AL/HR-TP-1995-0037

EXPLORING THE CONCEPT OF ACCEPTABILITY AS A CRITERION FOR EVALUATING PERFORMANCE MEASURES

Jerry W. Hedge

Personnel Decisions Research Institutes, Inc.
100 S. Ashley Drive
Suite 1230
Tampa, FL 33602

Mark S. Teachout

HUMAN RESOURCES DIRECTORATE
TECHNICAL TRAINING RESEARCH DIVISION
7909 Lindbergh Drive
Brooks AFB, Texas 78235-5352

December 1995

Interim Technical Paper for Period September 1992 - June 1995

ARMSTRONG LABORATORY

AIR FORCE MATERIEL COMMAND
BROOKS AIR FORCE BASE, TEXAS

NOTICE

Publication of this paper does not constitute approval or disapproval of the ideas or findings. It is published in the interest of scientific and technical information (STINFO) exchange.

When Government drawings, specifications, or the data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in anyway supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Office of Public Affairs has reviewed this paper, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This paper has been reviewed and is approved for publication.

Mark S. Teachout

MARK S. TEACHOUT, Ph.D.
Project Scientist
Technical Training Research Division

R Bruce Gould

R. BRUCE GOULD, Ph.D.
Technical Director
Technical Training Research Division

James B. Bushman

JAMES BUSHMAN, Lt Col, USAF
Chief, Technical Training Research Division

Please notify this office, AL/HRPP, 7909 Lindbergh Drive, Brooks AFB TX 78235-5352, if your address changes, or if you no longer want to receive our techinical reports. You may write or call the STINFO office at DSN 240-3853 or commercial (210) 536-3853.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	December 1995	Interim Paper - September 1992 - June 1995

4. TITLE AND SUBTITLE

Exploring the Concept of Acceptability as a Criterion for Evaluating Performance Measures

5. FUNDING NUMBERS

PE - 62205F
PR - 1121
TA - 12
WU - 00

6. AUTHOR(S)
Jerry W. Hedge
Mark S. Teachout

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Personnel Decisions Research Institutes, Inc.
100 S. Ashley Drive, Suite 1230
Tampa, FL 33602

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Armstrong Laboratory
Human Resources Directorate
Technical Training Research Division
7909 Lindbergh Drive
Brooks AFB, TX 78235-5352

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

AL/HR-TP-1995-0037

11. SUPPLEMENTARY NOTES

Technical Monitor: Dr. Mark S. Teachout, DSN: 240-2932
Comm: (210) 536-2932

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release: distribution is unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This paper explores the construct of acceptability as a criterion for evaluating rater reactions to several different rating forms. Self and peer job performance ratings were completed by enlisted Air Force incumbents, in addition to the supervisors of those incumbents. Questionnaires were completed by participants to determine their perceptions of rating form acceptability and factors related to acceptability, including motivation to rate, job satisfaction, situational constraints and rater trust. Results indicated that motivation to rate, trust in others, and situational constraints were predictive of acceptability for both supervisors and job incumbents. In addition, there were differences in rating form acceptability by rating source and rating form. Overall, supervisors' perceptions were more favorable than incumbents, and the task-level rating form was significantly less acceptable to all rating sources, compared to the three other forms. Results are discussed in terms of the usefulness of acceptability as a criterion in applied research.

14. SUBJECT TERMS			15. NUMBER OF PAGES
Criterion Development	Organizational Justice	Rating Forms	
Distributive Justice	Performance Appraisal	Rater Motivation	
Job Performance	Procedural Justice	Rating Sources	25
Job Satisfaction	Rater Acceptability	Rater Trust	
	Rater Attitudes	Situational Constraints	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION O ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED

CONTENTS

Page

SUMMARY 1

I. INTRODUCTION 1

Attitudes About Performance Appraisal 2
Organizational Justice 3
Performance Appraisal Acceptability 3

II. METHOD 6

Background 6
Participants 6
Questionnaires 7
Rating Forms 8
Procedures 9

III. RESULTS 9

Factor Analyses 9
Regression Analysis 13
Analysis of Variance 15

IV. DISCUSSION 15

REFERENCES 18

LIST OF TABLES

Page

1. Principal Components Analysis of the Job Incumbent Background and Rating Form - Questionnaire 10

2. Principal Components Analysis of Supervisor Background and Rating Form - Questionniare 12

3. User Acceptability Regression Analysis by Incumbent and Supervisor 14

4. Rating Source X Rating Form Analysis of Variance 15

PREFACE

This report documents research conducted on the acceptability of performance ratings as part of the Air Force Job Performance Measurement project. Portions of this research were completed under prime contract number F41689-84-D-0001 with Universal Energy Systems for the Training Systems Division of the Air Force Human Resources Laboratory, Brooks Air Force Base, TX. This paper was completed under in-house Work Unit No. 1121-12-00. Some of these results were presented at the annual meeting of the Society for Industrial and Organizational Psychology, Montreal, CA, April, 1992.

EXPLORING THE CONCEPT OF ACCEPTABILITY AS A CRITERION FOR EVALUATING PERFORMANCE MEASURES

SUMMARY

This study examined raters' reactions to the use of several different rating forms, and the notion of using acceptability as a criterion for evaluating appraisal systems or techniques. A total of 1581 self and peer job performance ratings were completed by enlisted Air Force job incumbents, in conjunction with ratings by 522 supervisors of those incumbents. Questionnaires were administered to determine rater perceptions of rating form acceptability and factors related to acceptability. Factor analyses identified a number of interpretable factors related to acceptability, motivation, job satisfaction, situational constraints, and rater trust. Regression analyses indicated that motivation to rate, trust in others, and situational constraints were predictive of acceptability for both supervisors and job incumbents. ANOVA and post-hoc tests indicated differences in acceptability across rating sources and rating forms. Supervisors' perceptions were more favorable than incumbents, and a task-level rating form was significantly less acceptable to all raters. Results are discussed in terms of usefulness of an acceptability criterion in applied research.

