AL-TP-1993-0010

ARMSTRONG LABORATORY

DETERMINANTS OF PERFORMANCE RATING ACCURACY: A FIELD STUDY

Mark S. Teachout

HUMAN RESOURCES DIRECTORATE
TECHNICAL TRAINING RESEARCH DIVISION
7909 Lindbergh Drive
Brooks Air Force Base, TX 78235-5352

Terry L. Dickinson

Old Dominion University
Norfolk, VA 23529

April 1993

Interim Technical Paper for Period January 1990 - December 1992

Approved for public release; distribution is unlimited.

AIR FORCE MATERIEL COMMAND
BROOKS AIR FORCE BASE, TEXAS

NOTICES

This technical paper is published as received and has not been edited by the technical editing staff of the Armstrong Laboratory.

When Government drawings, specifications, or other 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 any way 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
Project Scientist

Rodger D. Ballentine

RODGER D. BALLENTINE, Colonel, USAF
Chief, Technical Training Research Division

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	April 1993	Interim January 1990 – December 1992

4. TITLE AND SUBTITLE

Determinants of Performance Rating Accuracy: A Field Study

5. FUNDING NUMBERS

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

6. AUTHOR(S)

Mark S. Teachout
Terry L. Dickinson

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

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

8. PERFORMING ORGANIZATION REPORT NUMBER

AL-TP-1993-0010

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

10. SPONSORING MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The purpose of this research was to investigate the influence of rater and ratee characteristics, performance constraints, and rating system acceptability on the accuracy of supervisory performance ratings in a field setting. Participants were 212 raters and 405 ratees across three jobs in the United States Air Force. An hypothesized structural model of rating accuracy was tested using LISREL 7 to determine the relationships among nine latent variables. Although the goodness-of-fit statistics for the model were considered marginal, results indicated that motivation to rate accurately, trust in the appraisal process, rating form acceptability, rater cognitive ability, rater experience, and ratee experience were related to rating accuracy. Interpretations and suggestions for future research were discussed.

14. SUBJECT TERMS			15. NUMBER OF PAGES
Cognitive ability	Performance measurement	Supervisory ratings	98
Criterion development	Rating acceptability	Work samples	16. PRICE CODE
Job experience	Rating accuracy		

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500

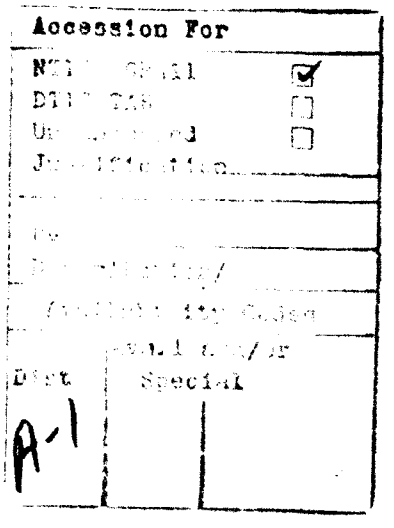
Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

CONTENTS

TABLE OF CONTENTS (concluded)

LIST OF FIGURES

LIST OF TABLES



PREFACE

During the 1980's the Armstrong Laboratory (AL\HR) developed a Job Perfromance Measurement System (JPMS) for the purpose of conducting research and development in the areas of selection test validation and training evaluation. The JPMS was developed and data were collected for eight Air Force specialties. Performance ratings were an important part of this measurement system. The present research examined the effects of rater and ratee characteristics, performance constraints, and rating system acceptability on the accuracy of supervisory performance ratings in a field setting across three of the AFSs.

This report documents preliminary work conducted on modeling factors that influence rating quality in a field setting. The work was conducted under in-house work unit No. 1121-1200, as the doctoral dissertation of the first author under the guidance of the second author. The authors are grateful to the other committee members, Dr Glynn Coates, Dr Al Glickman and Dr Jerry Hedge for their comments and support of this effort.

DETERMINANTS OF PERFORMANCE RATING ACCURACY: A FIELD STUDY

SUMMARY

This research examined the influence of rater, ratee and rating system characteristics on the accuracy of supervisory ratings across three Air Force Specialties. The hypothesized structural model was confirmed, indicating that each characteristic had a direct or indirect impact on accuracy. Practical implications and future research directions are discussed.

I. INTRODUCTION

Supervisory ratings are the most common method of assessing worker performance (Landy & Farr, 1980) because they can be used for a wide variety of jobs and appraisal functions. Given their widespread use and importance to human resource decision making, researchers must provide evidence of rating quality. Although there have been at least 25 different standards applied to performance ratings by various researchers over the years (Bernardin & Beatty, 1984), accuracy is considered to be the most important index of rating quality (Borman, 1979a; Dickinson, 1987; Kavanagh, Borman, Hedge, & Gould, 1986). In fact, Ilgen and Feldman (1983) termed accuracy the "ultimate goal" for determining the effectiveness of a performance appraisal system. Clearly, ratings must be

accurate if they are to play a role in enhancing organizational effectiveness.

Background

Research on the accuracy of performance ratings has focused on training raters in a laboratory setting to improve the accuracy of their ratings (e.g., Borman, 1979a; Dickinson & Silverhart, 1986; McIntyre, Smith, & Hassett, 1984; Pulakos, 1984). In these studies videotapes of ratee performance were rated and accuracy statistics computed to describe the success of training. Only a few research studies have investigated other variables that may account for accuracy in performance ratings (e.g., Borman, 1979b, 1980; Smithers & Reilly, 1987).

Several models suggest other variables that are likely to influence the accuracy of ratings (DeCotiis & Petit, 1978; DeNisi, Cafferty, & Meglino, 1984; Kavanagh et al., 1986; Landy & Farr, 1980). These variables can be thought of as "filters" through which performance information must pass (Landy & Farr, 1980) to effect rating accuracy. For example, Kavanagh et al. (1986) describe input and process variables that impact performance measurement quality (see Figure 1). Input variables include rater characteristics (e.g., rater cognitive ability, job experience), ratee characteristics (e.g., job experience), and perceived performance

