(12) LEVEL

(9) Technical Paper 342

AD

(10) ARI-TP-342

(6)

REVIEW OF PEER EVALUATION RESEARCH

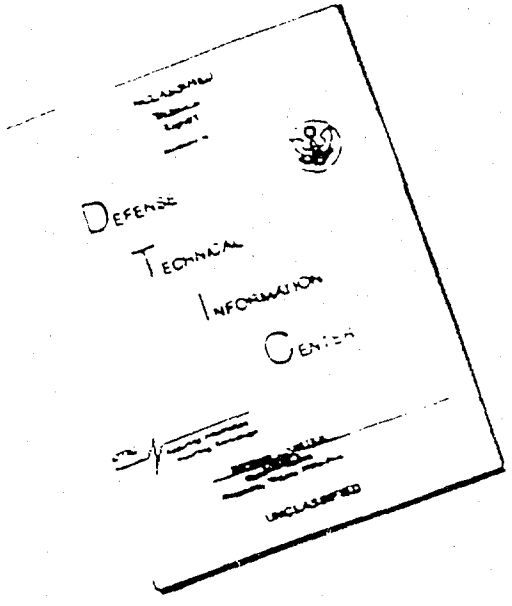
(10) Ronald G. Downey and Paul J. Duffy

PERSONNEL AND MANPOWER TECHNICAL AREA

U. S. Army
Research Institute for the Behavioral and Social Sciences

(11) Oct[illegible]978

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

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

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

JOSEPH ZEIDNER
Technical Director

WILLIAM L. HAUSER
Colonel, US Army
Commander

NOTICES

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

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

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

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Technical Paper 342	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): REVIEW OF PEER EVALUATION RESEARCH	5. TYPE OF REPORT & PERIOD COVERED
6. PERFORMING ORG. REPORT NUMBER	
7. AUTHOR(s): Ronald G. Downey and Paul J. Duffy	8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS: U.S. Army Research Institute for the Behavioral and Social Sciences (PERI-IL), 5001 Eisenhower Avenue, Alexandria, Va. 22333	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 2Q162717A766
11. CONTROLLING OFFICE NAME AND ADDRESS: Army Deputy Chief of Staff for Personnel, Washington, DC 20310	12. REPORT DATE: October 1978
	13. NUMBER OF PAGES: 28
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): --	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

--

18. SUPPLEMENTARY NOTES

--

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

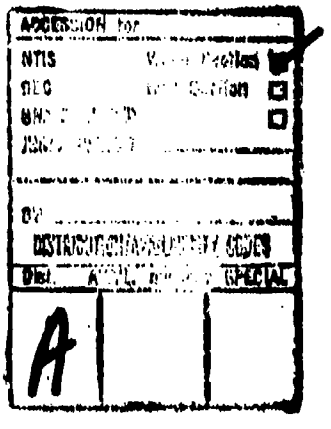
Peer evaluation
Peer rating
Validity
Reliability
Situational factors

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

Peer evaluation research was reviewed from the three major perspectives of validity studies, methodology, and situational factors. Most of the research programs were conducted in the course of developing procedures for evaluating training groups (e.g., in Officer Candidate School, U.S. Military Academy, and Ranger course). Substantial concurrent and predictive validity generally was found, with correlation coefficients in the .30 to .50 range. Different evaluation methods (rating, ranking, nominations, and combinations

20.

of these techniques) did not differ substantially in either reliability or validity. Evaluation methods did, however, differ in acceptability and feasibility. Situational factors have documented or potential effects on the evaluation process that developers and users of peer evaluations should be aware of. Although many issues surrounding peer evaluations remain unresolved, evidence suggests that these issues can be resolved, and that they do not detract from the conclusion that peer evaluations are a powerful tool in discriminating complex human behavior.



Technical Paper 342

REVIEW OF PEER EVALUATION RESEARCH

Ronald G. Downey and Paul J. Duffy

PERSONNEL AND MANPOWER TECHNICAL AREA

Submitted as complete and technically accurate, by:
Ralph R. Canter
Technical Area Chief

Approved By:

E. Ralph Dusek, Director
PERSONNEL AND TRAINING RESEARCH LABORATORY

Joseph Zeidner
TECHNICAL DIRECTOR

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES
5001 Eisenhower Avenue, Alexandria, Virginia 22333

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

October 1978

Army Project Number
2Q162717A766

Officer Careers

ARI Research Reports and Technical Papers are intended for sponsors of R&D tasks and other research and military agencies. Any findings ready for implementation at the time of publication are presented in the latter part of the Brief. Upon completion of a major phase of the task, formal recommendations for official action normally are conveyed to appropriate military agencies by briefing or Disposition Form.

FOREWORD

This research, carried out within the Personnel Accession and Utilization Technical Area of the Army Research Institute (ARI), includes a representative review of previous findings, both within the Army and otherwise, on the validity and reliability of peer evaluations. The research also reviews several situational or contextual factors that should be considered in conducting peer evaluations.

This research is an in-house effort and is responsive to Army Project 2Q162717A766 and to special requirements of the Office of Deputy Chief of Staff for Personnel.

JOSEPH ZEIDNER
Technical Director

REVIEW OF PEER EVALUATION RESEARCH

BRIEF

Requirement:

To review previous findings on the validity and reliability of peer evaluations as well as various situational moderators.

Procedure:

Peer evaluation research was reviewed from the four major perspectives of evaluation process, methodology, situational factors, and validity studies.

Findings:

Studies investigating the structure and nature of the peer evaluation process have generally found fairly clear factor structure across widely varying samples. There is some evidence that the structure may be as much in the nature of the rater as the ratee. A review of findings from research that utilized different methods indicated little evidence for substantial differences, in either reliability or validity, among techniques. Further, a review of the documented and potential effects of situational factors impacting on the evaluation process indicated that users of peer evaluation should be aware of these issues in designing programs. Research generally has found substantial concurrent and predictive validity, with correlations in the .30 to .50 range, but with most studies limited to training groups.

Utilization of Findings:

Several issues surrounding peer evaluations remain unresolved; however, evidence suggests that these issues can be resolved, and that peer evaluations are a powerful tool in discriminating complex human behavior.

REVIEW OF PEER EVALUATION RESEARCH

CONTENTS

LIST OF TABLES

LIST OF FIGURES

REVIEW OF PEER EVALUATION RESEARCH

INTRODUCTION

When confronted with the prospect of drawing order out of complex human behavior in the equally complex world of work, much traditional behavioral science research has been marked by two primary characteristics. First, heavy reliance has been placed upon human evaluations of other human beings. Second, this evaluative information has been typically gathered from a limited observational viewpoint, that of a superior toward a subordinate. The technique presented in this paper does not deviate from the first of these characteristics; it does rely on human evaluation of other human beings. However, it goes beyond the second characteristic by gathering evaluative information from the perspective of an individual's peers. For purposes of this paper, peers are operationally defined thus: (a) they have some common purpose or frame of reference (e.g., members of the same work group), and (b) generally speaking, they lack a formally recognized authority relationship between them. Although the term "peer rating" is most commonly applied to this technique, the present paper uses the more generic term "evaluation," reserving the term "rating" for one particular technique.

A source of much confusion in peer evaluation research has been a lack of clarity between the technique and the dimension or characteristic evaluated. Although previous work reviewed here substantially supports use of peer evaluation as a technique, issues surrounding the particular dimensions evaluated are not discussed in this review.

This paper contains three relatively complementary sections. First, a representative selection of typical validity research is reviewed, along with a brief history of the use of peer evaluations. The second section discusses various methodological issues underlying the peer evaluation technique, and the third section presents several situational or contextual factors that can affect a peer evaluation effort.

VALIDITY OF PEER EVALUATIONS

The history of the peer evaluation technique can be traced from the seminal work of Moreno (1934) and the development of the sociogram technique. However, the history of the technique as it is dealt with here is more conveniently traced to several efforts conducted during and after World War II (see, for example Clarke, 1946; U.S. Army Research Institute, 1943; Wherry, 1945). One of the earliest investigations published in the professional literature is that by Williams and Leavitt (1947).