I. INTRODUCTION

Research on the measurement of job performance remains a topic of considerable interest in the industrial/organizational psychology literature, and conceptual and methodological advances continue to be made (Borman, 1991). While criterion measurement is essential for almost any personnel research application, choosing an adequate criterion or set of criteria remains a relatively casual process. As a number of researchers (e.g., Kavanagh, 1982; Sulsky & Balzer, 1992) have noted, there are multiple criteria that can and should be used for judging the quality of measurement instruments, procedures, and systems.

Over the years researchers have identified and/or developed many examples of "criteria for criteria" as standards on which to assess the quality of criterion measures (Weitz, 1961). Bellows (1961) suggested that criterion measures be reliable, realistic, representative, related to other criteria, acceptable to the job analyst, acceptable to management, consistent from one situation to another, and predictable. Blum and Naylor (1968) proposed that criterion measures should also be inexpensive, understandable, measurable, relevant, uncontaminated and bias-free, and discriminating. Bernardin and Beatty (1984) compiled a large list of variables and clustered these variables into three primary categories of criteria: quantitative (e.g., reliability, validity, discriminability), utilization (e.g., feedback, merit pay, adverse impact), and qualitative (e.g., amount of documentation, user acceptability, maintenance costs).

While researchers have periodically called attention to the availability of multiple criteria to judge criteria, operational definitions of these variables are not frequently supplied. If some form of empirical evaluation is included it is typically dominated by reliability and validity considerations. Several researchers have attempted to apply a more systematic process to the use of multiple criteria to judge performance measures.

McAfee and Green (1977) described a set of 16 criteria they applied to aid the selection of a performance appraisal method for use in a large midwestern hospital. Ten different appraisal methods were rated on criteria such as usefulness for counseling and employee develoment, expense to develop, reliability, and freedom from psychometric errors. McAfee and Green then rated each method on each criterion, and used a weighted sum to identify the best method for the job and organization under consideration.

Drawing on the work of McAfee and Green (1977), Kavanagh (1980, 1982) proposed a list of 19 criteria against which to judge the value of performance appraisal systems. Each of these criteria was operationally defined, and included psychometric quality, developmental costs, user acceptance, periodic review/feedback, meets EEOC guidelines, and susceptibility to inflation of ratings. User acceptance, or acceptability was seen as critical to the appraisal system's effect on employee motivation and management control.

In the remainder of this paper we intend to examine more fully the concept of acceptability, and demonstrate how it may be used as a criterion to evaluate the worth of a performance appraisal system or technique.

Attitudes About Performance Appraisal

Recent reviews of performance appraisal have emphasized a broader focus on criteria. Dickinson (1993) reviewed the literature on attitudes about performance appraisal, and suggested that if negative attitudes about performance appraisal prevail among organizational members, performance appraisal will be unacceptable to many members, and its use may hinder rather than help achieve outcomes. In addition, Dickinson's review supported Lawler's contention that appraisal system characteristics, the individual, and the organization are all determinants of attitudes about performance appraisal.

Murphy and Cleveland (1991) noted that the dominance of psychometric and accuracy criteria have diverted researchers' attention away from three classes of criteria that might be critical in determining the success of an appraisal system, namely, 1) reactions, 2) practicality, and 3) decision process criteria.

They argue that reaction criteria (such as perceptions of fairness and accuracy of appraisal systems) probably place a ceiling on the possible effectiveness of the system, since acceptance of the system by raters and ratees may be necessary but not sufficient for the system to be effective. Practicality criteria such as time commitment, cost, political acceptability, and ease of installation are also cited by Murphy and Cleveland as useful but neglected criteria. Finally, the contribution of a performance appraisal system to the decision making process should be considered as well, both in terms of the degree to which decisions are accepted by members of the organization and the degree to which decisions are facilitated by the performance appraisal system.

Organizational Justice

A related body of literature that has received renewed attention recently is organizational justice (Thibaut & Walker, 1975), particularly with its translation into performance appraisal terms (e.g., Greenberg, 1986a). Studies on perceptions of justice and fairness in organizations are directed at identifying the features of organizational procedures that affect perceptions of fairness, work attitudes, and behavior. The literature suggests there are two dimensions of perceived justice to any policy -- distributive justice and procedural justice.

Distributive justice refers to normative standards for evaluating the fairness of the allocation of outcomes between parties (Leventhal, 1976). Distributive justice interpreted in performance appraisal terms focuses on the fairness of the evaluations received relative to the work performed. Distributive justice comes into play when evaluation decisions are emphasized as a means to an end, for example, when persons view performance ratings as a means of obtaining promotions, salary increases, etc. Distributive fairness is reflected in equitable distribution of reward outcomes across persons rather than in the determination of performance ratings.

Procedural justice refers to normative standards for evaluating the manner in which a decision is reached. Procedural justice related to performance appraisal focuses on the fairness of the evaluation procedures used to determine the ratings. In other words, as Greenberg (1986a) suggests, beliefs about fair performance evaluations may be based on the procedures by which the evaluations are determined apart from the evaluation received. Thus, procedural fairness comes into play when performance evaluations are considered as "ends in themselves" (Greenberg, 1986b). Procedural fairness is reflected in the perceived validity of performance measurement procedures and the opportunity for employees to provide a complete picture of their performance to supervisors before the evaluation. Thus, behaviors and components of job performance that contribute to evaluations are inputs.

Within this broad organizational framework, the current study would be classified as addressing performance appraisal issues related to procedural justice. Because our study was for research purposes only, rating outcomes were not of primary concern, but rather how rater attitudes about appraisal related to the processes and procedures of appraisal, as well as to other salient variables.

Performance Appraisal Acceptability

In spite of the common sense logic that acceptance of a personnel procedure is crucial to its effective use, it was not until 1967 that Lawler noted that attitudes toward performance ratings could affect their validity. Lawler (1967) proposed a model of the factors that affect the construct validity of ratings. Central to the model was the belief that attitudes toward the equity and acceptability of a rating system are a function of organizational and individual characteristics, as well as the rating format.