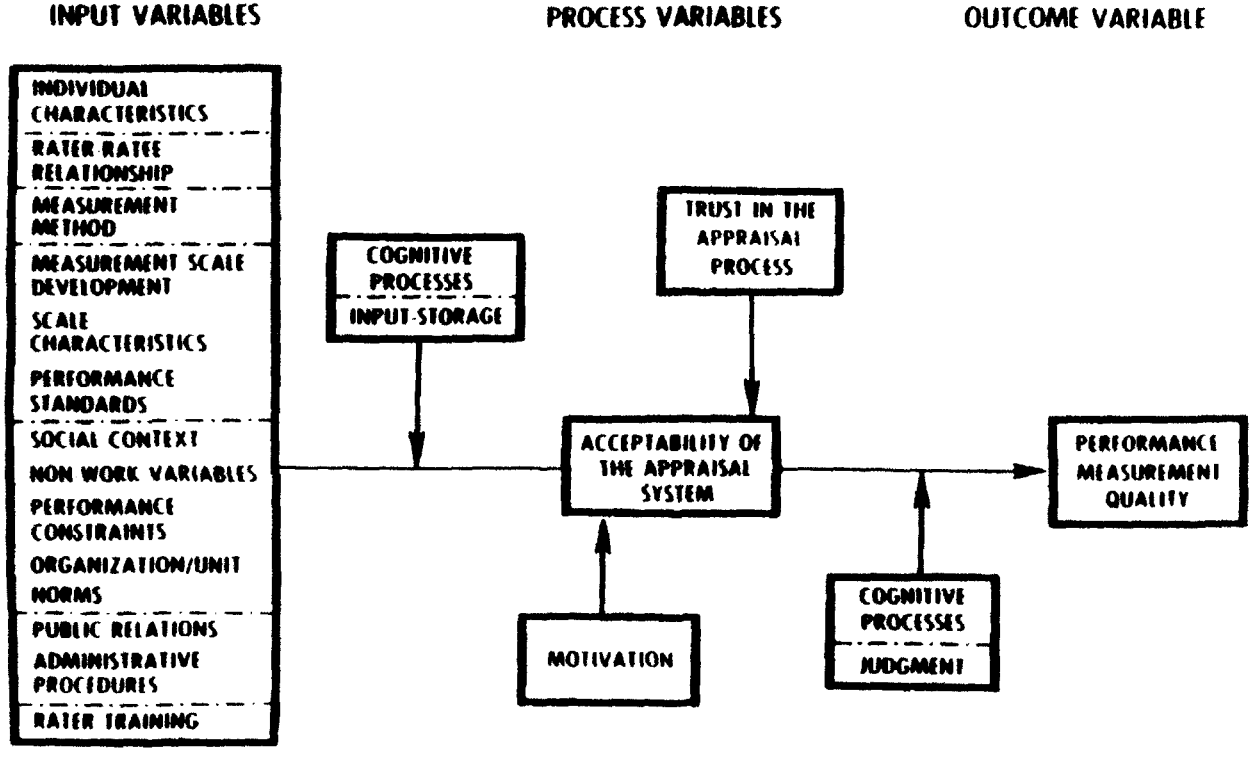


Figure 1. A job performance measurement classification scheme for validation research (From Kavanagh et al., 1986).

constraints (e.g., availability of tools, equipment, and manuals). Process variables involve acceptability of the appraisal system (e.g., acceptability of the rating instrument and procedures, motivation to rate, and trust in the appraisal process). To date, these variables have not been studied simultaneously in a research investigation. The present research examines the effects of these variables on the accuracy of supervisory performance ratings in a field setting.

Rater Characteristics

Although research on rater characteristics has not been systematic, the evidence available suggests that some characteristics influence performance ratings (Landy & Farr, 1983) including demographic factors (e.g., age, sex), individual difference characteristics (e.g., cognitive ability, personality), job-related factors (e.g., job experience, leadership style), and extent and type of rater training. For the purposes of this research, rater cognitive ability and job experience will be considered as characteristics that influence ability to rate.

Rater cognitive ability. Performance appraisal can be viewed as a cognitive task that requires the acquisition, organization, storage, retrieval, and integration of information (DeNisi et al., 1984). A simplified version of this model is displayed in Figure 2.

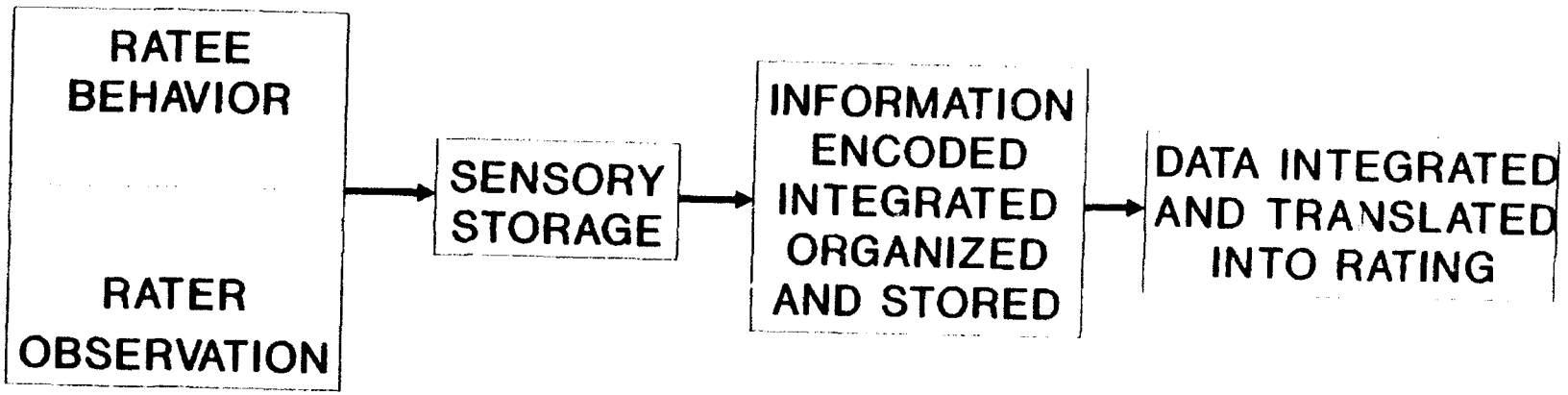


Figure 2. A simplified model of the performance appraisal process (Adapted from DeNisi et al., 1984).

Logically, individual differences in cognitive ability are important to the extent that they influence (improve or distort) the quality of information at any point in the process of making a performance rating. Of course, these individual differences would subsequently influence the accuracy of the final judgment. Although researchers (DeNisi et al., 1984; Feldman, 1981; Landy & Farr, 1983; Wherry & Bartlett, 1982) have emphasized the role of cognitive processes in evaluating others and have recognized the importance of understanding how raters process information (e.g., Bernardin, Cardy, & Carlyle, 1982; Borman, 1983; Murphy & Balzer, 1986), relatively little attention has been given to individual difference correlates of rating accuracy.

The major study investigating the relationship between individual differences and rating accuracy was conducted by Borman (1979b). He developed videotapes and "true scores" (i.e., target scores) of performance effectiveness for two jobs -- recruiting interviewer and manager. Borman found that 12 personal characteristics correlated significantly with rating accuracy, with the highest correlations between accuracy and intelligence, personal adjustment, and detail orientation. Importantly, the 12 characteristics accounted for 17% of the variance in accuracy, suggesting that individual differences play a significant role in determining rating accuracy.

Similarly, Smither and Reilly (1987) utilized videotapes and "true scores" and found that rater intelligence, as measured by the Wesman Personnel Classification Test, was positively related to rating accuracy.

Two additional studies support the relationship between cognitive variables and rating accuracy. Cardy and Kehoe (1984) used a measure of field dependence/ independence to categorize individuals in terms of selective attention ability. Cardy and Kehoe found that raters high on selective attention ability (i.e., more field independent) provided more accurate ratings than raters low on selective attention ability (i.e., more field dependent). In another study, Mount and Thompson (1987) measured the extent to which subordinates perceived their managers to be performing role responsibilities in the way the subordinates believed they should be performed. Results indicated that subordinate ratings of performance were more accurate when the behaviors of the manager were consistent with the expectations of the subordinate rater.

Several additional studies lend support to the notion that cognitive ability is related to rating quality. Schneider and Bayroff (1953) and Bayroff, Haggerty, and Rundquist (1954) found that high aptitude United States

Army trainees provided ratings of their fellow trainees that were more valid in predicting subsequent job performance than trainees with low aptitude test scores. Finally, Mullins and Force (1962) compared ratings to paper-and-pencil test scores and found evidence of a generalized ability of peers to rate their co-workers accurately across several performance dimensions.

Collectively, these studies indicate a relationship between general cognitive ability and rating quality that requires further examination.

Other cognitive variables. Few additional cognitive variables have been examined as rater attributes that affect rating quality, and then, the studies have examined their effect on rating errors and not on rating accuracy. For example, Schneier's (1977) theory of cognitive compatibility has received considerable attention. According to this theory, a rater's accuracy on a particular rating format (e.g., behaviorally anchored rating scale versus graphic rating scale) depends upon the rater's cognitive complexity. Cognitive complexity is the extent to which a rater has the ability to perceive behavior in a multidimensional fashion. Presumably, cognitively complex raters are more capable of perceiving, storing, and recalling information about others compared to cognitively simple raters. Schneier found that cognitively complex raters made fewer rating errors and

were more confident about their ratings. However, subsequent studies (Bernardin et al., 1982; Lahey & Saal, 1981; Sauser & Pond, 1981) have failed to replicate or extend Schneier's findings. Together, the evidence indicates that cognitive complexity is not related to rating quality. However, as Lahey and Saal (1981) have noted, other cognitive variables are likely to be important in understanding performance judgments, but researchers have failed to identify or properly measure the critical ones.