Since that time, peer evaluations have been used for two primary purposes. The first of these purposes is evaluative in the criterion sense: The concern is in judging the extent or adequacy of some individual characteristic (e.g., leadership effectiveness, job performance). The second purpose is evaluative in the sense of gaining information with which to predict some future outcome (individual potential, motivation to work, etc.). Both purposes have guided the efforts in research as well as operational settings, although typically only one purpose has been the focus in any given situation.

Table 1 summarizes the results and major characteristics of a representative sampling of studies which report validity information for peer evaluations. This overview is intentionally not exhaustive, since several other more specialized reviews are available elsewhere (e.g., Gibb, 1969; Hollander, 1954a; Boulger & Coleman, 1964; & Nadal, 1968). Lindzey and Byrne (1968) have also presented an excellent review of the use of social choice methodology of which peer evaluations are one type.

There are several noteworthy features in Table 1. First, the magnitude of the validity coefficients is generally strong in both concurrent and predictive studies. Peer evaluations have shown rather strong predictive ability even for periods up to 5 years (Hollander, 1965). Furthermore, in those studies that included measures in addition to peer evaluations, the peer evaluations tended to have the highest concurrent or predictive validity.

Also, the majority of the evidence for the value of peer evaluations has been gathered in a training situation, particularly in the military environment. In fact, only two of the studies in Table 1 (Weitz, 1958; Downey, Medland, & Yates, 1976) used a sample from other than a training or educational environment. With a few exceptions, most evidence has been gained from people relatively low in the hierarchy of their organizational setting.

A third major feature of Table 1 is the variety of dimensions that peers have been required to evaluate and the variety of criteria with which peer evaluations have been related. The peer evaluation dimensions have included leadership potential, personality traits, and supervisory skill, to name but a few.

Table 1

Some Representative Studies on the Validity of Peer Evaluations

Investigators	Type of subject	Dimensions evaluated	Criteria	Correlation
Amir, Kovarsky, & Sharan (1970)	Enlisted military trainees	Promotion potential	Promotion to NCO[a]	.44** (1,979)[b]
	NCO trainees	Promotion potential	Promotion to officer[a]	.63** (1,918)
Berkshire & Nelson (1958)	Military officer trainees	Promotion potential	Graduation[c] Performance[a]	--[d] (1,152) --[d] (1,152)
Butler (1974)	West Point trainees	Leadership	Performance[a] Promotion[a]	.38** (547) .24** (547)
Doll (1963)	Military officer trainees	Promising officers	Pass/fail[c]	.20** (606)
	Military cadets	Promising cadets	Pass/fail[c]	.36** (660)
Downey (1973)	Senior military officer trainees	Promotion potential	Promotion[a]	--[d]** (246)
Downey, Medland, & Yates (1976)	Senior military officers	Promotion potential	Promotion[a]	.53** (242)
Haggerty (1963)	West Point trainees	Leadership traits	Performance[a]	.38** (120)
		Leadership traits	Performance[a]	.26** (253)
Hollander (1954[b])	Military officer trainees	Leadership	Graduation[c]	.27** (268)

Table 1 (continued)

Investigators	Type of subject	Dimensions evaluated	Criteria	Correlation
Hollander (1965)	Military officer trainees	Leadership	Grades[c]	.51** (229)
			Performance[a]	.37** (229)
Klieger, deJung, & Dubuisson (1962)	Enlisted military trainees	Performance potential	Discharge[a]	.42** (1,571)
Kraut (1975)	Manager trainees	Impact--10 scales	Promotion[a]	.31** (82)
		Tactfulness--3 scales	Promotion[a]	.02 (82)
	Executive trainees	Impact--10 scales	Performance[a]	.35** (83)
		Tactfulness--3 scales	Performance[a]	.37** (83)
Kubany (1957)	Medical students	Medical performance potential	Instructor evaluations[c]	.48** (87)
Levi, Torrance, & Pletts (1958)	Enlisted military trainees	13 dimensions of personality & potential	Dropout rate[c]	--[d]** (770)
			Performance[a]	--[d]** (770)
Peterson, Lane, & Ambler (1966)	Military officer trainees	Carefulness	Pass/fail[c]	.22** (462)
Ricciuti (1955)	Military officer trainees	Leadership	Performance as midshipmen[c]	.32** (324)
			Performance training cruise[c]	.26** (324)
Roadman (1964)	Management trainees	13 dimensions of personality, achievement, & leadership	Promotion[a]	--[d]** (56)

Table 1 (continued)

Investigators	Type of subject	Dimensions evaluated	Criteria	Correlation
Smith (1967)	College students	Extraversion	GPA[c]	.05 (348)
		Strength of character	GPA[c]	.43** (348)
Tupes (1957)	Military officer trainees	Composite of 30 personality factors	Performance[a]	.51** (615)
			Grades[c]	.31** (615)
Waters & Waters (1970)	Sales trainees	Agreeable	Performance[a]	-.27* (53)
		Sales potential	Performance[a]	.31* (53)
Weitz (1958)	Salesmen	Promotion potential	Performance[a]	.40** (100)
Wherry & Fryer (1949)	Military officer	Leadership	Retention[c]	.70** (134)
			Graduation[c]	.49** (.34)
Wiggins, Blackburn, & Hackman (1969)	College graduate students	Academic success	GPA[c]	.56** (46)
		Academic success	GPA[c]	.69** (58)
Williams & Leavitt (1947)	Military officer trainees	Future potential	Performance[a]	.47** (100)
Willingham (1958)	Military officer trainees	17 leadership traits	Pass/fail[c]	.28** (994)

[a]Predictive criterion.
[b]Numbers in parentheses are number of subjects.
[c]Concurrent criterion.
[d]Significant group differences found.
*p < .05.
**p < .01.

Attempts to implement peer evaluation programs have produced an impressive array of findings. However, several limitations also appear. For instance, there is only minimal evidence of the validity of peer evaluations among individuals at organizationally higher levels. There is also a limited, but growing, amount of evidence of the utility of peer evaluations in other than the training environment. In addition, in studies that use peer evaluations as a predictor of a concurrent or future criterion, virtually all the validity evidence is of a bivariate variety. Although a number of studies demonstrated that peer evaluations are often the best single predictor from among several predictors, no research was found that attempted to determine what other predictors might account for unique variance along with peer evaluations. An exception to this preoccupation with the bivariate paradigm is occasionally found in assessment center methodology. Mackinnon (1975) has elsewhere presented a comprehensive review of assessment centers, but even in assessment centers with a wealth of information available, the differential validity of peer evaluations has not always been adequately addressed.

METHODOLOGICAL ISSUES

Peer evaluations have been performed by means of four primary techniques: ratings, rankings, full nominations, and high nominations. The general paradigm of the rating technique calls for a group member to provide a rating of the relative amount or degree of the dimension under consideration possessed by every other group member. The ranking procedure simply requires each group member to rank-order all other group members from high to low (or some other relevant continuum) on the dimension under consideration. The full nomination technique requires that each group member choose a specified number or proportion of the group as being either high, medium, or low on a given dimension. The minor variation of this technique in which nominations of the middle are not required is also referred to as full nominations. However, the case in which only high nominations are elicited is reserved as a discriminably different technique, for reasons to be elaborated upon in later portions of the paper.

Several variations based on combinations of these basic techniques are forced distribution rankings, or combinations of rankings with ratings. General scoring algorithms for the four primary techniques follow.

Ratings:

$$\text{Score} = \frac{\Sigma^{r}_{Rt}}{N} .$$

Rankings:

$$\text{Score} = \frac{\Sigma^{r}_{Rk}}{N} \times \frac{100}{N_T} .$$

Full Nominations:

$$\text{Score} = \frac{\Sigma(r_L) + \Sigma(2r_M) + \Sigma(3r_H)}{N} .$$

High Nominations:

$$\text{Score} = \frac{\Sigma r_H}{N} ,$$

where r_{Rt} = rating,

r_{Rk} = ranking,

r_L = low nomination,

r_M = mid (or no) nomination,

r_H = high nomination,

N = number giving an evaluation, and

N_T = total number in the group.

All these techniques produce scores with means independent of group size, with the exception of the ranking formula, in which case adjustment must be made for group sizes greater than 100. The standard deviation of the various scores is a function of the reliability (consistency) of each group's evaluations; Gordon (1969) and Willingham (1959) deal with general issues related to reliability. Also, for a group using either a ranking or nomination technique, the average score is determined; the average score using the rating technique is free to vary.