Landy, Barnes, and Murphy (1978) were among the first researchers to empirically examine attitudinal factors as they relate to job performance measurement. They identified four significant predictors of perceived fairness and accuracy of performance appraisals: (a) frequency of appraisal, (b) plans developed with the supervisor for eliminating weaknesses, (c) supervisor's knowledge of the ratee's job duties, and (d) supervisor's knowledge of the ratee's level of performance. In a follow-up study with the same population, the level of the performance rating did not affect these relationships (Landy, Barnes-Farrell, & Cleveland, 1980).

Dipboye and de Pontbriand (1981) distinguished between employees' opinions of their performance appraisal system and employees' opinions of the appraisal itself. They found that four factors related to the two dependent variables: (a) favorability of the appraisal, (b) opportunity for employees to state their own perspective in the appraisal interview, (c) job relevance of appraisal factors, and (d) discussion of plans and objectives with the supervisor.

A series of studies by Kavanagh and colleagues extended the examination of users' perceptions of performance appraisal systems (Hedge, 1983; Kavanagh & Hedge, 1983; Kavanagh, Hedge, Ree, Earles, & DeBiasi, 1984). Although users attitudes toward the appraisal form and the broader concept of the appraisal system did not seem to differ (i.e., virtually identical regression models were found), several attitudes toward the appraisal system were significant predictors of appraisal acceptability across studies. These included attitudes about whether: (a) the appraisal system facilitates fair and accurate appraisals, (b) the appraisal system allows raters to distinguish between workers' proficiencies, (c) the appraisal system provides clear performance standards, (d) ratees receive satisfactory feedback, and (e) ratees receive a satisfactory performance evaluation.

While the study by Hedge and Kavanagh (1983) and Kavanagh et al. (1984) focused exclusively on factors related to acceptability, Hedge (1983) used an acceptability measure (i.e., how acceptable do you find your current performance appraisal system?), in conjunction with more traditional performance appraisal criterion measures to evaluate the implementation of a new performance appraisal system at a large hospital. He discovered that ratees found the new appraisal system more acceptable than the system previously in use.

The objective of the present research was to focus on one criterion, acceptability, that has been relatively under-investigated, develop an operational definition for that variable, and follow a systematic procedure for collecting and evaluating such data. While Kavanagh (1980, 1982), Bernardin and Beatty (1984), and others have discussed *acceptability* as an important criterion, it has rarely been used.

The present study had four interrelated purposes. Because of the scarcity of research that uses or examines the use of reaction criteria, our first purpose was to extend the development of a reaction criterion beyond what has been done to date. Early research focused on single-item measures (e.g., Landy et al., 1978; Landy et al., 1980) of perceived fairness and accuracy. Other researchers (Dobbins, Cardy, & Platz-Vieno, 1990; Giles & Mossholder, 1990) chose satisfaction with appraisal as their single-item reaction measure, arguing that a satisfaction criterion assesses both fairness cognitions and affect, thus offering a broader indicator of appraisal reactions (Giles & Mossholder, 1990). Dipoye and de Pontbriand (1980) used a 3-item measure of satisfaction and understanding of the appraisal process,

and a 4-item measure of whether the appraisal system facilitates employee evaluation. Kavanagh et al. (1984) focused on both the system and the form, broadening the concept to emphasize overall acceptability, but once again using single-item measures.

Following the advice of Kavanagh (1982), Bernardin and Beatty (1984), and Murphy and Cleveland (1991), we chose to focus on the construct of acceptability as the measure that would best capture reactions to appraisal. Building on previous research findings on reactions to appraisal, items were written to reflect broadly the concept of acceptability, including: a) facilitates identification of performance differences between employees, b) facilitates capturing the true picture of job performance, c) overall acceptability of the form, d) ease of form use and understanding, e) facilitates confidence in ratings, and f) facilitates fair evaluation of performers.

A second purpose of the study was to examine the relationship between perceptions of appraisal acceptability and variables both internal and external to the appraisal process. Because of the validation research purpose of our study, variables prevalent in previous appraisal reaction studies (e.g., setting performance objectives, devising action plans, counseling employees, discussing salary issues) were irrelevant and thus not included. We did, however, identify two appraisal process factors from the literature that seemed relevant to our study, rater trust and rater motivation.

Rater motivation has been largely ignored by performance appraisal researchers. Although DeCotiis and Petit (1978) incorporated rater motivation as an important part of their model of the appraisal process, they cited only Taft's (1971) theory of interpersonal judgments as support for the inclusion of this variable in their model. Recently, Bernardin and his colleagues (Bernardin & Cardy, 1982; Bernardin, Orban, & Carlyle, 1981) focused on rater motivation, but only in terms of how it might be affected by the level of trust a rater has in the appraisal system. Bernardin, Orban, & Carlyle (1981) developed a measure they labeled "trust in the appraisal process," and found that both trust and motivation were linked to the perceptions of fairness and accuracy of appraisal. Consequently, for the current study, items were written to tap facets of these factors including: a) general motivation to rate, b) motivation to rate accurately, c) rater trust in the appraisal process, d) trust in other raters, and e) trust in researchers.

Past research within the performance appraisal domain has also identified variables that appear to have relevance for our study. We focused on three particular variables that could impact ratings. Peters and O'Connor (1980) hypothesized that constraints on performance may lead to lower effectiveness levels, and some support has been found for such a notion (e.g., O'Connor, Peters, Rudolf, & Pooyan, 1984; Olson & Borman, 1989). Extending this logic for raters, it was hypothesized that situational constraints may affect raters' ability to rate accurately, thereby affecting perception of appraisal acceptability, and items were written to tap constraints related to tool availability and job manual availability and clarity.

Two other sets of items were also developed to examine the influence of other external variables on performance appraisal acceptability. The two factors that appeared to have some relevance, and had been used in past research studies were supervisory support and job satisfaction.

Dickinson (1993) noted that perhaps the single most important determinant of employee attitudes about performance appraisal is the supervisor. He suggested that when the supervisor is seen as trustworthy and supportive, then attitudes about performance appraisal are favorable. In addition, Olson and Borman (1989) found relationships between supervisory support and job performance, suggesting that supervisory support could be related to attitudes about performance appraisal.