Rater job experience. Rating incumbent performance can be considered one of many tasks that a supervisor is asked to perform. Since general job experience is related to performance across a wide range of tasks, jobs, and lengths of service (Gordon & Johnson, 1982; Hunter & Hunter, 1984; McDaniel, Schmidt, & Hunter, 1988; Schmidt, Hunter, & Outerbridge, 1986), it should also improve rating performance. Raters need to know requirements of the ratee's job and have knowledge of ratee's job performance in order to evaluate that performance accurately. Increased job experience should provide raters with this information.

Several studies indicate that rater job experience improves rating quality, although results have been mixed. While Mandell (1956) found that raters with more than four

years of experience as supervisors tended to be more lenient in their ratings than raters with less experience, and Jurgensen (1950) determined that more experienced raters provided more reliable ratings, Klores (1966) obtained no significant effect of rater experience. Further, Cascio and Valenzi (1977) found a significant effect for rater experience in a study conducted for research purposes only, but experience only accounted for a small percentage of the rating variance. Finally, Huber, Neale, and Northcraft (1987) found that the relationship between objective performance (as displayed in "paper people" scenarios) and overall performance ratings was greater for raters with more tenure.

Overall, rater job experience as measured in these studies (e.g., job tenure) appears to have a small, but positive effect on rating quality. Apparently, rater job tenure is a general measure of job experience that increases knowledge of performance requirements, standards and procedures, and knowledge of ratee performance.

Ratee Characteristics

Ratee characteristics have also been shown to influence the quality of performance ratings (Landy & Farr, 1980). Ratee job experience will be considered in this research.

Ratee job experience. Ratee job experience is a characteristic that should influence rating quality, since it creates a preconceived notion (DeNisi et al., 1984) or frame-of-reference (Wyer & Srull, 1980) that the rater uses to interpret information and make performance judgments. In addition, the longer a ratee has been with an organization, the greater the likelihood the rater has had the opportunity to observe relevant job behaviors and make more accurate ratings.

Although some studies have found no relationship (Klores, 1966; Schwab & Heneman, 1978) or a negative relationship (Schneier & Beusse, 1980; Svetlik, Prien, & Barrett, 1964) between ratee experience and performance ratings, other studies (Bass & Turner, 1973; Cascio & Valenzi, 1977; Jay & Copes, 1957; Vance, Coovert, MacCallum, & Hedge, 1989; Zedeck & Baker, 1972) suggest that job experience and performance ratings are positively, but weakly correlated. For example, in a study where ratee experience ranged from 8 to 48 months, Vance et al. (1989) found that incumbents with more tenure were rated more highly in three separate structural models of task ratings (i.e., aggregated self, peer, and supervisor ratings).

In contrast, to Vance et al., (1989), Schmidt et al., (1986) found that when rating workers with up to five years of experience, supervisors do not rely on ratee job

experience, but rather on job knowledge and work sample performance capabilities. However, in evaluating workers with more than five years of experience, supervisors gave higher ratings than workers merited in comparison to job knowledge and work sample performance. Perhaps ratees with more experience are given a "bonus" for their experience, or the "benefit of the doubt" with respect to their performance, since the longer an individual has been a member of an organization, the better he or she is expected to perform. These results suggest that at some point, increasing amounts of ratee experience could contribute to rating inaccuracy.

Performance Constraints

The theoretical framework developed by Peters and O'Connor (1980) indicates that situational constraints impact performance, as well as job satisfaction and organizational withdrawal. These constraints could also affect rating quality. Research has shown that specific situational factors can limit individual work performance in laboratory experiments (e.g., Peters, O'Connor, & Rudolf, 1980), as well as in civilian (e.g., O'Connor, Peters, Pooyan, Weekley, Frank, & Erenkrantz, 1984), and military settings (Broedling, Crawford, Kissler, Mohr, Newman, White, Williams, Young, & Koslowski, 1980; Kane, 1979, 1981; O'Connor, Eulberg, Peters, & Watson, 1984).

Eulberg, O'Connor, Peters, and Watson (1984) summarized 14 categories of situational constraints that include the availability of job-related information (e.g., technical data), tools, equipment, materials, supplies, and parts. The appropriate technical manuals and the correct tools, equipment, parts, and supplies are necessary to achieve maximum job proficiency. Although the number and type of constraints are likely to differ across organizational and work environments, these constraints can be recognized, differentiated, and verified by job incumbents (Eulberg et al., 1983). However, supervisors may fail to recognize, or may misperceive the extent to which these constraints interfere with ratee job performance, and consequently, supervisors may assign inaccurate ratings. However, constraints are likely not to influence rating quality for raters with greater cognitive ability or more job experience (Mitchell & Kalb, 1982), because these raters have greater ability to assess the work environment and the impact of constraints on performance.

In sum, there may be dysfunctional consequences of situational constraints for rating accuracy, but these constraints may be ameliorated by rater characteristics.

Appraisal System Acceptability

Although acceptance of a personnel procedure is crucial to its effective use, acceptability is a relatively uninvestigated criterion for judging the quality of a measurement system. Lawler (1967) was the first to note that attitudes toward performance ratings could affect their quality and developed a model of the factors that affect the construct validity of ratings. He proposed 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.

Several studies (Dipboye & de Pontbriand, 1981; Hedge, 1983; Hedge, Teachout, & Dickinson, 1987; Kavanagh & Hedge, 1983; Kavanagh, Hedge, Ree, Earles, & DeBiasi, 1985; Landy, Barnes, & Murphy, 1978; Landy, Barnes-Farrell, & Cleveland, 1980) have investigated Lawler's proposal. Landy et al., (1978) identified four 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 spuriously affect these relationships (Landy et al., 1980).

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

A series of studies by Kavanagh and Hedge (Hedge, 1983; Kavanagh & Hedge, 1983; Kavanagh et al., 1985) extended the examination of acceptability to users of performance appraisal systems in order to clarify and expand the nature of acceptability and its predictors. Importantly, organizational and appraisal system factors accounted for 81% of the total variance in ratings. Although users did not differentiate between the appraisal instrument and appraisal process, several attitudes toward the appraisal system were significant predictors of appraisal acceptability. These included attitudes about fair and accurate appraisals, performance differences between workers, a satisfactory evaluation, satisfactory feedback, and clear performance standards.

Hedge et al. (1987) examined the concept of user acceptability of the appraisal process and correlates of

user acceptability as criteria for choosing among several rating formats in a research purposes only environment. They found that acceptability is a complex, multifaceted construct involving perceptions of appraisal fairness, clarity of format instructions, ability to discriminate performance with the format, accuracy of ratings, and confidence in ratings. In addition, rater motivation and trust in the appraisal process accounted for substantial variance (i.e., 38%) in user acceptability.

Unfortunately, rater motivation has received little attention by performance appraisal researchers. Although DeCotiis and Petit (1978) consider rater motivation to be an important component in the performance appraisal process, only Taft's (1971) theory of interpersonal judgments is cited to support this contention. More recently, Bernardin and colleagues (Bernardin & Cardy, 1982; Bernardin, Orban, & Carlyle, 1981) have studied rater motivation, but only in terms of the trust raters have in the appraisal process.