Metric and Distribution

The metric and distributional properties of associate evaluations are directly related to the particular technique employed. With respect to scaling properties, the rankings and both nomination procedures produce an ordinal scale (Stevens, 1951). The ratings from an evaluator are the most nearly equal interval data, although here also it can be argued that these are merely an ordinal scale. The scaling properties of the summated scores from the various techniques approximate interval data as the number in the evaluation group increases.

The four common procedures will generally produce different distributions, examples of which are displayed in Figure 1. Given the relatively free response mode, ratings will often produce negatively skewed distributions largely because group norms tend to inflate any evaluative procedure. The ranking procedure, if it were perfectly reliable, would produce a rectangular distribution with one person at each rank. Generally, less than perfectly reliable rank scores will tend to be normally distributed, with very unreliable scores producing a more leptokurtic curve, and a perfectly unreliable procedure producing a point distribution with everyone receiving an average rank equal to the middle rank. Full nomination scores produce a distribution which, if perfectly reliable, is trimodal, with one group receiving all high nominations, another group all low nominations, and the remainder middle nominations or none at all. High nominations produce a bimodal distribution (not shown in Figure 1).

Basis of Comparison

Scores resulting from the four primary techniques vary along another important dimension--the evaluative process evoked in the evaluator upon which judgments are made. Drucker (1957) initially pointed out the duality of focus with which peer evaluations can be executed: whether the frame of reference or standard upon which the evaluations are made is internal or external to the group. In one case, the evaluator compares the particular individual against a frame of reference external to the group and assigns the individual to a category. In the second case, the evaluator compares the particular individual against a frame of reference internal to the group and makes a judgment of more or less, and assigns the individual to the appropriate category. The external process can be used only with the rating procedure. The internal process can also be used with ratings; with rankings and nominations, it is required. The internal process, in general, requires a moderate number of individuals in the group (more than five). The direct implication of this distinction is that the external frame of reference allows both comparison between individuals across peer groups and the comparison of peer groups. The internal process does not allow comparison between individuals across peer groups unless the assumption is accepted that the groups are equal on the particular ability, trait, or behavior.

A corollary of this implication is that population norms can be developed only through the use of a rating procedure and an external frame of reference, again unless group equality is assumed or assured.

Reliability

The reliability of associate evaluations has generally been determined by one of two methods, estimation of internal consistency or test-retest correlation. Both methods are analogous to the same procedures in classical test theory (Lord & Novick, 1968).

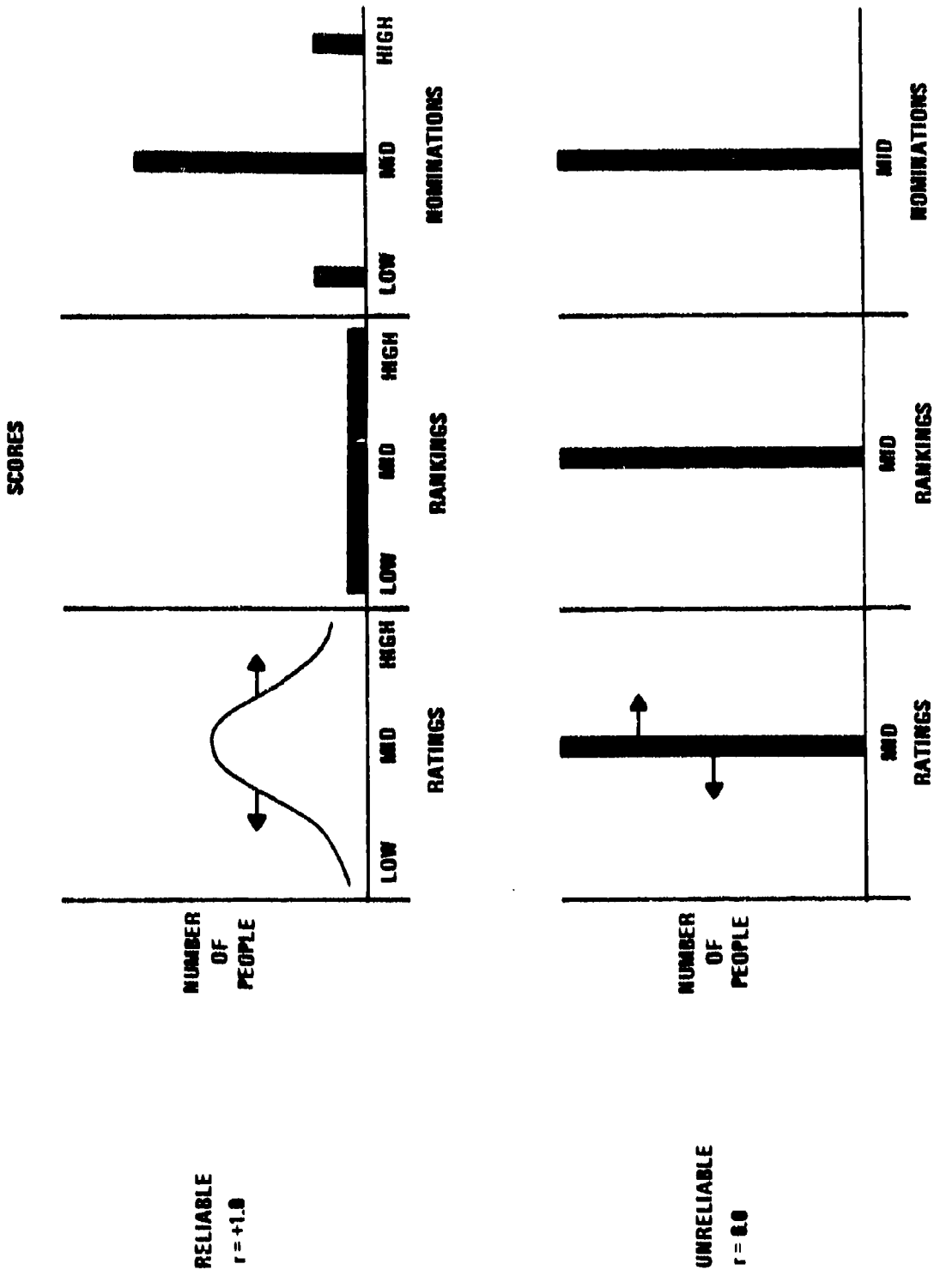


Figure 1. Score distributions for reliable and unreliable evaluations.

The internal consistency of peer evaluations is the degree to which members of a peer group agree with one another when observing an individual in a similar situation and at the same time. Using the multiple-choice test paradigm, the evaluators are comparable to test items and those who are being evaluated are comparable to persons taking the test. Although Gordon (1969) has recommended the use of the alpha coefficient for estimating the internal consistency or reliability of peer evaluations, the most common procedure has been a split-half (or group) estimate. The split-half estimate is made by randomly assigning peer group members to one of two groups, computing scores in each group for all group members, and then correlating the scores for each ratee from each group (see Hollander, 1957, & Downey, 1974). The correlation coefficient is then adjusted for total group size using the Spearman-Brown formula (Gulliksen, 1950). If small groups are used, a random split may not be possible, and some technique for averaging the intercorrelations between evaluators could be used (Gulliksen, 1950).

The test-retest method of estimating reliability requires that group members evaluate each other at two different times. Scores from the two different evaluations are then correlated. Examples of this type of estimate are given in Hollander (1957) and Downey (1974, 1976). Perhaps the most rigorous examination of reliability was done by Gordon and Medland (1965), in which they varied both time of administration and group doing the evaluations and found reliability coefficients in the 80's.

Research has generally demonstrated the reliability of peer evaluations to be in the .70 to .90 range, regardless of the type of reliability estimate employed. Research comparing the various evaluative methodologies is rare but has generally supported the view that all four methods are quite similar, with perhaps a slight advantage to ratings (Suci, Vallance, & Glickman, 1954; Downey, 1974; Hammer, 1963). Even the use of a paired comparison procedure does not significantly improve reliability (Bolton, 1971).

Acceptability

A major factor in the success or failure of any peer evaluation procedure, whether for operational or research purposes, is the degree to which participants accept the purpose of the evaluations. Acceptability is generally studied as a specific issue of the particular program under investigation rather than comparative analyses of acceptability across techniques or situations. There is therefore little formal evidence of differences between techniques in this respect, but inferences can be drawn from the particular qualities of the technique.