Similarly, while only modest relationships have been found between job satisfaction and job performance (e.g., Iaffaldano & Muchinsky, 1985; Podsakoff & Williams, 1986), we felt it would be useful to explore whether attitudes about the job would be related to attitudes about appraisal system acceptability. For example, Giles and Mossholder (1990) found modest relationships between job satisfaction and satisfaction with the appraisal system.

A third purpose of the present study was to examine the link between rating source and performance appraisal acceptability. While previous studies of performance appraisal attitudes have almost exclusively focused on ratee reactions, the focus of our study was on the reactions of the *raters* to the forms they had been asked to use. In addition, because both job incumbents and supervisors were asked to provide performance ratings and responses to other attitudinal questions, we were able to examine whether the variables associated with appraisal acceptability differed by rating source, and whether levels of appraisal acceptability differed by rating source.

A fourth purpose of our research was to examine whether rater acceptability differed across rating forms. As noted earlier, McAfee and Green (1977) evaluated 10 appraisal methods against a list of 16 criteria as a way to select an appraisal method for nurses in a hospital. Relying on their own knowledge of the different methods, they rated the effectiveness of the methods on the 16 criteria and arrived at a final decision about which type of appraisal method to use. Kavanagh (1982) also recommended that such a procedure be used, but to our knowledge, no published study has gathered attitudinal information (from the individuals who would be asked to use the forms) as one component of the measurement method selection process. Thus, a final aim of our research study was to collect data on rater attitudes about the acceptability of four separate performance appraisal forms that had been developed for possible use in a validation project.

In summary, there is little empirical research concerning perceptions of appraisal system acceptability. The purpose of the present research was to identify and clarify the construct of appraisal acceptability, and examine factors related to this acceptability construct. We also wanted to examine whether attitudes about acceptability differ across rating sources and rating forms.

II. METHOD

Background

Betweeen 1984 and 1989 the Air Force Human Resources Laboratory[1] conducted a large-scale research project to develop a variety of performance measures for use in the validation of

[1] This is now the United States Air Force Armstrong Laboratory, Human Resources Directorate.

selection and classification tests and evaluation of training programs (Hedge & Teachout, 1986; Teachout & Pellum, 1990). As part of this project, a variety of different rating forms were developed to evaluate the job performance of enlisted personnel in their first four years of military service.

Participants

Personnel from seven Air Force specialties participated in this research. These specialties included Air Traffic Control Operator, Aircrew Life Support Specialist, Information Systems Radio Operator, Aerospace Ground Equipment Mechanic, Personnel Specialist, Precision Measurement Equipment Laboratory Specialist, and Avionic Communications Specialist. A total of 1581 job incumbents (ratees and peers[2]) and 522 supervisors completed self, peer, or supervisor ratings (5530 ratings were completed), as well as Background and Rating Form Questionnaires. Job incumbents averaged 27.5 months of Total Active Federal Military Service; 79.0% were male, and 75.4% were caucasian.

Questionnaires

Two questionnaires were developed to gather information from job incumbents and supervisors both before they made ratings (using a Background Questionnaire), and after they made ratings (using a Rating Form Questionnaire). The Background Questionnaire included 10 items hypothesized to measure three different constructs. Three items measured situational constraints (e.g., "The technical manuals and other written materials that I use in my job are available when I need them."). Five items measured job satisfaction (e.g., "I get a sense of accomplishment from my job."). Two items measured supervisory support (e.g., "I feel that my supervisor gives me the support I need to do my job.").

The Rating Form Questionnaire contained 20 items hypothesized to measure three constructs. Seven items measured the rater's motivation to rate (e.g., "How motivated were you to complete the rating forms?"; "Did you make an 'extra effort' to carefully pay attention to all of the instructions and examples in order to make accurate ratings?"). Seven items measured rater trust in the appraisal process, in other raters, and in the researchers conducting the research (e.g., "Will your supervisor have access to any information about you collected from the rating forms?"; "Do you believe other persons involved really tried to follow the rules in completing their ratings?"; "Do you believe that the true purpose of the ratings was the one explained to you during the rater orientation?"; several of these items were similar to those used by Bernardin, Orban, & Carlyle, 1981).

Six items measured acceptability of the appraisal process. The six acceptability items were designed to tap perceptions of appraisal form (a) fairness, (b) clarity of instructions, (c) contributions to rating accuracy, (d) contributions to discrimination between ratees, (e) overall acceptability to raters, and (f) confidence raters had in their ratings. Raters responded to the same six acceptability questions for each of the four rating forms.

---

[2] Job incumbents could be asked to provide a self rating, a peer rating, or both types of ratings. However, regardless of the rating requirements, all *job incumbents* were in their first four years of military service.

Scales for all items were five-point, adjectivally anchored graphic rating scales. Across the seven specialties, Background and Rating Form Questionnaire items were identical, with one exception. The Air Traffic Control Operator Background Questionnaire omitted one "constraint item" that asked about tool and equipment availability.

Rating Forms

A series of four rating forms were developed to measure job performance. All rating forms were constructed using a 5-point, adjectivally anchored rating scale. In addition, specific behavioral examples were included for three of the four rating forms to provide detailed information to assist the raters in making accurate judgments.

Task Rating Form. This form consisted of a comprehensive listing of tasks representative of the job content domain. Task identification was based on an extensive stratified random sampling plan (Lipscomb & Dickinson, 1988) that used information obtained from the Air Force's Occupational Survey Program (Christal, 1974). The relative amount of time spent performing these tasks, learning difficulty, and emphasis given to the tasks in training were used to select a representative set of tasks. The number of tasks included on a Task Rating Form varied between 25 and 40 across the seven Air Force specialties. Ratings were made on a 5-point graphic rating scale, with numerical and adjectival anchors at each of these five points. The scale ranged from "1" -- never meets acceptable level of proficiency to "5" -- always exceeds acceptable level of proficiency.