In spite of the lack of research, appraisal system acceptability is important because raters must not only be capable, but they must also be willing to provide accurate ratings (Banks & Murphy, 1985). In this regard, acceptability of the rating instrument and process, motivation to rate, and trust in the procedures, purpose,

and appraisal process are necessary (although not sufficient) conditions for obtaining accurate ratings.

Hypothesized Model of Rating Accuracy

The present research examined the influence of rater characteristics, ratee characteristics, performance constraints, and appraisal acceptability on the accuracy of performance ratings across three jobs in a field setting. Based on the preceding review, the model displayed in Figure 3 was hypothesized and tested. Nine latent variables (encircled) are depicted. Each was assessed by one or more measured variables. Three of the latent variables are criteria and include: (a) work sample performance, represented by the average score on a hands-on work sample test designed to measure maximum job proficiency on several job tasks; (b) supervisory ratings, represented by the average rating of performance on the same tasks measured by the work sample test; and (c) acceptability of rating forms, represented by rater perceptions of the acceptability of the task rating forms. Six latent variables were hypothesized to influence supervisory ratings directly or indirectly, and hence, indirectly influence rating accuracy as indicated by the relationship between supervisory ratings and work sample performance. These six variables are: (a) rater cognitive ability, measured by a composite of three subtest scores of the Armed Services Vocational Aptitude

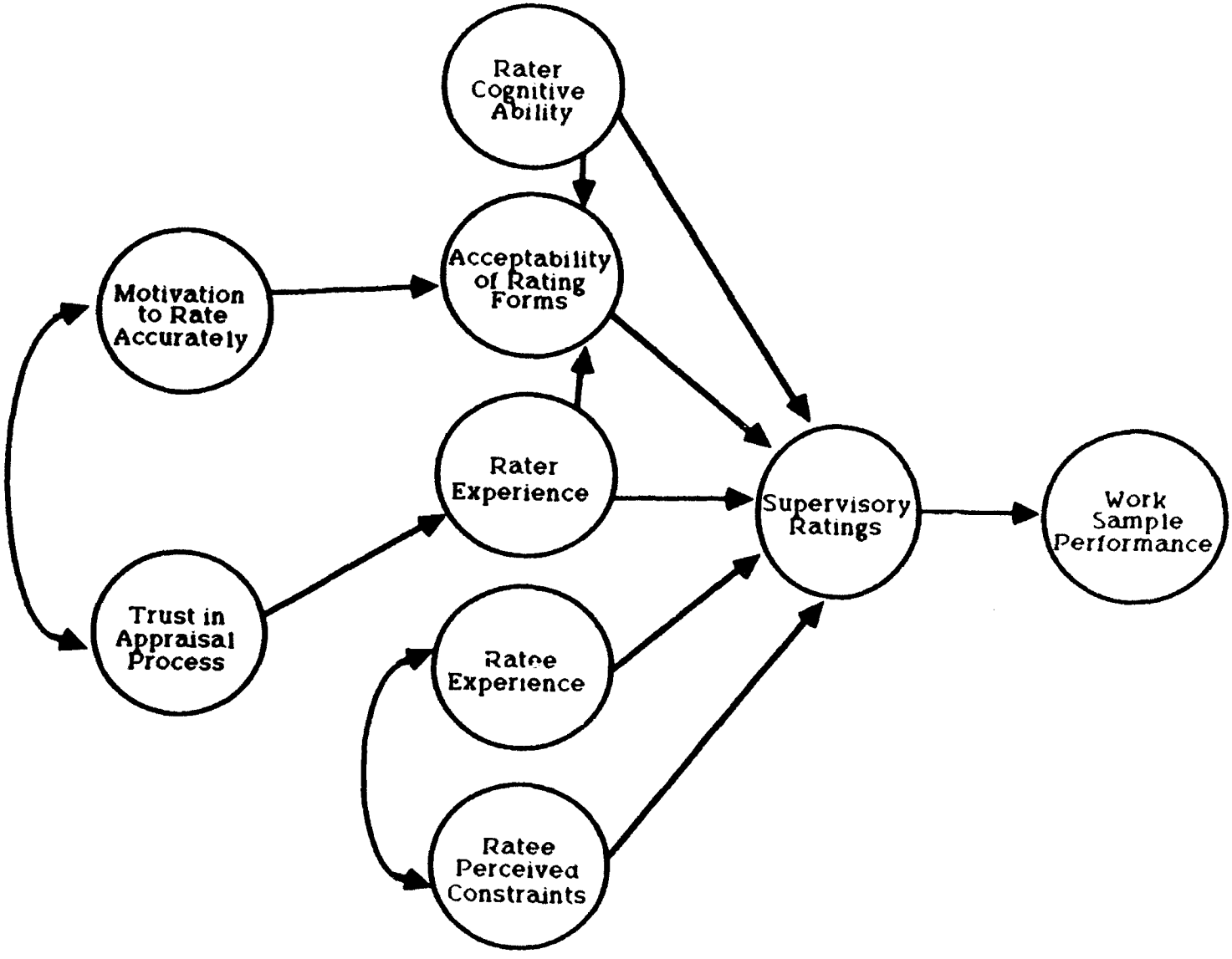


Figure 3. An hypothesized model of rating accuracy.

Battery; (b) rater job experience, represented by months on the job; (c) motivation to rate, represented by rater perceptions; and (d) trust in the appraisal process, represented by rater perceptions; (e) ratee experience, represented by months on the job; and (f) situational constraints, represented by ratee perceptions of constraints to job performance.

The latent predictor variables were hypothesized to account for variance in one or more of the latent criterion variables, either directly or indirectly. Motivation to rate accurately and trust in the appraisal process were hypothesized to relate directly to rating form acceptability based on Hedge et al. (1987), and they were also expected to correlate (as indicated by the curved line in Figure 3). Rater cognitive ability and rater job experience were hypothesized to relate directly to rating form acceptability and supervisory ratings. Kavanagh et al. (1986) suggest (see Figure 1) that cognitive processes play a key role in the input and storage of information as well as during the rating judgment. For this reason, it was expected that rater cognitive ability and rater experience relate to both of these criteria. Finally, ratee job experience and perceived constraints to performance were hypothesized to relate directly to supervisory ratings, and they were expected to correlate. Ratee experience was shown to be

directly related to task ratings for three models of task performance (Vance et al., 1989), and only weakly related ($r = .17$) to work sample performance for one of the models. Since Vance et al. examined a job that is part of the same program of research as the present research (see Hedge & Teachout, 1986), ratee experience was hypothesized to relate directly to supervisory ratings, but not to work sample performance.

In sum, this research investigated the simultaneous effects of rater and ratee characteristics, performance constraints, and appraisal acceptability on the accuracy of performance ratings in a field setting.

II. METHOD

Participants

Participants were 617 enlisted personnel in the United States Air Force who were located at 54 bases around the world. They included 87 raters and 178 ratees employed as air traffic control operators, 54 raters and 94 ratees employed as avionic communications specialists, and 71 raters and 133 ratees employed as information systems radio operators. Ratees had an average of 27 months in the Air Force; 99% of them were high school graduates; 80% were male; and 79% were caucasian. All raters were first-level supervisors who had an average of 128 months in the Air Force.

Work Sample Test Performance

A hands-on work sample test was constructed to assess incumbent job proficiency on tasks representative of the job. The domain of tasks for each job is identified and defined in the Air Force Occupational Survey data base (Christal, 1974). A domain task sampling plan was developed (Lipscomb, 1984), and tasks were sampled with stratified random sampling procedures (Lipscomb, 1984; Lipscomb & Dickinson, 1987). Test developers used technical orders and manuals (i.e., descriptions of work procedures), as well as input from subject matter experts (SMEs) to define and describe the procedural steps required for successful task completion. A hands-on work

sample test was constructed for each task, reviewed by SMEs, and field tested at several Air Force bases. A "yes/no" format was used to score each step to be performed in a test. The proportion of steps performed correctly was calculated for each test and averaged across hands-on task scores on the work sample tests. This average was used as a measure of work sample performance. Examples of the hands-on work sample tests are contained in Appendix A.