A major factor in the acceptability of a technique is the degree of perceived difficulty. From this point of view, both the rating and ranking of large numbers of individuals (more than 20) can be time-consuming and makes for difficult discriminations, particularly among group members who are more or less average on the particular dimension. On the other hand, the nomination procedure allows the individual to place a large number of people in a desired category and does not require such difficult discriminations.

The rating procedure is quite acceptable to the raters where the rated group is small and cohesive. The full nomination technique is acceptable to the nominators for moderate-size to large groups in which not all individuals are well known to one another. The high nomination technique is even more acceptable because it does not require an individual to make negative evaluations.

Another determinant of the degree of acceptability is the degree to which group members are knowledgeable about the evaluation procedure, process, background, and use. Downey (1975) found that acceptability improved as a function of an educational program. Two different considerations were noted: (a) the degree to which peer evaluations were felt to be valuable and accurate estimates and (b) the degree to which the evaluations were acceptable for particular uses. Downey also found that a person's peer evaluation score and degree of acceptance of the peer evaluation process were positively correlated; larger correlations were found in the group who knew less about the peer evaluation process.

Feasibility

Closely linked with the concept of acceptability is feasibility, or costs associated with the implementation and execution of a particular peer evaluation system. The major costs associated with a peer evaluation system are (a) preparation of evaluation materials, (b) administration time, and (c) scoring cost. Prior to the advent of automatic data processing procedures, the costs associated with use of any peer evaluation system in large groups or on a large scale were prohibitive. Merely in terms of bits of information collected, it can be seen that the number of evaluations is typically equal to n (n - 1) where n is the number in the group. Thus, peer evaluation systems are relatively costly efforts, which typically require more than minimal sophistication with data processing procedures. Unfortunately, little systematic information on cost is available.

SITUATIONAL FACTORS

In addition to the methodological concerns of the various techniques, several situational or contextual factors can affect a peer evaluation system, often without regard to the specific technique under discussion. These factors include group size, informal group structures, demographic

characteristics, group boundaries, hierarchical characteristics, friendships, length of association, and types of interaction.

Group Size

Very few attempts have been made to study the independent effects of group size. More often than not, what evidence there is has been reported as a byproduct in research directed elsewhere. For example, Downey, Medland, and Yates (1976) used a peer nomination technique with groups of Army colonels in 14 career groups that varied in size from 22 to 321. Reliability coefficients varied from .63 to .94 and the rank order coefficient between group size and reliability was .03. Downey (1976), in a sample of Army Rangers, compared peer ratings collected within squads ($n \cong 10$) with peer nominations collected on the same men within platoons ($n \cong 40$). Coefficients between the two scores were in the .60's. However, platoon scores were both more reliable and more predictive of job performance.

As mentioned previously, from the standpoint of feasibility both ratings and rankings would seem to be most appropriate for relatively small group sizes (approximately a dozen), whereas the nomination technique is virtually mandatory for large groups (more than 50). From the standpoint of empirical results, it appears that small groups may produce somewhat unreliable scores, with reduced validity. Alternatively, although it is rational to believe that there is an optimal upper size peer group, scant evidence exists to support this view.

Informal Group Structures

Within any formally defined group, there may exist one or more informal subgroups defined by some sort of mutual self-interest. The issue then arises as to the effect these informal subgroups may have on a peer evaluation procedure conducted in the total group.

The worst case would be one in which two equal-sized informal subgroups existed within a total group, and each group member was exclusively in one subgroup or the other. In such a situation, one or both subgroups might make their evaluations solely on the basis of subgroup membership, i.e., on a basis other than the one intended. The net effect of such behavior is to attenuate the validity of the peer evaluation procedure; attenuation is most pronounced when both subgroups engage in such behavior. The effect diminishes if one of the groups does, in fact, provide evaluations over the whole group on the dimension intended. The effect also diminishes as informal subgroup size decreases or as the number of subgroups increases.

In terms of technique, the effect of subgroup behavior is pronounced if ratings or rankings are used. Resultant scores are most likely to be negatively skewed. The use of full nominations will tend to produce scores with decreased variance, and high nominations will produce the worst case with a drastic reduction in variance. An important point when using nominations is that the use of too many nominations relative to total group size may increase the effect of subgroup behavior (see Downey, 1974).

It is clear that subgroups of sufficient size can have an effect upon the final scores. The problem is the incidence of such effects and whether there exists a mechanism for detecting them. If the evaluation process is part of an ongoing process, the simplest procedure for checking for these problems is the repetitive production of reliability indices as part of the procedure for producing peer scores. If the reliability coefficients were to drop below .60, it would probably indicate a problem, and care should be taken in use of the evaluations. Alternatively, a two-way analysis of variance design, one factor being the type of raters and the other factor being the same type of ratees could be used. If a significant interaction were found, then a strong case could be made for considering the peer scores as at least partially the result of group membership.

Demographic Characteristics

The use of peer evaluations with their reliance upon fallible human observers immediately raises the possibility of racial and sexual bias on the part of evaluators. This concern is especially crucial in view of recent problems associated with demonstrating the absence of bias in employment selection and classification measures as well as in criterion measures.

The evidence concerning racial bias in peer evaluations is mixed and inconclusive. In a study dealing with Air Force recruits, Cox and Krumboltz (1958) found that subjects were rated higher by members of their own race, but the effect varied across groups, and there was substantial agreement on rank order across races ($r = .76$). They concluded that any bias was far from complete and suggested that prior acquaintanceship of group members might account for the differences. In a similar study in the Army, deJung and Kaplan (1962) found similar results: Ratings differed as a function of the rater's race. However, an analysis of covariance adjusting for a combined interest and math score showed that whites did not give higher adjusted scores to whites or blacks, but that blacks gave higher adjusted scores to blacks. Results were interpreted in terms of assignment of higher scores to close acquaintances--a result had most impact upon blacks rating blacks (because of the smaller group size).

In a more recent study in an industrial training context, Schmidt and Johnson (1971) used a forced-choice rating distribution in groups made up of approximately equal numbers of blacks and whites. No differences due to race were found.

The evidence suggests that peer evaluations can be subject to racial bias, but the effect is perhaps more strongly related to the interaction between friendship or acquaintanceship and the particular evaluation method used than to the fact of race itself. The presence of substantial correlation between the rank orderings from each race indicates that the ordering was not much affected by race. But the use of ratings allows evaluators to assign unrelated scores to individuals whom they consider special in some way.

In terms of sexual bias, Mohr and Downey (1977) recently reported results from a small sample of Army officers, in which females scored lower than males on evaluations received from both males and females. If bias occurred, it was on the part of both groups. An interesting finding was that females' self-ratings were not related to either male or female evaluations, but males' self-ratings were related to these evaluations.

This admittedly small number of studies appears to indicate that differences based upon race and sex can occur, but does not make clear whether these differences are attributable to race or sex group differences, to interaction patterns (e.g., friendships), to the specific methodology, or to some combinations of these factors. It would certainly be safe to say that researchers should be sensitive to the potential for such bias.

Group Boundaries

The discussion of peer evaluations has proceeded to this point as if it were clear just what is meant by a peer or associate group. Most researchers report their procedures in sufficient detail to show the general characteristics of the groups in the study. However, given the variety of overlapping and higher order groups in most real-life settings, the issue becomes that of defining some basic guidelines for selecting the appropriate rating group. It is clear that the selection of the evaluative group can be affected by such factors as length and type of interaction, formal organizational structure, informal group structure, friendship patterns, and, of course, the particular dimension being evaluated.

There are few empirical findings to guide selection of the peer group. Rather, guidelines must be best guesses based on partial information from related data.

In a 1976 study, Downey found that platoon evaluations produced more reliable and slightly more valid scores than did squad evaluations, but the differences were potentially confounded by differences in method and group size. Gordon and Medland's 1965 study, in which individuals were evaluated at two different times by totally different groups, indicated a high degree of stability across the two evaluations. Even the method used to compute reliability indices, random splits of the primary group, supported the notion that group composition can be drastically altered without giving rise to major problems in the reliability and validity of scores.