Dimensional Rating Form. This rating form consisted of 4 to 10 technical dimensions designed to encompass the domain of job performance within each specialty. Potential dimensions were identified through factor analysis of co-performance ratings for tasks that are performed by first-term enlisted personnel. Subject-matter experts (SMEs) used this information in preliminary workshops to identify and define technical dimensions, and to generate and categorize specific behavioral examples for each dimension. In a series of follow-up workshops, the set of dimensions was reviewed, revised, and confirmed, and the specific behavioral examples were developed and revised. These examples were then assigned to dimensions and scale values through a standard retranslation process. The behavioral examples were developed using a variant of the Behavior Summary Scale (BSS) approach (Borman, 1979), where valid SME-generated behavioral anchors at each level were combined to form paragraph descriptors of that proficiency level. For example, these paragraphs described technical effectiveness, technical efficiency, and amount of supervision relevant to each proficiency level.

Air Force-wide Rating Form. This rating form consisted of eight performance dimensions descriptive of success across all Air Force specialties. Because of this cross-specialty focus, workshop participants were resource managers who have oversight responsibility for, and knowledge of, many different specialties. Their combined knowledge provided the details for constructing a 5-point BSS rating form applicable across all specialties. This form contained a broad range of dimensions, including technical ability, initiative/effort, adherence to regulations, leadership, military appearance, self-development, and self-control. In addition, behavioral examples, specific to each dimension, anchored each of the five scale values.

Global Rating Form. This 2-item rating form consisted of an overall technical and an overall interpersonal rating. Once again, a series of workshops with SMEs from each specialty was used to generate 5-point BSS rating scales. Just as with the Dimensional Rating Form, the behavioral examples for the technical item depicted technical effectiveness, technical efficiency, and amount of supervision relevant to each specialty. The behavioral examples for the interpersonal item described initiative, effort, and teamwork relevant to each specialty.

Procedures

Prior to the completion of all rating forms and questionnaires, raters were introduced to the purpose of data collection, participation requirements were explained, and they were familiarized with each measure used in the project. This orientation session was followed by approximately 1 hour of frame-of-reference and rater error training (for a detailed description see Bierstedt & Hedge, 1987). Immediately following this group session, rating booklets were distributed, and raters were asked to complete all measures. The rating booklets were organized such that each rater completed the Background Questionnaire followed by the Global, Dimensional, Task, and Air Force-wide rating forms, and then the Rating Form Questionnaire. Supervisors were asked to rate up to three job incumbents under their supervision. Job incumbents were asked to rate themselves and/or up to three of their co-workers. Thus, a job incumbent could be a self rater, a peer rater, or both. Regardless of the number of ratings completed by a rater, Background and Rating Form Questionnaire data were collected only once per rater.

III. RESULTS

Factor Analyses

The 10 Background Questionnaire items, the 6 acceptability items (for each of four rating forms), 7 motivation to rate items, and 7 appraisal trust items from the Rating Form Questionnaire were factor analyzed separately to clarify and refine the hypothesized constructs. Each analysis used the principle components extraction technique, with orthogonal rotation of factors having eigenvalues of 1.0 or greater to a varimax solution. These factor analyses were performed separately on supervisor and job incumbent data. Because acceptability data (six items) were collected on each rating form, separate factor analyses were computed for each form.[3]

---

[3] In each case, all six items loaded quite similarly and strongly on one acceptability construct across the seven specialties. Subsequently, separate factor and regression analyses were computed using acceptability data from each of the four rating forms. That is, eight factor analyses and eight regression analyses (four rating forms by two sources) were computed. However, because the acceptability factor loadings were quite similar, only the results using Task Rating Form data are presented in Tables 1, 2, and 3. Results using the other rating form acceptability composites are available from the first author.

Following the separate questionnaire analyses, a higher order factor analysis was performed to combine the four sets of factors into one general set of appraisal-related dimensions. Because factors such as acceptability, trust in the appraisal process, and motivation to rate were hypothesized to be intercorrelated, the higher order analysis employed a principle components model with the direct oblimin method of oblique rotation. Once again, data from supervisors and job incumbents were analyzed separately.

Tables 1 and 2 present loadings of variables on factors for job incumbents and supervisors, respectively. Variables are ordered and grouped by size of loading to facilitate interpretation. Loadings under .45 (20% of variance) were excluded. Nine interpretable factors were identified for job incumbents and supervisors, although the factors were not identical across the two sources. The interpretable 9-factor solution for the job incumbent data set included the following factors: a) motivation to rate accurately, b) job satisfaction, c) acceptability, d) situational constraints, e) trust in other raters, f) supervisory support, g) trust in the appraisal process, h) trust in researchers, and i) general motivation to rate. The 9-factor solution from the supervisor data set produced eight factors (a - h above) in common with the job incumbent set. However, supervisors did not distinguish between general motivation to rate and motivation to rate accurately, but (unlike job incumbents) they did distinguish between job satisfaction and "esprit de corp." High loadings across data sets and factors suggest relatively well-defined constructs.

Table 1. Principal Components Analysis of the Job Incumbent Background and Rating Form - Questionnaire.

Factor Label and Items	Loading
1. Motivation to Rate Accurately	
a: satisfied ratings were accurate	.76
b: extra effort to pay attention	.74
c: care about rating accuracy	.72
d: important to make accurate ratings	.71
e: in general, accurate ratings important	.68
2. Job Satisfaction	
a: job is interesting	.90
b: satisfied with job	.85
c: sense of accomplishment from job	.83
d: job important to AF mission	.67
e: able to use skills/talents in job	.65

Factor Label and Items	Loading
3. Acceptability of Rating Form	
a: allow true picture of performers	.83
b: show differences between performers	.81
c: acceptable to users	.81
d: evaluate job proficiency fairly	.72
e: easy to use and understand	.67
f: instill confidence in ratings	.45
4. Situational Constraints	
a: technical manuals are available	.86
b: tools and equipment available	.80
c: technical manuals clear/understandable	.46
5. Trust in Other Raters	
a: others tried to follow rating rules	-.75
b: others cared about accurate ratings	-.72
c: others gave higher ratings than deserved	.52
6. Supervisor Support	
a: supervisor gives support I need	.96
b: supervisor concerned about well-being	.95
7. Trust in Appraisal Process	
a: others comfortable giving low ratings	.73
b: supervisor access to this information	.64
8. General Motivation to Rate	
a: motivated to complete rating forms	.73
b: rating process interesting	.71
9. Trust in Researchers	
a: ratings used for research purposes	.84
b: true purpose of rating explained	.71

Table 2. Principal Components Analysis of the Supervisor Background and Rating Form - Questionnaire.