Test Administrator Training

The work sample tests were administered to job incumbents by active-duty noncommissioned officers who had extensive work experience in the jobs tested. These administrators received two weeks of observation and scorer training (Hedge, Lipscomb, & Teachout, 1988). This type of training has been shown to produce accurate and reliable test administrator scoring (Hedge, Dickinson, & Bierstedt, 1985). Hedge et al. (1985) calculated scorer agreement and correlational accuracy indices between test administrator scores and videotape target scores. Average inter-scorer agreement ($\bar{r}$= .81) and accuracy ($\bar{r}$= .85) were quite high.

In the present research, videotapes of work sample test performance with known target scores were also used as a training device to improve observational skills.

After viewing and scoring the videotapes, the administrators engaged in detailed discussions to identify the key behaviors that an incumbent should perform and avoid for successful task completion.

In addition, a technique referred to as "shadow scoring" was used during data collection in the field. In this technique, two test administrators independently observed and scored an individual performing a task. The technique was effective in maintaining agreement in the scoring of the work sample tests. The average scorer agreement was 95% across 58 individuals from the three specialties.

Rating Scales

Graphic rating scales were constructed to measure performance on the same tasks measured by the hands-on work sample tests. Each task was described by its statement from the Air Force Occupational Survey. Performance was rated on a 5-point adjectivally anchored scale ranging from 1, with an anchor of "never meets acceptable level of proficiency," to 5, with an anchor of "always exceeds acceptable level of proficiency." The rating scales used to rate task proficiency are contained in Appendix B.

Measures Influencing Accuracy

The information used to measure the constructs hypothesized to influence rater accuracy were obtained from two questionnaires and personnel records.

General background questionnaire. The general background questionnaire contained items measuring perceived performance constraints (e.g., technical manuals are clear and understandable). The three items used from this questionnaire are contained in Appendix C.

Rating form questionnaire. The rating form questionnaire contained items measuring acceptability of the rating process. The items evaluated motivation to rate accurately (e.g., motivation to complete forms), trust in the appraisal process (e.g., did others try to follow the rules), and acceptability of the rating forms (e.g., fairness; ease of use; discriminability). Based on previous research (Hedge et al., 1987), five items were selected to measure rating form acceptability, four items to measure motivation to rate accurately, and three items to measure trust in the appraisal process. These items are contained in Appendix D.

Personnel records. General measures of rater and ratee experience (i.e., months in the Air Force) were obtained from personnel records. In addition, the measure of rater cognitive ability was obtained from the Armed Services Vocational Aptitude Battery (ASVAB) which is used

for military enlistment and classification decisions. The ASVAB contains 10 subtests that reflect subject areas predictive of training criteria (United States Department of Defense, 1984). The Armed Forces Qualification Test (AFQT), a composite of three ASVAB subtests, was used as the measure of rater cognitive ability. The AFQT is used by all of the Armed Services as an indicator of general trainability. This composite score contains measures of verbal ability, mathematical knowledge, and arithmetic reasoning.

Procedure

Instruments were developed, and data were collected as part of a large-scale research and development project to validate selection and classification tests (Hedge & Teachout, 1986). In a group orientation session, the research project was described, participation conditions were explained, and all rating measures were shown and described to raters. This orientation was followed by one hour of frame-of-reference and rater error training (McIntyre et al., 1984). Two rating exercises facilitated understanding and use of rating forms by identifying performance behaviors at varying levels and by associating these behaviors with the rating-scale anchors. Participants practiced rating the performance of the incumbents described in the two exercises. Following

these ratings, they received target-score feedback about accuracy. In addition, a third exercise highlighted rating errors, and suggestions were made on how to improve rating accuracy.

Immediately following rater training, rating booklets were distributed, and raters were asked to complete all measures. The booklets were organized such that raters completed the general background questionnaire, followed by rating forms, and then the rating form questionnaire.

Subsequent to this group session, job incumbents were individually administered the work sample tests. Time limits were specified for each of these tests. The total time for administration of the tests was approximately four to seven hours, depending on the Air Force job.

Data Analyses

Data were analyzed using the PRELIS (Joreskog & Sorbom, 1988) and LISREL 7 computer programs (Joreskog & Sorbom, 1989). PRELIS is a preprocessor for LISREL. It was used to compute polychoric, polyserial, and Pearson product-moment correlations among continuous and ordinal variables. The polychoric correlation is an estimate of the correlation between two categorical variables assumed to have underlying continuous distributions, while the polyserial correlation is an estimmate of a correlation between a continuous and categorical variable. Thus, PRELIS produces the appropriate covariance matrix for

input into LISREL. In the present research, variables representing acceptability of the rating form, motivation to rate accurately, and trust in the appraisal process were considered ordinal variables, while the remaining variables were considered to be continuous.

LISREL simultaneously estimates two models: A measurement model that describes the relationships among latent variables and their observable indicators; and a structural model that defines causal relations among latent variables. The measurement and structural model parameters to be estimated are determined by the hypothesized model. These estimated parameters are termed free, while parameters not estimated are termed fixed. Initial estimates of free parameters are obtained (see Joreskog & Sorbom, 1989), followed by an iterative procedure to improve the initial estimates according to a maximum likelihood criterion. This analysis approach corrects for unreliability in the measured variables that represent the latent variables (Widaman, 1985), making it unnecessary to correct the relationships among the variables for attenuation. Whenever the reliability of the measured variables is known, or can reasonably be estimated, these values should be used to estimate the corresponding error variance. In this research, the reliabilities of the experience measures were considered

to be 1.0, because they were taken from personnel records. The estimated or known reliabilities of other variables were: .92 for the AFQT, which was the measure of rater cognitive ability, (Palmer, P., Hartke, D. D., Ree, M. J., Welsh, J. R., & Valentine, L. D., Jr., 1986); and .70 for work sample performance tests and for supervisor ratings (Kraiger, 1990).

The goodness-of-fit statistics associated with LISREL analyses indicate the likelihood that the hypothesized model could have produced the observed data. They reflect the extent to which the covariance matrix fitted to the observed variables (i.e., the covariances estimated by the model) reproduces the actual sample covariances among the observed variables. Three statistics were used to assess the fit of the model. First, the LISREL output provides an overall chi-square based on the difference between the observed and estimated covariance matrices. A non-significant chi-square is desired, since it indicates no significant difference between the observed and estimated matrices, and, hence, a good-fitting model. However, since the chi-square test is very sensitive to sample size, and will almost always reject a model on statistical bases (Bentler & Bonett, 1980), other methods have been developed to provide practical estimates of how well the model fits the data. A second goodness-of-fit statistic, rho (Bentler & Bonett,

1980) compares the ratio of the obtained chi-square relative to its degrees-of-freedom (df) with the same ratios representing two other models that serve as reference points: (a) the null, or worst case model, which hypothesizes that the measured variables are uncorrelated in the population; and (b) the idealized model, which may not really exist but which is a useful reference point and has an expected chi-square/degrees-of-freedom ratio of 1. Thus, rho describes where the hypothesized model lies on a continuum from the null to the idealized model. Rho should be equal to or greater than .90 to indicate a practical fit of the model to the data (Bentler & Bonett, 1980). The third goodness-of-fit statistic is the goodness-of-fit index (GFI) produced as part of the LISREL output. The GFI measures the fit of the model to the covariance matrix and requires a value of .90 or greater to indicate practical fit of the model.