Hierarchical Characteristics

A concept related to that of group boundaries is that of hierarchies. Suppose one were to perform a peer evaluation procedure in a traditionally hierarchical organization. If work groups at the subordinate level are chosen as the peer groups, what effect does inclusion of their immediate superiors have on the resulting evaluations? Conventional wisdom tends to hold that inclusion of such individuals can contaminate the procedure, and therefore they should be excluded from the worker peer groups and included in a peer group of first-level supervisors.

Again, results bearing upon hierarchical inclusion are mixed. Research by Levi, Torrance, and Pletts (1958) indicated no effects from including the formal leader in the peer evaluation process. Research by Downey in 1975, in which the leaders of small combat units were included in the peer nomination process, indicated that the leaders spanned the full range of peer evaluation scores. There was a positive relationship between formal position and peer evaluation scores of leadership potential (as there should be, if the original selection procedure for leaders had any validity). These data were experimental, and the introduction of an operational system might change the result.

A rational solution to the boundary/hierarchical problem should be guided by the following suggestions:

1. The group selected should be large enough to overcome problems associated with primary groups.

2. The group should not be so large as to include subgroups who may be relatively unknown to each other or may be competing for similar resources and rewards.

3. The function of the group selected should be reasonably related to the dimension to be evaluated; e.g., if evaluation of leadership in a work setting is desired, a work group and not a social group should be selected.

Friendship

Friendship has been a major research issue in the history of peer evaluations. According to folklore, peer evaluations are the product of friendship or popularity and are therefore not valid indications of the dimension under consideration. The impact of this bit of folklore has been that, with the exception of simple validity studies, this is probably the single most researched question associated with peer evaluations.

Wherry and Fryer (1949) were the first to address the issue of friendship in peer ratings. They reported that although there was a moderate degree of relationship between friendship and a leadership criterion, the major portion of the predicted criterion variance was independent of friendship. They concluded that peer evaluations of leadership are not popularity contests. Studies by Gibb (1950) and Horrocks and Wear (1953) in college samples supported Wherry and Fryer's findings. Borgatta (1954) also reported that leadership and popularity evaluations were related, but he failed to draw any conclusions. Several other investigations have documented a moderate degree of relationship between friendship and peer evaluations of leadership (Hollander, 1956; Hollander & Webb, 1955; Theordorson, 1957).

Downey (1974) presented evidence that the use of full nominations (with small numbers of high and low nominations required) reduced the correlation between friendship and leadership evaluations compared with forced distribution ratings.

It seems that when an evaluator is faced with the task of evaluating several people, some of whom he or she considers friends, the evaluator will tend to select a friend rather than another person considered to be of equal, or at least indistinguishable, merit. Therefore, the variance associated with friendship may be a source of systematic error primarily in the middle of the distribution. This systematic error variance will increase in large groups, in which some members are relatively unknown to each other or the interaction patterns are not fully established for all members.

However, in spite of the impressive array of research findings as to the minimal effect of friendship, the "popularity contest" issue remains the argument most consistently offered against the use of peer evaluations in an operational setting.

Length of Association

When peer evaluations are considered for use in any situation, an important question is how long group members must be associated with each other before they can provide reliable and valid evaluations. This issue is often raised in the context of transient training groups.

Research fairly consistently finds that peers can make reliable and valid evaluations after a relatively short period of time--typically 3 to 6 weeks (Hollander, 1957).

Subsidiary to the overall issue is the effect of including a new group member in an intact group. Mayfield (1975) has suggested that in such a situation there may be reason to suspect that a longer period of acquaintanceship is necessary for sufficient integration into the group. A more generalized way of approaching the question is to determine which person is known or not well known to other members of the group. Evidence has shown that an individual not well known to other members of the group will typically be evaluated as near the middle of the distribution of peer evaluation scores within the group (Downey, 1974).

In terms of technique, a nomination procedure is most likely to decrease the error variance associated with acquaintanceship; ratings or rankings tend to capitalize on the error variance and show a greater degree of relationship with acquaintanceship.

Type of Interaction

Although peer evaluations have been used and reported over a span of more than 25 years, they have been applied in rather limited situations. Most of the research has been conducted with junior personnel in a military training context such as Officer Candidate School (OCS). A recent effort to use a peer nomination process in a senior Army officer promotion system produced supportive results (Downey, Medland, & Yates, 1976). Outside the military, Weitz (1958) and subsequently Mayfield (1970; 1975) have worked in industry with insurance salesmen.

Freeberg (1969) reported a project in which peer evaluations were more highly related to a performance criterion when the interaction between peers was relevant to the dimension being evaluated. Bayroff and Machlin (1950) found that leadership evaluations could be made in an academic environment and were highly related to evaluations made after exposure to a situation where leadership was displayed. Lewin, Dubno, and Akula (1971) indicated that video tapes supplied sufficient information for reliable evaluations and that these evaluations were highly related to evaluations from group members.

Until more extensive research is conducted in broader organizational contexts with a wider selection of subject populations, the generality of the peer evaluation process is largely a matter of conjecture. However, it would be safe to assume that peer evaluations of a variety of complex human behaviors can be rendered reliably after exposure of the peers to each other in situations that require the individual to interact either with the environment or with others in relevant situations. Further, the validity of the evaluations will be a function of the degree to which the particular behaviors are relevant

to the dimension under study. Hollander (1956) found that reliable evaluations were given after 1 hour of discussion between peers in a naval OCS class, but the scores had only moderate relationship with evaluations obtained 3 weeks later, and were even less predictive of eventual job performance. This convergence of views by peers after a short period of exposure is probably a function of similar psychological maps of behavior on the part of peers, and the preliminary evaluations are subject to revision based upon further information. There seems to be little advantage in using one evaluative technique over another, so long as the technique does not require the evaluator to make finer discriminations than are possible, based on the type of interaction and the amount of information that can be gathered from the interaction.

SUMMARY

Researchers have used the peer evaluation technique both as a criterion of complex human behavior and as an index of future potential. The particular dimension measured has varied considerably. The validity research summarized presents an impressive array of findings with correlation coefficients in the .30 to .50 range either in a concurrent or a predictive situation. Research on extending the generality of the peer evaluation procedure to a more diverse sampling of peer group types, particularly nontraining groups, has been limited.

The four major techniques have also demonstrated important similarities and differences in their psychometric properties. For example, only ratings can produce comparable scores across different groups without extensive assumptions. Research results indicate little differences in measurement reliability between techniques. The limited findings also indicate that, in general, ratings and rankings are less acceptable than either of the nomination techniques.

In view of the documented and likely effects of various situational factors on the evaluation process, it is important that the researcher be aware of potential problems in the use of peer evaluations. No direct relationship was found between group size and the reliability or validity of the evaluations, but it can be assumed that very small or very large groups will produce less reliable and less valid scores. Group structure and demographic characteristics were found to be sources of potential difficulties. With respect to the popular issues of friendship, acquaintanceship, and type of personal interaction, there is little evidence that these have a major impact on the validity of the scores. Indications are that all techniques are relatively impervious to a variety of situational factors, the nomination technique being perhaps the most versatile.

One possible adjustment in future work with this technique is to begin referring to it as associate evaluation rather than peer evaluation. The term peer evaluation, or more commonly peer rating, has acquired overtones of meaning and often has a negative connotation among those required to perform the evaluations. Moreover, the more generalized rubric "associate evaluation" conceptually embraces more individuals; the distinction should not be merely semantic.

In brief, peer evaluations, or associate evaluations, have been shown to be fruitful tools in both research and application. Several issues regarding their use remain to be resolved, but there is sufficient evidence to suggest that these issues can be resolved, and that they do not detract from the conclusion that associate evaluations are a very powerful tool for discriminating complex human behavior.

REFERENCES

Amir, Y., Kovarsky, Y., & Sharan, S. Peer Nominations as a Predictor of Multistage Promotions in a Ramified Organization. Journal of Applied Psychology, 1970, 54, 462-469.

Bayroff, A. G., & Machlin, C. T. Development of Criteria of Leadership in ROTC. Unpublished manuscript, 1950. (Available from R. G. Downey, Kansas State University.)

Berkshire, J. R., & Nelson, P. D. Leadership Peer Ratings Related to Subsequent Proficiency in Training and in the Fleet (Special Report No. 58-20). Pensacola, Fla.: Naval School of Aviation Medicine, 1958.

Bolton, W. L. An Application of Constant Sum Paired Comparison Technique to Peer Ratings at an Army ROTC Summer Camp. Unpublished master's thesis, University of Tennessee at Knoxville, 1971.