Factor Label and Items	Loading
1. Motivation to Rate Accurately	
a: important to make accurate ratings	.88
b: care about rating accuracy	.86
c: extra effort to pay attention	.79
d: in general, accurate ratings important	.68
e: satisfied ratings were accurate	.57
2. Job Satisfaction	
a: sense of accomplishment from job	.82
b: able to use skills/talents in job	.77
c: satisfied with job	.69
3. Acceptability of Rating Form	
a: show differences between performers	.86
b: allow true picture of performers	.82
c: acceptable to users	.81
d: evaluate job proficiency fairly	.75
e: easy to use and understand	.73
f: instill confidence in ratings	.57
4. Supervisor Support	
a: supervisor gives support I need	.92
b: supervisor concerned about well-being	.89
5. Trust in Other Raters	
a: others tried to follow rating rules	.87
b: others cared about accurate ratings	.84
6. Trust in Researchers	
a: ratings used for research purposes	-.80
b: true purpose of rating explained	-.68

Factor Label and Items	Loading
7. Situational Constraints	
a: technical manuals are available	.79
b: tools and equipment available	.64
c: technical manuals clear/understandable	.50
8. Trust in Appraisal Process	
a: supervisor access to this information	.79
b: others comfortable giving low ratings	.64
c: others gave higher ratings than deserved	.50
9. Esprit de Corp	
a: job important to AF mission	-.68
b: sense of pride being in AF	-.64

Regression Analyses

In an effort to identify factors predictive of acceptability, multiple regression analyses were conducted separately for supervisors and job incumbents. Based on the previously-derived factor solutions, an overall acceptability dependent measure was formed by unit weighting the six items loading on that factor for each of the four rating forms. Recall that attitudes about acceptability of the four rating forms were gathered separately for each form. Thus, our overall acceptability measure was formed by combining scores on the six acceptability items across the four rating forms, yielding a 24-item acceptability composite. Likewise, independent measures were formed by unit weighting the items loading on each factor identified in the principle components analysis, yielding composites for eight supervisor and eight job incumbent factors.

In addition, to assess the contribution of Air Force specialty to variance in the dependent measure, specialties were dummy coded as an independent variable, and forced into the regression equation first, followed by the remainder of the independent variables entering in a forward inclusion manner. The results of the multiple regression analysis are presented in Table 3.

Table 3. User Acceptability Regression Analysis by Incumbent and Supervisor.

Factor	Beta	Cumulative multiple R	Cumulative R squared
		Job Incumbent	
Specialty	.018	.055	.003
Motivation to Rate Accurately	.200	.429	.184
Trust in Researchers	.221	.502	.252
General Motivation to Rate	.193	.534	.286
Trust in Other Raters	.111	.548	.300
Trust in the Appraisal Process	.106	.557	.311
Situational Constraints	.088	.564	.318
		Supervisor	
Specialty	.039	.077	.006
Motivation to Rate Accurately	.236	.387	.150
Trust in Researchers	.173	.438	.192
Trust in Other Raters	.160	.467	.218
Trust in the Appraisal Process	.137	.487	.237
Esprit de Corp	.112	.503	.253
Situational Constraints	.091	.510	.260

For job incumbents, six factors (listed by order of entry into the regression equation) were identified as predictors of acceptability. These included motivation to rate accurately, trust in researchers, general motivation to rate, trust in other raters, trust in the appraisal process, and situational constraints. These six measures accounted for 32% of the variance in acceptability. For supervisors, six factors were identified as predictors of acceptability: motivation to rate accurately, trust in researchers, trust in other raters, trust in the appraisal process, esprit de corp, and situational constraints, which accounted for 26% of the variance in the dependent measure. In general, rater motivation, rater trust, and situational constraints on work performance were significantly related to acceptability in both rater groups. Supervisor support and job satisfaction variables did not account for appreciable variance in acceptability, although supervisors did believe that feelings of esprit de corp could influence attitudes about appraisal acceptability. These findings of modest relationships between appraisal attitudes and job satisfaction or supervisory support are consistent with results reported by Giles and Mossholder (1990). Finally, the Air Force specialty was not an important factor in the variance of the dependent measure (accounting for only .003% and .006% of job incumbents' and supervisors' acceptability respectively).

Analysis of Variance

In order to investigate differences in acceptability across rating sources and rating forms a Rating Source (2) x Rating Form (4) analysis of variance (ANOVA) was computed, using the 6-item acceptability composite as the dependent measure. Table 4 displays the results of this analysis.

Table 4. Rating Source X Rating Form Analysis of Variance.

Factor	DF	MS	F
Between subjects			
Rating Source (S)	1	492.96	9.22*
Subjects witin group	2101	53.46	—
Within subjects			
Rating Form (F)	3	286.26	51.99*
S X F	3	9.98	1.81
Subjects within groups	6303	5.51	—

*p < .01

The Rating Source and Rating Form main effects were found to be significantly different than chance ($p < .01$). Scheffe's post hoc tests for differences among means were conducted on each significant effect. For the Source effect, supervisors were found to be more accepting of the measurement system than job incumbents. The rating form post hoc analysis found significant mean differences between the Task Rating Form and all other forms, with the Task Rating Form less acceptable to raters.

IV. DISCUSSION

The present study examined the concept of acceptability of performance ratings, and correlates of acceptability. It then used acceptability as a criterion to assess differences in rater perceptions across rating sources and forms.