In addition to goodness-of-fit information, LISREL provides detailed information about the fit of individual parameters in the model. A squared multiple correlation is produced for each measured variable as an indication of the proportion of variance in the measured variable predicted by the latent variable. A large squared multiple correlation indicates that the measured variables provide a good measure of the latent variable. In

addition, LISREL provides tests of significance for the parameter estimates (i.e., T-values), using approximate standard errors of those estimates, and modification indices (MI) indicating expected overall improvement in the chi-square measure of fit of the model resulting from allowing fixed parameters to be free. Model modifications constitute a specification search (MacCallum, 1986). A specification search is a sequential process of modifying a model to improve its fit and/or parsimony. This process involves adding new parameters to the original model, or deleting variables. These decisions are usually based upon inspection of the modification indices produced as part of the LISREL output. A specification search with excessive modifications can lead to overfitting and capitalization on chance. Therefore, such searches are considered exploratory and resulting models should be cross-validated before validity can be inferred (Bentler, 1980; Cliff, 1983; Cudeck & Brown, 1983). More important than the statistics used to evaluate model fit is the theoretical basis used by the investigator. Modifications should be made only if they are substantively meaningful and theoretically justified.

III. RESULTS

Overview

The original model depicted in Figure 3 fit the data marginally. A specification search (MacCallum, 1986) led to modifications of this model. Although modifications improved the model fit, and most of the hypothesized relationships were retained, the modified model was still marginal. However, several paths indicated significant relationships among latent traits and these were interpreted.

Hypothesized Model

The test of the hypothesized accuracy model produced a significant chi-square equal to 723.48 (df = 163; $p < .01$) with marginal indices of fit (GFI = .853, rho = .75). A specification search was conducted, because several hypothesized paths were not significant (i.e., T-values were less than 2.0), and because several modification indices indicated that the fit of the model could be improved. Inspection of the squared multiple correlations and measurement model loadings for perceived constraints items indicated that the perceived constraints construct was poorly measured. In addition, there were nonsignificant paths for perceived constraints with supervisory ratings and incumbent experience. Therefore, this construct was eliminated from further consideration.

The subsequent solution indicated that an item from trust in appraisal process should be eliminated, because its measurement model loading indicated that the item measured the trust construct poorly. Successive solutions indicated nonsignificant paths from rater cognitive ability to rating form acceptability, and from rater experience to supervisory ratings. These paths were also eliminated. Finally, modification indices indicated that a path be allowed between ratee experience and work sample test performance. Although previous research indicated that the most viable path was from ratee experience to supervisory ratings (Vance et al., 1989), there has also been some support for a direct relationship between ratee experience and work sample performance (Schmidt et al., 1986; Vance et al., 1989). Thus, this change was deemed appropriate. These modifications produced a model with a significant, but reduced, chi-square of 527.15 (df = 100; p, < .01). Nonetheless, the indices of fit were still marginal (GFI = .863, rho = .79). The results of this model are displayed in Figure 4. The corresponding measurement model for the criterion and predictor variables are contained in Tables 1 and 2, respectively. The T-values for all path and maximum likelihood estimates were significant (p < .05).

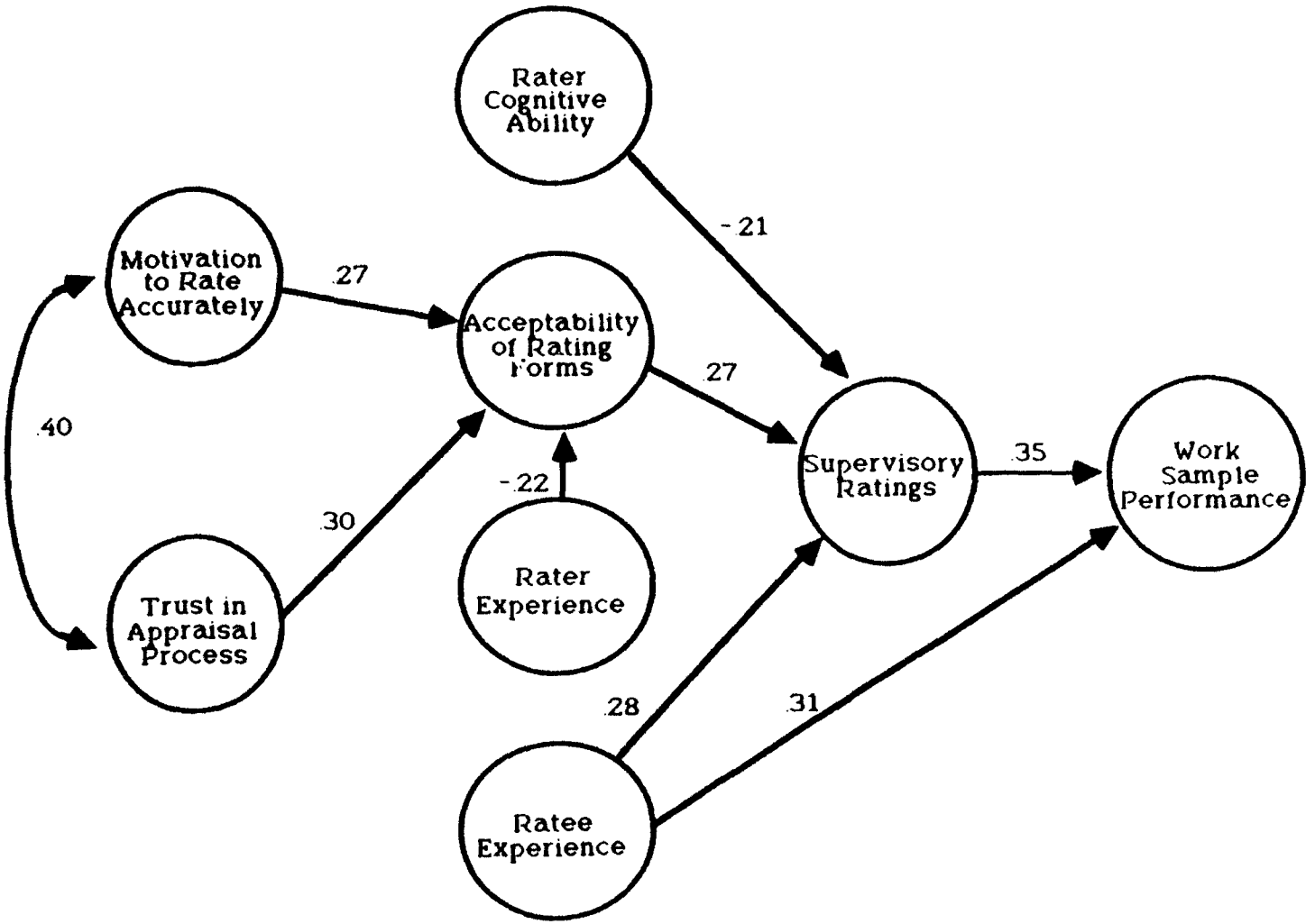


Figure 4. A modified model of rating accuracy

Table 1. Measurement Model Factor Loadings for the Criterion Variables

Measured Variable	Criterion latent variables: Rating Form Acceptability	Supervisory Ratings	Work Sample Performance
Fairness	.78		
Understandable	.60		
Discrimination	.65		
Accuracy	.68		
Acceptability	1.00[a]		
Sup Ratings		1.00[a]	
Work Sample Perf			1.00[a]

Note. Sup Ratings, Supervisory ratings; Work Sample Perf, Work sample performance.

[a]Variables were fixed to 1.00 to provide a scale for the factor loadings.