Borgatta, E. F. Analysis of Social Interaction and Sociometric Perception. Sociometry, 1954, 17, 7-32.

Boulger, J. R., & Coleman, J. Research Findings with Peer Ratings (Research Note 8). Washington, D.C.: Division of Research, Peace Corps, 1964.

Butler, R. P. Correlates of Officer Performance (Research Report 74-021). West Point, N.Y.: U.S. Military Academy, 1974.

Clarke, J. M. Picking the 9,000. Infantry Journal, 1946, 59, 7-13.

Cox, J. A., & Krumboltz, J. D. Racial Bias in Peer Ratings of Basic Airmen. Sociometry, 1958, 21, 292-299.

deJung, J. E., & Kaplan, H. Some Differential Effects of Race of Rater and Ratee on Early Peer Ratings of Combat Aptitude. Journal of Applied Psychology, 1962, 46, 370-374.

Doll, R. E. Officer Peer Ratings as a Predictor of Failure to Complete Flight Training. Aerospace Medicine, 1963, 34, 130-131.

Downey, R. G. Associate Ratings and Senior Service School Selection. ARI Research Memorandum 73-4, 1973.

Downey, R. G. Associate Evaluations: Nominations vs. Ratings. ARI Technical Paper 253, 1974.

Downey, R. G. Associate Evaluations: Improving Field Acceptance. ARI Research Memorandum 75-5, 1975.

Downey, R. G. Utilization of Associate Nominations in a Training Environment: Ranger Course. ARI Research Problem Review 76-8, October 1976.

Downey, R. G., Medland, F. F., & Yates, L. G. Evaluation of a Peer Rating System for Predicting Subsequent Promotion of Senior Military Officers. Journal of Applied Psychology, 1976, 61, 206-209.

Drucker, A. J. Predicting Leadership Ratings in the United States Army. Educational and Psychological Measurement, 1957, 17, 240-263.

Freeberg, N. E. Relevance of Rater-Ratee Acquaintance in the Validity and Reliability of Ratings. Journal of Applied Psychology, 1969, 53, 518-524.

Gibb, C. A. The Sociometry of Leadership in Temporary Groups. Sociometry, 1950, 13, 226-243.

Gibb, C. A. Leadership. In G. Lindzey & E. Aronson (Eds.), Handbook of Social Psychology (2nd ed., Vol. 4). Reading, Mass.: Addison-Wesley, 1969, 205-282.

Gordon, L. V. Estimating the Reliability of Peer Ratings. Educational and Psychological Measurement, 1969, 29, 305-313.

Gordon, L. V., & Medland, F. F. The Cross-Group Stability of Peer Ratings of Leadership Potential. Personnel Psychology, 1965, 18, 173-177.

Gulliksen, H. Theory of Mental Tests. New York: John Wiley and Sons, 1950.

Haggerty, H. R. Status Report on Research for the U.S. Military Academy. ARI Technical Research Report 1133, 1963.

Hammer, C. H. A Simplified Technique for Evaluating Basic Trainees on Leadership Potential. ARI Research Memorandum 63-10, 1963.

Hollander, E. P. Buddy Ratings: Military Research and Industrial Implications. Personnel Psychology, 1954, 7, 385-393. (a)

Hollander, E. P. Peer Nominations on Leadership as a Predictor of the Pass-Fail Criterion in Naval Air Training. Journal of Applied Psychology, 1954, 38, 150-153. (b)

Hollander, E. P. The Friendship Factor in Peer Nominations. Personnel Psychology, 1956, 9, 435-447.

Hollander, E. P. The Reliability of Peer Nominations Under Various Conditions of Administration. Journal of Applied Psychology, 1957, 41, 85-90.

Hollander, E. P. Validity of Peer Nominations in Predicting a Distant Performance Criterion. Journal of Applied Psychology, 1965, 49, 434-438.

Hollander, E. P., & Webb, W. B. Leadership, Followership, and Friendship: An Analysis of Peer Nominations. The Journal of Abnormal and Social Psychology, 1955, 50, 163-167.

Horrocks, J. E., & Wear, B. A. An Analysis of Interpersonal Choice Relationships of College Students. The Journal of Social Psychology, 1953, 38, 87-98.

Klieger, W. A., deJung, J. E., & Dubuisson, A. U. Peer Ratings as Predictors of Disciplinary Problems. ARI Technical Research Note 124, 1962.

Kraut, A. I. Prediction of Managerial Success by Peer and Training-Staff Ratings. Journal of Applied Psychology, 1975, 60, 14-19.

Kubany, A. J. Use of Sociometric Peer Nominations in Medical Education. Journal of Applied Psychology, 1957, 41, 389-394.

Levi, M., Torrance, E. P., & Pletts, G. O. Sociometric Studies of Combat Air Crews in Survival Training. Sociometry, 1958, 21, 304-328.

Lewin, A., Dubno, P., & Akula, W. Face-to-Face Interaction in the Peer-Nomination Process. Journal of Applied Psychology, 1971, 55, 495-497.

Lindzey, G., & Byrne, D. Measurement of Social Choice and Interpersonal Attractiveness. In G. Lindzey & E. Aronson (Eds.), Handbook of Social Psychology (2nd ed., Vol. 2). Reading, Mass.: Addison-Wesley, 1968.

Lord, F. M., & Novick, M. R. Statistical Theories of Mental Test Scores. Reading, Mass.: Addison-Wesley, 1968.

MacKinnon, D. W. An Overview of Assessment Centers (CCL Technical Report No. 1). Greensboro, N.C.: Center for Creative Leadership. May 1975.

Mayfield, E. C. Management Selection: Buddy Nominations Revisited. Personnel Psychology, 1970, 23, 377-391.

Mayfield, E. C. Peer Nominations in the Life Insurance Industry. In R. G. Downey & F. F. Medland (Chair), Peer Ratings: Beyond Validation Studies. Symposium presented at the meeting of the American Psychological Association, Chicago, 1975.

Mohr, E. S., & Downey, R. G. Are Women Peers? Journal of Occupational Psychology, 1977, 50, 53-57.

Moreno, J. L. Who Shall Survive? Nervous and Mental Disease Monograph, 1934 (No. 58).

Nadal, R. A. A Review of Peer Rating Studies (Research Report No. 68-8). West Point, N.Y.: Office of Military Psychology and Leadership, U.S. Military Academy, 1968.

Peterson, F. E., Lane, N. E., & Ambler, R. K. Carefulness Peer Ratings as a Predictor of Success in Naval Aviation Training (Special Report 66-1). Pensacola, Fla.: U.S. Naval Aerospace Medical Institute, U.S. Naval Aviation Medical Center, 1966.

Ricciuti, H. N. Ratings of Leadership Potential at the U.S. Naval Academy and Subsequent Officer Performance. Journal of Applied Psychology, 1955, 39, 194-199.

Roadman, H. E. An Industrial Use of Peer Ratings. Journal of Applied Psychology, 1964, 48, 211-214.

Schmidt, F. L., & Johnson, R. H. Effect of Race on Peer Ratings in an Industrial Situation. Journal of Applied Psychology, 1971, 57, 237-241.

Smith, G. M. Usefulness of Peer Ratings of Personality in Educational Research. Educational and Psychological Measurement, 1967, 27, 967-984.

Stevens, S. S. Mathematics, Measurement, and Psychophysics. In S. S. Stevens (Ed.), Handbook of Experimental Psychology. New York: Wiley, 1951, Ch. 1.

Suci, G. J., Vallance, T. R., & Glickman, A. S. An Analysis of Peer Ratings (Technical Bulletin No. 54-9). Newport, R.I.: Bureau of Naval Personnel, 1954.

Theordorson, G. A. The Relationship Between Leadership and Popularity Roles in Small Groups. American Sociological Review, 1957, 18, 58-67.

Tupes, E. C. Relationships Between Behavior Trait Ratings by Peers and Later Officer Performance of USAF Officer Candidate School Graduates (Research Report AFPTRC-TN-57-125). Lackland Air Force Base, Tex.: Personnel Laboratory, Air Force Personnel and Training Research Center, 1957.

U.S. Army Research Institute for the Behavioral and Social Sciences. Selection of Leaders: The Status of the Measurement of Leadership. ARI Report No. 444, April 1943.