Factor analysis identified a number of interpretable factors: rater acceptability, rater motivation, job satisfaction, supervisor support, situational constraints, and rater trust. Comparable factor patterns suggest that job incumbents and supervisors structure perceptions similarly. High factor loadings across these factors indicate relatively well-defined constructs.

It is especially interesting to note the high factor loadings for all items under the acceptability factor. The importance of attitudes toward appraisal system equity and acceptability was noted by Lawler (1967) over 25 years ago. More recently, Landy and his colleagues (Landy, Barnes, & Murphy, 1978; Landy, Barnes-Farrell, & Cleveland, 1980) and Dipboye and de Pontbriand (1981) operationalized Lawler's notion, focusing on perceived fairness and accuracy of the system, and ratee attitudes about the usefulness of the system and the process. Our findings suggest that acceptability is a broader, multi-faceted construct involving perceptions of appraisal fairness, clarity of instruction, accuracy, discriminability, and confidence.

Regression analyses indicate that the same basic information influences job incumbent and supervisor acceptability of the appraisal process: rater motivation, rater trust, and situational constraints. Rater motivation and trust are variables internal to the appraisal process. As Bernardin and his colleagues (Bernardin & Cardy, 1982; Bernardin, Orban, & Carlyle, 1981) have noted, individual rater motivation and trust in the appraisal process may be strongly linked to perceived accuracy and fairness in appraisal. Our results support empirically the link between appraisal process variables and acceptability. This suggests that organizations should foster conditions for motivation and trust in the appraisal process. Rater orientation and training strategies should be helpful in this regard.

External work impediments also seem to effect perceptions of system acceptability. Evidently, raters believe that problems with tool and technical manual availability, and technical manual clarity interfere with not only ratee proficiency but also raters' performance judgments. Previously, Peters and O'Connor (1980) suggested that situational constraints affect job performance. Our findings suggest that constraints may also interfere with *rater* ability to judge job proficiency fairly, accurately, and confidently.

The ANOVA and post hoc test results identified differences in levels of acceptability, with the Task Rating Form significantly less acceptable to all raters, and supervisors' perceptions of the appraisal system more favorable than incumbents'perceptions. These results raise questions about the usefulness of the Task Rating Form, and warrant further investigation since this finding is contrary to expectations. It seems logical to assume that a detailed rating form would allow raters to assess performance more accurately than a more general form, and therefore it would be more acceptable. However, raters may dislike rating individuals on twenty-five to forty items, and more is not better. Perhaps length of time required to rate, rating specificity, or both could be the primary reason(s) for lower acceptability.

Mean differences were also found between rating sources, with supervisor perceptions of the appraisal system more favorable than incumbent perceptions. Why was the appraisal system more acceptable to supervisors than it was to job incumbents? Perhaps supervisor familiarity with the rating process might affect acceptability. Because of the nature of their jobs, supervisors have much more experience rating performance than do job incumbents and, therefore, might be more likely to understand and accept the process. Another possibility could be incumbent skepticism toward the rating process. Since job incumbents have spent their careers being the "target" of ratings perhaps they are more skeptical of the process, and their perceptions of rating acceptability are lower.

Research on the measurement of job performance has been prominent in the industrial/organizational psychology literature for many years. Most of this work has used validity, reliability, and rating error measures. As various authors have noted, however (e. g., Bernardin & Beatty, 1984; Jacobs, Kafry, & Zedeck, 1980; Kavanagh, 1982), there are multiple criteria to use for judging the quality of measurement instruments, procedures, and systems. A relatively uninvestigated criterion is acceptability.

Jacobs et al. (1980), in an examination of the behaviorally anchored rating scale (BARS) literature, noted their own disappointing experiences with organizations abandoning recently-developed appraisal systems. They suggested that many organizations revert back to evaluation systems in use prior to intervention because of organization policy, and the excessive personnel time and energy requirements associated with BARS.

These frustrations, in a very applied way, speak to the issue of acceptability, and suggest the importance of including this variable as a criterion when evaluating worth of an appraisal system. After all, if a psychometrically-sound system is developed, but is unacceptable to its users, it may never be used, or it might be used improperly.

This research has attempted to clarify the concept of acceptability and identify factors related to acceptability. We believe that a rater acceptability criterion can contribute valuable information about the worth of a particular measurement instrument or an appraisal system, and should be used *in conjunction with* other, more frequently used appraisal critera. As Banks and Murphy (1985) have noted, raters must not only be capable, but they must also be *willing* to provide accurate ratings.

V. REFERENCES

Balzer, W. K., & Sulsky, L. M. (1990). Performance appraisal effectiveness. In (K. R. Murphy and F. E. Saal, Eds.), *Psychology in Organizations: Integrating Science and Practice* (pp. 133-156). Hillsdale, NJ: Lawrence Erlbaum.

Banks, C. G., & Murphy, K. R. (1985). Toward narrowing the research-practice gap in performance appraisal. *Personnel Psychology, 38*, 335-345.

Bellows, R.M. (1954). *Psychology of personnel in business and industry* (2nd ed.). Englewood Cliffs, N.J.: Prentice-Hall.

Bernardin, H. J., & Beatty, R. W. (1984). *Performance appraisal: Assessing human behavior at work.* Boston, MA: Kent Publishing.

Bernardin, H. J., & Cardy, R. L. (1982). Appraisal accuracy: The ability and motivation to remember the past. *Public Personnel Management Journal, 119*, 352-357.

Bernardin, H. J., Orban, J. A., & Carlyle, J. J. (1981). Performance ratings as a function of trust in appraisal, purpose for appraisal, and rater individual differences. *Proceedings of the Academy of Management, 311-315.*

Bierstedt, S. A., & Hedge, J. W. (1987, September). *Job performance measurement system trainer's manual* (AFHRL-TP-86-34). Brooks AFB, TX: Training Systems Division, Air Force Human Resources Laboratory.

Blum, M.L., & Naylor, J.C. (1968). *Industrial psychology.* New York: Harper and Row.