Table 2. Measurement Model Factor Loadings for the Predictor Variables

	Predictor latent variables				
Measured Variable	Rater Cog Abil	Rater Exp	Motiv to rate Accur	Trust in Appraisal Process	Ratee Exp
AFQT	1.00[a]				
Sup Months		1.00[a]			
Care			1.00[a]		
Importance			.99		
Satisfied			.80		
Acc in Project			.71		
Others Follow Instruct				.89	
Others Care about Accuracy				1.00[a]	
Ratee Months					1.00[a]

Note. Rater Cog Abil, Rater cognitive ability; Rater Exp, Rater experience; Motiv to rate Accur, Motivation to rate accurately; Ratee Exp, Ratee experience; Sup Months, Supervisor months in Air Force; Care, Care about accuracy; Importance, Importance of accuracy; Satisfied, Satisfied with accuracy; Acc in Project, Important of accuracy in this project; Others Follow Instruct, Others follow instructions; Ratee Months, Ratee months in Air Force.

[a]Variables were fixed to 1.00 to provide a scale for the factor loadings.

Results displayed in Figure 4 indicate that most of the hypothesized paths among latent variables were retained. Motivation to rate accurately and trust in the appraisal process were directly and positively related to rating form acceptability, and they were correlated. Rater cognitive ability and rater experience were hypothesized to relate directly to rating form acceptability and to supervisory ratings. However, rater cognitive ability was negatively related to supervisory ratings and did not relate to rating form acceptability, while rater experience was negatively related to rating form acceptability and not directly to supervisory ratings. Ratee job experience was positively related to supervisory ratings and work sample performance. The latent variable of perceived constraints was not included in this model, because it was not related to incumbent job experience or supervisor ratings. Finally, supervisory ratings were positvely related to work sample performance indicating that the ratings were accurate with respect to ordering individuals similarly to the work sample tests.

IV. DISCUSSION

Overview

The purpose of this research was to investigate the effects of rater and ratee characteristics, performance constraints, and appraisal acceptability on the accuracy of performance ratings in a field setting. Six latent variables were hypothesized to influence supervisory ratings, directly or indirectly, and to influence rating accuracy as indicated by the accuracy relationship between supervisory ratings and work sample performance. Although the overall fit of the model to the data was marginal, most of the hypothesized relationships were retained, and these relationships have important implications for performance rating accuracy.

Rating Accuracy

Supervisory ratings were positively related to the work sample test scores, supporting the hypothesis that these ratings are accurate with respect to the ordering of individuals similarly to the work sample test scores. For the purposes of the present research, relationships of the remaining constructs with the supervisory ratings are considered to be related positively or negatively to accuracy. However, these relationships should be considered indirect influences on rating accuracy.

Appraisal System Acceptability

The results clearly support the role of appraisal system acceptability as an important component of the rating process. Acceptability of the rating form is an important criterion that is influenced by the rater's motivation to rate accurately and trust in others involved in the rating process. The more supervisors were motivated to rate accurately and the more they trusted others involved in the rating process, the more acceptable they found the rating form. This, in turn, was positively related to the ratings assigned. Thus, motivation to rate accurately and trust in others indirectly affect rating quality through their impact on rating form acceptability. In addition, these two variables were linked in a correlative, but not causal fashion.

Until recently, rater's acceptance of a performance appraisal system had been ignored. However, Lawler's (1967) conceptual model hypothesized a link between employee's attitudes about the acceptability of a performance appraisal system and it's validity. Subsequent studies (Dipboye & de Pontbriand, 1981; Hedge, 1983; Hedge et al., 1987; Kavanagh & Hedge, 1983; Kavanagh et al., 1985; Landy et al., 1978; Landy et al., 1980) examined this concept further. However, the present research is the first to examine the link between appraisal system acceptability and rating quality in a

field setting. These results support that link. Moreover, the results support the conceptual framework proposed by Kavanagh et al. (1986) in which measurement system characterisitcs (e.g., motivation and trust) indirectly affect the quality of measurement through their impact on acceptability.

Future research should focus on other attitudes, reactions, and perceptions that are likely to influence rating quality. While the present research dealt with rater perceptions of acceptability (i.e., motivation, trust, and rating form acceptability), Kavanagh and Taber (1987) described additional correlates and causes of employee acceptability of performance appraisal such as situational favorability (e.g., rater training, purpose of appraisal explicit and observable, and required goal-setting) and supervisor-subordinate attributes (e.g., joint goal-setting, age, race, and sex congruence). These variables should also be examined in relation to rating quality.

Rater Characteristics

Rater cognitive ability and rater job experience were considered as characteristics that influence the ability to rate accurately. It was hypothesized that each of these characteristics would influence rating accuracy indirectly through rating form acceptability and supervisory ratings. Results indicated that rater

cognitive ability was related negatively to supervisory ratings but not related to rating form acceptability, while rater job experience was negatively related to rating form acceptability but not to supervisory ratings.

Rater cognitive ability. There are several explanations and interpretations for the results for rater cognitive ability. First, the negative relationship between rater cognitive ability and supervisory ratings indicates that raters with more ability assigned lower ratings. However, this does not mean that raters with more ability are necessarily less accurate. Since ratings are oftentimes considered lenient, lower ratings could be considered more realistic and an indication of greater accuracy for raters with more cognitive ability.

Secondly, the present research used a measure of general cognitive ability to represent that rater ability. This is in contrast to previous research that has typically included more specific cognitive variables such as personal judgment and detail orientation (Borman, 1979b), field dependence/independence (Cardy & Kehoe, 1984), perceived role congruence (Mount & Thompson, 1987), and cognitive complexity (Schneier, 1977). Nonetheless, general abilities are more predictive of performance criteria than specific abilities due to the cognitive processes involved in learning and thinking (Hunter, 1986). Thus, general cognitive ability should be a more

appropriate predictor of rating ability than specific cognitive variables because of the cognitive processes involved in making performance judgments.

Finally, these results indicate that rater cognitive ability influences supervisory ratings, but it does not influence perceptions of rating form acceptability. While previous research (e.g., Borman, 1979b) has demonstrated a positive relationship between rater cognitive ability and rating accuracy, no research has addressed the relationship between rater cognitive ability and aspects of the rating process. Although, the present research examined the relationship between rater cognitive ability and rating form acceptability and found that the ability is not related to this aspect of the rating process, future research should focus on the relationship of rater cognitive ability with other aspects of the rating process. Potential aspects include time pressure for completion of ratings, number of ratings to be completed, and effectiveness of rater training and instructions (Landy & Farr, 1983).

Rater experience. Rater experience was negatively related to rating form acceptability and unrelated to supervisory ratings. Thus, the more experienced the rater, the less acceptable the rating form. However, rater experience did not directly influence the ratings assigned. There are several plausible explanations for

these findings. Since more experienced supervisors have used the military performance appraisal system (AFR 39-62, 1989) more frequently, the performance rating approach used in the present research might have been less acceptable, because it represents a greater departure from the status quo for the more experienced supervisors than it was for the less experienced supervisors. In fact, the performance appraisal system was designed to be dissimilar to the military system, and this accounts for experience being negatively related to rating form acceptability.

The present research also indicates that rater job experience did not have a direct influence on the ratings assigned. Job experience was measured as job tenure, and this is a general measure that may not reflect specific knowledge of job requirements, standards and procedures, and knowledge of ratee performance. Perhaps this global index of experience was not adequate and sufficient in the present research to explain variance in ratings. Future research should consider developing specific measures of experience for knowledge of job requirements, standards, and procedures, and knowledge of ratee performance.

Finally, the present research is the first to examine the influence of rater experience simultaneously on two aspects of the rating process (i.e., perceptions of rating form acceptablity and the rating itself). While individual differences in rater experience influenced

perceptions of rating form acceptability there was no direct influence on the accuracy of the rating. Future research should examine why rater ability influences the final rating but not perceptions of the rating form, while experience influences attitudes and perceptions of the rating form but not the final rating.