Waters, L. K., & Waters, C. W. Peer Nominations as Predictors of Short-Term Sales Performance. Journal of Applied Psychology, 1970, 54, 42-44.

Weitz, J. Selecting Supervisors with Peer Ratings. Personnel Psychology, 1958, 11, 25-35.

Wherry, Robert J. Validation of a Program for Selection of Officers for Retention in the Peacetime Army. ARI Research Report 704, July 1945.

Wherry, R. J., & Fryer, D. H. Buddy Ratings: Popularity Contest or Leadership Criteria? Personnel Psychology, 1949, 2, 147-159.

Wiggins, N., Blackburn, M., & Hackman, J. R. Prediction of First-Year Graduate Success in Psychology: Peer Ratings. The Journal of Educational Research, 1969, 63, 81-85.

Williams, S. B., & Leavitt, H. J. Group Opinion as a Predictor of Military Leadership. Journal of Consulting Psychology, 1947, 11, 283-291.

Willingham, W. W. A Note on Peer Nominations as a Predictor of Success in Naval Flight Training (Research Report). Pensacola, Fla.: U.S. Naval School of Aviation Medicine, 1958.

Willingham, W. W. Estimating the Internal Consistency of Mutual Peer Nominations. Psychological Reports, 1959, 5, 163-167.

DISTRIBUTION

ARI Distribution List

4 OASD (M&RA)
2 HQDA (DAMI-CSZ)
1 HQDA (DAPE-PBR)
1 HQDA (DAMA-AR)
1 HQDA (DAPE-HRE-PO)
1 HQDA (SGRD-ID)
1 HQDA (DAMI-DOT-C)
1 HQDA (DAPC-PMZ-A)
1 HQDA (DACH-PPZ-A)
1 HQDA (DAPE-HRE)
1 HQDA (DAPE-MPO-C)
1 HQDA (DAPE-DW)
1 HQDA (DAPE-HRL)
1 HQDA (DAPE-CPS)
1 HQDA (DAPD-MFA)
1 HQDA (DARD-ARS-P)
1 HQDA (DAPC-PAS-A)
1 HQDA (DUSA-OR)
1 HQDA (DAMO-RQR)
1 HQDA (DASG)
1 HQDA (DA10-PI)
1 Chief, Consult Div (DA-OTSG), Adelphi, MD
1 Mil Asst, Hum Res, OCDR&E, OAD (E&LS)
1 HQ USARAL, APO Seattle, ATTN: ARAGP-R
1 HQ First Army, ATTN: AFKA-OI-TI
2 HQ Fifth Army, Ft Sam Houston
1 Dir, Army Stf Studies Ofc, ATTN: OAVCSA (DSP)
1 Ofc Chief of Stf, Studies Ofc
1 DCSPER, ATTN: CPS/OCP
1 The Army Lib, Pentagon, ATTN: RSS Chief
1 The Army Lib, Pentagon, ATTN: ANRAL
1 Ofc, Asst Sec of the Army (R&D)
1 Tech Support Ofc, OJCS
1 USASA, Arlington, ATTN: IARD-T
1 USA Rsch Ofc, Durham, ATTN: Life Sciences Dir
2 USARIEM, Natick, ATTN: SGRD-UE-CA
1 USATTC, Ft Clayton, ATTN: STETC-MO-A
1 USAIMA, Ft Bragg, ATTN: ATSU-CTD-OM
1 USAIMA, Ft Bragg, ATTN: Marquat Lib
1 US WAC Ctr & Sch, Ft McClellan, ATTN: Lib
1 US WAC Ctr & Sch, Ft McClellan, ATTN: Trng Dir
1 USA Quartermaster Sch, Ft Lee, ATTN: ATSM-TS
1 Intelligence Material Dev Ofc, EWL, Ft Holabird
1 USA SE Signal Sch, Ft Gordon, ATTN: ATSO-EA
1 USA Chaplain Ctr & Sch, Ft Hamilton, ATTN: ATSC-TE-RD
1 USATSCH, Ft Eustis, ATTN: Educ Advisor
1 USA War College, Carlisle Barracks, ATTN: Lib
2 WRAIR, Neuropsychiatry Div
1 DLI, SDA, Monterey
1 USA Concept Anal Agcy, Bethesda, ATTN: MOCA-WGC
1 USA Concept Anal Agcy, Bethesda, ATTN: MOCA-MR
1 USA Concept Anal Agcy, Bethesda, ATTN: MOCA-JF
1 USA Artic Test Ctr, APO Seattle, ATTN: STEAC-MO-ASL
1 USA Artic Test Ctr, APO Seattle, ATTN: AMSTE-PL-TS
1 USA Armament Cmd, Redstone Arsenal, ATTN: ATSK-TEM
1 USA Armament Cmd, Rock Island, ATTN: AMSAR-TSC
1 FAA-NAFEC, Atlantic City, ATTN: Library
1 FAA-NAFEC, Atlantic City, ATTN: Hum Engr Br
1 FAA Aeronautical Ctr, Oklahoma City, ATTN: AAC-44D
2 USA Fld Arty Sch, Ft Sill, ATTN: Library
1 USA Armor Sch, Ft Knox, ATTN: Library
1 USA Armor Sch, Ft Knox, ATTN: ATSB-DI-E
1 USA Armor Sch, Ft Knox, ATTN: ATSB-DT-TP
1 USA Armor Sch, Ft Knox, ATTN: ATSB-CD-AD
2 HQUSACDEC, Ft Ord, ATTN: Library
1 HQUSACDEC, Ft Ord, ATTN: ATEC-EX-E-Hum Factors
2 USAEEC, Ft Benjamin Harrison, ATTN: Library
1 USAPACDC, Ft Benjamin Harrison, ATTN: ATCP-HR
1 USA Comm-Elect Sch, Ft Monmouth, ATTN: ATSN-EA
1 USAEC, Ft Monmouth, ATTN: AMSEL-CT-HDP
1 USAEC, Ft Monmouth, ATTN: AMSEL-PA-P
1 USAEC, Ft Monmouth, ATTN: AMSEL-SI-CB
1 USAEC, Ft Monmouth, ATTN: C, Fast Dev Br
1 USA Materiels Sys Anal Agcy, Aberdeen, ATTN: AMXSY-P
1 Edgewood Arsenal, Aberdeen, ATTN: SAREA-BL-H
1 USA Ord Ctr & Sch, Aberdeen, ATTN: ATSL-TEM-C
2 USA Hum Engr Lab, Aberdeen, ATTN: Library/Dir
1 USA Combat Arms Trng Bd, Ft Benning, ATTN: Ad Supervisor
1 USA Infantry Hum Rsch Unit, Ft Benning, ATTN: Chief
1 USA Infantry Bd, Ft Benning, ATTN: STEBC-TE-T
1 USASMA, Ft Bliss, ATTN: ATSS-LRC
1 USA Air Def Sch, Ft Bliss, ATTN: ATSA-CTD-ME
1 USA Air Def Sch, Ft Bliss, ATTN: Tech Lib
1 USA Air Def Bd, Ft Bliss, ATTN: FILES
1 USA Air Def Bd, Ft Bliss, ATTN: STEBD-PO
1 USA Cmd & General Stf College, Ft Leavenworth, ATTN: Lib
1 USA Cmd & General Stf College, Ft Leavenworth, ATTN: ATSW-SE-L
1 USA Cmd & General Stf College, Ft Leavenworth, ATTN: Ed Advisor
1 USA Combined Arms Cmbt Dev Act, Ft Leavenworth, ATTN: DepCdr
1 USA Combined Arms Cmbt Dev Act, Ft Leavenworth, ATTN: CCS
1 USA Combined Arms Cmbt Dev Act, Ft Leavenworth, ATTN: ATCASA
1 USA Combined Arms Cmbt Dev Act, Ft Leavenworth, ATTN: ATCACC-E
1 USA Combined Arms Cmbt Dev Act, Ft Leavenworth, ATTN: ATCACC-CI
1 USAECOM, Night Vision Lab, Ft Belvoir, ATTN: AMSEL-NV-SD
3 USA Computer Sys Cmd, Ft Belvoir, ATTN: Tech Library
1 USAMERDC, Ft Belvoir, ATTN: STSFB-DQ
1 USA Eng Sch, Ft Belvoir, ATTN: Library
1 USA Topographic Lab, Ft Belvoir, ATTN: ETL-TD-S
1 USA Topographic Lab, Ft Belvoir, ATTN: STINFO Center
1 USA Topographic Lab, Ft Belvoir, ATTN: ETL-GSL
1 USA Intelligence Ctr & Sch, Ft Huachuca, ATTN: CTD-MS
1 USA Intelligence Ctr & Sch, Ft Huachuca, ATTN: ATS-CTD-MS
1 USA Intelligence Ctr & Sch, Ft Huachuca, ATTN: ATSI-TE
1 USA Intelligence Ctr & Sch, Ft Huachuca, ATTN: ATSI-TEX-CS
1 USA Intelligence Ctr & Sch, Ft Huachuca, ATTN: ATSI-CTS-CR
1 USA Intelligence Ctr & Sch, Ft Huachuca, ATTN: ATSI-CTD-DT
1 USA Intelligence Ctr & Sch, Ft Huachuca, ATTN: ATSI-CTD-CS
1 USA Intelligence Ctr & Sch, Ft Huachuca, ATTN: DAS/SRD
1 USA Intelligence Ctr & Sch, Ft Huachuca, ATTN: ATSI-TEM
1 USA Intelligence Ctr & Sch, Ft Huachuca, ATTN: Library
1 CDR, HQ Ft Huachuca, ATTN: Tech Ref Div
2 CDR, USA Electronic Prvg Grd, ATTN: STEEP-MT-S
1 CDR, Project MASSTER, ATTN: Tech Info Center
1 Hq MASSTER, USATRADOC, LNO
1 Research Institute, HQ MASSTER, Ft Hood
1 USA Recruiting Cmd, Ft Sheridan, ATTN: USARCPM-P
1 Senior Army Adv., USAFAGOD/TAC, Elgin AF Aux Fld No. 9
1 HQ USARPAC, DCSPER, APO SF 96558, ATTN: GPPE-SE
1 Stimson Lib, Academy of Health Sciences, Ft Sam Houston
1 Marine Corps Inst., ATTN: Dean-MCI
1 HQUSMC, Commandant, ATTN: Code MTMT 51
1 HQUSMC, Commandant, ATTN: Code MPI-20
2 USCG Academy, New London, ATTN: Admission
2 USCG Academy, New London, ATTN: Library
1 USCG Training Ctr, NY, ATTN: CO
1 USCG Training Ctr, NY, ATTN: Educ Svs Ofc
1 USCG, Psychol Res Br, DC, ATTN: GP 1/62
1 HQ Mid-Range Br, MC Det, Quantico, ATTN: Pstl Div