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

Borman, W.C. (1991). Job behavior, performance, and effectiveness. In M. Dunnette and L. Hough (Eds.), *Handbook of Industrial and Organizational Psychology* (2nd edition, Vol. 2, pp. 271-326), Palo Alto, CA: Consulting Psychologists Press.

Christal, R. E. (1974, January). *The United States Air Force occupational research project* (AFHRL-TR-73-75, AD-774 574). Lackland AFB, TX: Occupational Research Division , Air Force Human Resources Laboratory.

DeCotiis, T., & Petit, A. (1978). The performance appraisal process: A model and some testable propositions. *Academy of Management Review, 3*, 635-646.

Dickinson, T. L. (1993). Attitudes about performance appraisal. H. Schuler, J. L. Farr, and M. Smith (Eds.), *Personnel selection and assessment: Individual and organizational perspectives,* Hillsdale, NY: Lawrence Erlbaum.

Dipboye, R., & de Pontbriand, R. (1981). Correlates of employee reactions to performance appraisals and appraisal systems. *Journal of Applied Psychology, 66*, 248-251.

Dobbins, G. H., Cardy, R. L., & Platz-Vieno, S. J. (1990). A contingency approach to appraisal satisfaction: An initial investigation of the joint effects of organizational variables and appraisal characteristics. *Journal of Management, 16*, 619-632.

Giles, W. F., & Mossholder, K. W. (1990). Employee reactions to contextual and session components of performance appraisal. *Journal of Applied Psychology, 75*, 371-377.

Greenberg, J. (1986a). Determinants of perceived fairness of performance evaluations. *Journal of Applied Psychology, 71*, 340-342.

Greenberg, J (1986b). The distributive justice of organizational performance evaluations. In H.W. Biefhoff, R.L. Cohen, & J. Greenberg (Eds.), *Justice in social relations* (pp. 337- 351), New York: Plenum.

Hedge, J. W. (1983, August). *A focus on global measures of appraisal system success/failure*. Paper presented at the annual meeting of the Academy of Management, Dallas.

Hedge, J.W., & Teachout, M.S. (1986, November). *Job performance measurement: A systematic program of research and development* (AFHRL-TP-86-37, AD-A174 175). Brooks AFB, Texas: Air Force Human Resources Laboratory.

Iaffaldano, M. R., & Muchinsky, P. M. (1985). Job satisfaction and job performance: A meta-analysis. *Psychological Bulletin, 97*, 251-273

Jacops, R., Kafry, D., & Zedeck, S. (1980). Expectations of behaviorally anchored rating scales. *Personnel Psychology, 33*, 595-640.

Kavanagh, M.J. (1980). *Criteria for the evaluation of performance measurement techniques and performance systems.* Paper presented at the First Annual Scientist-Practitioner Conference in Industrial/Organizational Psychology, Virginia Beach, VA.

Kavanagh, M. J. (1982). Evaluating performance. In K. M. Rowland & G. R. Ferris (Eds.), *Personnel management*. Boston, MA: Allyn & Bacon.

Kavanagh, M. J., & Hedge, J. W. (1983, May). *A closer look at correlates of performance appraisal system acceptability*. Paper presented at the annual meeting of the Eastern Academy of Management, Pittsburg, PA.

Kavanagh, M. J., Hedge, J. W., Ree, M., Earles, J., & DeBiasi, G. L. (1985, May). *Clarification of some issues in regard to employee acceptability of performance appraisal: Results from five samples*. Paper presented at the annual meeting of the Eastern Academy of Management, Albany, NY.

Landy, F. J., Barnes-Farrell, J. R., & Cleveland, J. N. (1980). Perceived fairness and accuracy of performance evaluation: A follow-up. *Journal of Applied Psychology, 65,* 355-356.

Landy, F. J., Barnes, J. R., & Murphy, K. R. (1978). Correlates of perceived fairness and accuracy of performance evaluation. *Journal of Applied Psychology, 63,* 751-754.

Lawler, E. E. (1967). The multitrait-multirater approach to measuring managerial job performance. *Journal of Applied Psychology, 51,* 369-381.

Leventhal, G. S. (1976). The distribution of rewards and resources in groups and organizations. In L. Berkowitz and E. Walster (Eds.), Advances in experimental social psychology (Vol. 9). New York: Academic Press.

Lipscomb, M. S., & Dickinson, T. L. (1988, June). The Air Force domain specification and sampling plan. In M. S. Lipscomb & J. W. Hedge (Eds.), *Job performance measurement: Topics in the performance measurement of Air Force enlisted personnel* (AFHRL-TP-87-58, AD-A195 630). Brooks AFB, TX: Training Systems Division, Air Force Human Resources Laboratory.

McAfee, B., & Green, B. (1977). Selecting a performance appraisal method. *The Personnel Administrator, 22,* 61-64.

Murphy, K. R., Cleveland, J. (1991). *Performance appraisal: An organizational perspective.* Boston, MA: Allyn & Bacon.

O'Connor, E. J., Peters, L., Rudolf, C. J., & Pooyan, A. (1982). Situational constraints and employee affective reactions: A field replication. *Group and Organizational Studies, 7,* 418-428.

Olson, D. M., & Borman, W. C. (1989). More evidence on relationships between the work environment and job performance. *Human Performance, 2,* 113-130.

Peters, L. H., & O'Connor, E. J. (1980). Situational constraints and work outcomes: The influence of a frequently overlooked construct. *Academy of Management Review, 5,* 391-397.

Podsakoff, P. M., & Williams, L. J. (1986). The relationship between job performance and job satisfaction. In E. A. Locke (Ed.), *Generalizing from laboratory to field settings.* Lexington, MA: Lexington.

Teachout, M.S., & Pellum, M.W. (1991, February). *Air Force research to link standards for enlistment to on-the-job performance* (AFHRL-TR-90-90). Brooks AFB,TX: Training Systems Division, Air Force Human Resources Laboratory.

Thibaut, J., & Walker, L. (1975). *Procedural justice: A psychological analysis.* Hillsdale, NJ: Erlbaum.

Weitz, J. (1961). Criteria for criteria. *American Psychologist, 16,* 228-232.