How adequately did the S3 follow the commander's guidance in the following areas:

   a. Location of TOWs?

1	2	3	4	5	6	7
Totally Inadequate	Largely Inadequate	Somewhat Inadequate	Borderline	Somewhat Adequate	Largely Adequate	Totally Adequate

   b. Effective use of the Recon PLT?

1	2	3	4	5	6	7
Totally Inadequate	Largely Inadequate	Somewhat Inadequate	Borderline	Somewhat Adequate	Largely Adequate	Totally Adequate

   c. Insure security and positive control for all PZ/LZ operations?

1	2	3	4	5	6	7
Totally Inadequate	Largely Inadequate	Somewhat Inadequate	Borderline	Somewhat Adequate	Largely Adequate	Totally Adequate

   d. Insure integration of engineer support?

1	2	3	4	5	6	7
Totally Inadequate	Largely Inadequate	Somewhat Inadequate	Borderline	Somewhat Adequate	Largely Adequate	Totally Adequate

7. How adequately did the S3 organize the briefing so that listeners could follow the change of topics?

1	2	3	4	5	6	7
Totally Inadequate	Largely Inadequate	Somewhat Inadequate	Borderline	Somewhat Adequate	Largely Adequate	Totally Adequate

8. How adequately did the S3 pace the tempo of the briefing to allow listeners time to take notes if desired?

1	2	3	4	5	6	7
Totally Inadequate	Largely Inadequate	Somewhat Inadequate	Borderline	Somewhat Adequate	Largely Adequate	Totally Adequate

9. How adequately did the S3 use maps and/or visual aids to insure that the content of the briefing was understood?

1	2	3	4	5	6	7
Totally Inadequate	Largely Inadequate	Somewhat Inadequate	Borderline	Somewhat Adequate	Largely Adequate	Totally Adequate

APPENDIX D

QUESTIONNAIRE

Questionnaire

Please answer the following questions in the space provided. Confidentiality of all responses is guaranteed by the researchers.

1. Name _____

2. Rank _____

3. Check the highest level of civilian education completed.

    _____ High school diploma or equivalent
    _____ Associate degree
    _____ Bachelors degree
    _____ Masters degree
    _____ Ph.D.

4. Check all of the following military education programs which you have attended.

    _____ Officers Basic Course
    _____ Officers Advanced Courses
    _____ Combined Arms & Services Staff School (CAS^3)
    _____ CGSC
    _____ Army War College

5. Have you had any prior training in rating, performance measurement, appraisal, or evaluation methods?

              _____ Yes   _____ No

6. Other than OER or EER evaluations, have you had any prior experience as a performance evaluator? _____ Yes _____ No

    a. If so, what was the nature of the evaluation(s) you performed?

    _____

    _____

    b. On how many occasions have you performed evaluations? _____

7. Number of years in the military? _____

8. Have you ever held a position as a battalion staff officer? ___ Yes ___ No

    a. If so, which position(s) did you hold? (S1, S2, FSO, etc) _____

    b. Total number of years experience as a battalion staff officer _____

9. Please rate the effect you feel that the rater training material had on the accuracy of the ratings you were asked to make.

1	2	3	4	5
Extremely Harmful	Somewhat Harmful	Neutral	Somewhat Helpful	Extremely Helpful

Give any suggestions you may have concerning ways in which the rater training material might be improved.

_____

_____

_____

APPENDIX E

ANAGRAM TASK

Instructions

On the attached sheet you will find a list of anagrams (scrambled words). Your task is to rearrange the letters of the anagrams to spell a word, as in the following example.

Example:

Anagram	Solution
TUSDY	STUDY

You will have 20 minutes to complete as many anagrams as possible.

Anagram	Solution
ENKLA	
GRVAE	
RUTCK	
RAIFY	
UNCHL	
ATONB	
BNUCH	
WUNOD	
CRWDO	
NDBRA	
LEFAM	
VLORE	
TOGIB	
IGLUT	
AELBD	
TIWHD	
BIROT	
PMUIO	
NHACR	
RFCTA	

If you complete this task before time limit, please wait for further

APPENDIX F

MULTIVARIATE ANALYSIS OF VARIANCE TABLES

Multivariate Analysis of Variance
Summaries

Analysis 1 - S2 Factors

Source	Wilks Lambda	S, M, N	Approx. F.	Significance
Group	.934	1, 1, 7 1/2	.297	.88

Analysis 2 - S3 Factors

Source	Wilks Lambda	S, M, N	Approx. F.	Significance
Group	.959	1, 1, 7 1/2	.181	.95

F-1

APPENDIX G

MEANS AND STANDARD DEVIATIONS

OF RATINGS ON INDIVIDUAL ITEMS

Means and Standard Deviations of Ratings on Individual Items

Form	Item #	Factor	Group	Mean	SD
S2	1	Clarity	treatment	5.64	1.03
			control	5.36	1.21
	2		treatment	4.82	1.33
			control	4.64	1.36
	3	Adequacy of	treatment	5.82	1.08
		Information	control	5.00	1.55
	4		treatment	5.55	.69
			control	5.55	1.13
	5		treatment	5.36	1.03
			control	4.73	1.35
	6		treatment	5.18	1.78
			control	5.27	1.10
	7		treatment	5.27	1.35
			control	4.18	1.66
	8		treatment	5.00	1.41
			control	4.73	1.27
	9		treatment	4.27	1.49
			control	4.09	1.30
	10		treatment	4.73	1.56
			control	4.09	1.30
	11	Following	treatment	5.00	1.10
		Commander's Guidance	control	4.45	1.64
	12		treatment	3.64	2.25
			control	3.27	1.62
	13		treatment	4.91	1.14
			control	4.27	1.35
	14	Briefing	treatment	4.45	1.21
		Delivery	control	4.09	1.76
	15		treatment	4.00	2.00
			control	3.27	1.49
	16		treatment	3.45	1.73
			control	2.73	1.56
	17		treatment	2.82	1.33
			control	2.27	1.56

Form	Item #	Factor	Group	Mean	SD
S3	1	Clarity	treatment	5.91	.70
			control	5.64	.81
	2		treatment	5.00	.89
			control	5.09	1.22
	3	Adequacy of	treatment	5.55	1.04
		Information	control	4.36	1.63
	4		treatment	5.45	1.75
			control	5.73	.65
	5		treatment	3.91	1.58
			control	4.27	1.68
	6		treatment	4.73	1.56
			control	5.18	.75
	7		treatment	3.82	1.99
			control	4.36	1.50
	8		treatment	4.45	1.37
			control	4.73	1.27
	9		treatment	3.91	1.58
			control	4.36	1.43
	10		treatment	5.27	1.42
			control	5.00	1.61
	11	Following	treatment	5.09	1.38
		Commander's Guidance	control	4.55	1.75
	12		treatment	4.80	1.14
			control	4.55	1.51
	13		treatment	4.64	1.57
			control	4.18	1.60
	14	Briefing	treatment	4.00	1.61
		Delivery	control	3.73	1.56
	15		treatment	3.64	1.69
			control	3.00	1.95
	16		treatment	3.00	1.41
			control	2.36	1.57