1 US Marine Corps Liaison Ofc, AMC, Alexandria, ATTN: AMCGS-F
1 USATRADOC, Ft Monroe, ATTN: ATRO-ED
6 USATRADOC, Ft Monroe, ATTN: ATPR-AD
1 USATRADOC, Ft Monroe, ATTN: ATTS-EA
1 USA Forces Cmd, Ft McPherson, ATTN: Library
2 USA Aviation Test Bd, Ft Rucker, ATTN: STEBG-PO
1 USA Agcy for Aviation Safety, Ft Rucker, ATTN: Library
1 USA Agcy for Aviation Safety, Ft Rucker, ATTN: Educ Advisor
1 USA Aviation Sch, Ft Rucker, ATTN: PO Drawer O
1 HQUSA Aviation Sys Cmd, St Louis, ATTN: AMSAV-ZDR
2 USA Aviation Sys Test Act., Edwards AFB, ATTN: SAVTE-T
1 USA Air Def Sch, Ft Bliss, ATTN: ATSA TEM
1 USA Air Mobility Rsch & Dev Lab, Moffett Fld, ATTN: SAVDL-AS
1 USA Aviation Sch, Res Trng Mgt, Ft Rucker, ATTN: ATST-T-RTM
1 USA Aviation Sch, CO, Ft Rucker, ATTN: ATST-D-A
1 HQ, DARCOM, Alexandria, ATTN: AMXCD-TL
1 HQ, DARCOM, Alexandria, ATTN: CDR
1 US Military Academy, West Point, ATTN: Serials Unit
1 US Military Academy, West Point, ATTN: Ofc of Milt Leadrshp
1 US Military Academy, West Point, ATTN: MAOR
1 USA Standardization Gp, UK, FPO NY, ATTN: AASE-QC
1 Ofc of Naval Rsch, Arlington, ATTN: Code 452
3 Ofc of Naval Rsch, Arlington, ATTN: Code 458
1 Ofc of Naval Rsch, Arlington, ATTN: Code 450
1 Ofc of Naval Rsch, Arlington, ATTN: Code 441
1 Naval Aerospc Med Res Lab, Pensacola, ATTN: Acous Sch Div
1 Naval Aerospc Med Res Lab, Pensacola, ATTN: Code L51
1 Naval Aerospc Med Res Lab, Pensacola, ATTN: Code L5
1 Chief of NavPers, ATTN: Pers-OR
1 NAVAIRSTA, Norfolk, ATTN: Safety Ctr
1 Nav Oceanographic, DC, ATTN: Code 6251, Charts & Tech
1 Center of Naval Anal, ATTN: Doc Ctr
1 NavAirSysCom, ATTN: AIR-5313C
1 Nav SubMed, ATTN: 713
1 NavHelicopterSubSqua 2, FPO SF 96601
1 AFHRL (FT) William AFB
1 AFHRL (TT) Lowry AFB
1 AFHRL (AS) WPAFB, OH
2 AFHRL (DOJZ) Brooks AFB
1 AFHRL (DOJN) Lackland AFB
1 HQUSAF (INYSD)
1 HQUSAF (DPXXA)
1 AFVTG (RD) Randolph AFB
3 AMRL (HE) WPAFB, OH
2 AF Inst of Tech, WPAFB, OH, ATTN: ENE/SL
1 ATC (XPTD) Randolph AFB
1 USAF AeroMed Lib, Brooks AFB (SUL-4), ATTN: DOC SEC
1 AFOSR (NL), Arlington
1 AF Log Cmd, McClellan AFB, ATTN: ALC/DPCRB
1 Air Force Academy, CO, ATTN: Dep't Beh Sen
5 NavPers & Dev Ctr, San Diego
2 Navy Med Neuropsychiatric Rsch Unit, San Diego
1 Nav Electronics Lab, San Diego, ATTN: Res Lab
1 Nav Trng Cen, San Diego, ATTN: Code 9000-Lib
1 NavPostGraSch, Monterey, ATTN: Code 55As
1 NavPostGraSch, Monterey, ATTN: Code 2124
1 NavTrngEquipCtr, Orlando, ATTN: Tech Lib
1 US Dept of Labor, DC, ATTN: Manpower Admin
1 US Dept of Justice, DC, ATTN: Drug Enforce Admin
1 Nat Bur of Standards, DC, ATTN: Computer Info Section
1 Nat Clearing House for MH-Info, Rockville
1 Denver Federal Ctr, Lakewood, ATTN: BLM
12 Defense Documentation Center
4 Dir Psych, Army Hq, Russell Ofcs, Canberra
1 Scientific Advsr, Mil Bd, Army Hq, Russell Ofcs, Canberra
1 Mil and Air Attache, Austrian Embassy
1 Centre de Recherche Des Facteurs, Humains de la Defense Nationale, Brussels
2 Canadian Joint Staff Washington
1 C/Air Staff, Royal Canadian AF, ATTN: Pers Stl Anal Br
3 Chief, Canadian Def Rsch Staff, ATTN: C/CRDS(W)
4 British Def Staff, British Embassy, Washington
1 Def & Civil Inst of Enviro Medicine, Canada
1 AIR CRESS, Kensington, ATTN: Info Sys Br
1 Militaerpsykologisk Tjeneste, Copenhagen
1 Military Attache, French Embassy, ATTN: Doc Sec
1 Medecin Chef, C.E.R.P.A.-Arsenal, Toulon/Naval France
1 Prin Scientific Off, Appl Hum Engr Rsch Div, Ministry of Defence, New Delhi
1 Pers Rsch Ofc Library, AKA, Israel Defense Forces
1 Ministerie van Defensie, DOOP/KL Afd Sociaal Psychologische Zaken, The Hague, Netherlands