Ratee Experience

Ratee experience was positively related to supervisory ratings and work sample test performance, indicating that the more experienced the incumbent, the higher the ratings that were assigned, the better the work sample performance and, hence, the more accurate the ratings. These results support previous studies (Bass & Turner, 1973; Cascio & Valenzi, 1977; Jay & Copes, 1957; Zedeck & Baker, 1972) that found a positive relationship between job experience and performance ratings. In addition, more experience increases the likelihood that the ratee performed the work sample previously as a job assignment, supporting the relationship of ratee experience to work sample performance. However, the present results are not consistent with findings by Schmidt et al. (1986) who found that job experience had no direct effect on supervisory ratings of job performance, but indirectly influenced ratings through job knowledge and work sample performance. The difference in findings

is likely due to the absence of job knowledge measures in the present research.

The results for ratee experience support the notion that experience provides a frame-of-reference (Wyer & Srull, 1980) against which the rater interprets information, forms impressions, and makes performance judgments. Future research should examine more specific indices of ratee experience that might influence performance ratings, such as the frequency and recency of performance on the job tasks to be rated.

Ratee Perceived Constraints

The construct of perceived constraints was measured poorly and was unrelated to ratee experience and supervisory ratings. Although no inferences can be made about the influence of perceived constraints on rating accuracy, the concept should still be considered viable. Fourteen categories of situational constraints have been described and categorized by researchers in this area (Eulberg et al., 1984). Perceived constraints is obviously a multifaceted construct that should be investigated in future research.

Further Research and Conclusions

Performance appraisal is a complex process that has been depicted and described in several different ways (e.g., Landy & Farr, 1983). Despite this complexity, much research has been conducted on different aspects of the

rating process that influence the quality of ratings, because appraisal information is important to human resource decision making. The present research is the first to examine several aspects of the rating process simultaneously in a field study.

The structural model of the process can only be considered marginally plausible, perhaps due to the complexity of the rating process, and the omission of other variables that might influence that process. Despite these marginal results, the relationships that were retained contribute to the understanding of factors that influence rating quality. Rater characteristics, ratee characteristics, and appraisal system acceptability all had a direct or indirect influence on the quality of supervisory ratings. This substantiates the importance of these constructs in understanding the rating process and indicates that further research is warranted. In this regard, several areas should be considered. First, the measurement of the constructs examined in the present research can be improved. More specific indices of rater and ratee experience can and should be used in addition to the general index of job experience (i.e., job tenure). Certainly, the situational constraints construct is multifaceted and should be measured differently in future research. Second, the use of work sample tests as target

scores in accuracy research should be encouraged. To date, accuracy research has been limited to laboratory study (cf. Sulsky & Balzar, 1988), but the present research suggests that rating accuracy can be investigated in field settings. Third, the results obtained here should be cross-validated. Final models resulting from specification searches must be cross-validated before validity can be confirmed (Bentler, 1980; Cliff, 1983; Cudeck & Browne, 1983). Fourth, after an accuracy model is cross-validated for supervisory ratings, the same should be accomplished for other sources of ratings (e.g., self-ratings). For example, research on the validity and accuracy of self-ratings has increased in the last decade due to the recognition that individuals can be a valuable source of information about their own performance (e.g., Mabe & West, 1982). A comparison of factors that influence the quality of supervisor versus self-ratings would appear fruitful. Finally, although the results reported here were part of a field study, the data were collected for research purposes only. Since the purpose of data collection has been shown to be a salient factor in performance appraisal research (e.g., McIntyre et al., 1984; Zedeck & Cascio, 1982), the extent to which the present results would change if data were collected for administrative purposes is a viable research question.

REFERENCES

AFR 39-62 (1989). The enlisted evaluation system. Department of the Air Force.

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

Bentler, P. M. (1980). Multivariate analysis with latent variables: Causal modeling. Annual Review of Psychology, 31, 419-456.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness fit in the analysis of covariance structures. Psychological Bulletin, 88, 588-606.

Bass, A. R., & Turner, J. N. (1973). Ethnic group differences in relationships among criteria of job performance. Journal of Applied Psychology, 57, 101-109.

Bayroff, A. G., Haggerty, H. R., & Rundquist, E. A. (1954). Validity of ratings as related to rating techniques and conditions. Personnel Psychology, 7, 93-114.

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

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., Cardy, R. L., & Carlyle, J. J. (1982). Cognitive complexity and appraisal effectiveness: Back to the drawing board? Journal of Applied Psychology, 6 151-160.

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 41st Annual Meeting of the Academy of Management, 311-315.

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

Borman, W. C. (1979b). Individual differences correlates of accuracy in evaluating others' performance effectiveness. Applied Psychological Measurement, 3, 103-115.

Borman, W. C. (1980). Performance judgments: The quest for accuracy in ratings of performance effectiveness. Paper presented at the First Annual Scientist-Practitioner Conference in Industrial-Organizational Psychology, Old Dominion University, Norfolk, VA.

Borman, W. C. (1983). Implications of personality theory and research for the rating of work performance in organizations. In F. Landy, S. Zedeck, & J. Cleveland (Eds.), Performance measurement and theory, (pp. 127-172). Hillsdale, NJ: Lawrence Erlbaum and Associates.

Broedling, L. A., Crawford, C. S., Kissler, G. D., Mohr, D. A., Newman, A. R., White, M. A., Williams, H., Young, L., & Koslowski, T. J. (1980). An examination of productivity impediments in the Navy industrial community (NPRDC-TR-81-2). San Diego, CA: Navy Personnel Research and Development Center.

Cardy, R. L., & Kehoe, J. F. (1984). Rater selective attention ability and appraisal effectiveness: The effect of a cognitive style on the accuracy of differentiation among ratees. Journal of Applied Psychology, 69, 589-594.

Cascio, W. F., & Valenzi, E. (1977). Behaviorally anchored rating scales: Effects of education and job experience of raters and ratees. Journal of Applied Psychology, 62, 278-282.

Christal, R. E. (1974). The United States Air Force occupational research project (AFHRL-TR-73-75). Lackland Air Force Base, TX: Occupational Research Division.

Cliff, N. (1983). Some cautions concerning the application of causal modeling methods. Multivariate Behavioral Research, 18, 115-126.

Cudeck, R., & Brown, M. W. (1983). Cross-validation of covariance structures. Multivariate Behavioral Research, 18, 147-187.

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

DeNisi, A., & Cafferty, T. P., & Meglino, B. M. (1984). A cognitive view of the performance appraisal process: A model and some research propositions. Organizational Behavior and Human Performance, 33, 360-396.

Dickinson, T. L. (1987). Designs for evaluating the validity and accuracy of performance ratings. Organizational Behavior and Human Decision Processes, 40, 1-21.

Dickinson, T. L., & Silverhart, T. A. (1986, August). Training to improve the accuracy and validity of performance ratings. Paper presented at the 94th annual convention of the American Psychological Association, Washington, DC.

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

Eulberg, J. R., O'Connor, E. J., Peters, L. H., & Watson, T. W. (1984). Situational constraints upon performance: A selective review of relevant literature. (AFHRL-TP-83-48). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Feldman, J. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. Journal of Applied Psychology, 66, 127-148.

Gordon, M. E., & Johnson, W. A. (1982). Seniority: A review of its legal and scientific standing. Personnel Psychology, 35, 255-280.

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, TX.

Hedge, J.W., Dickinson, T. L., & Bierstedt, S. A. (1988). The use of videotape technology to train administrators of walk-through performance testing (AFHRL-TP-87-71). Brooks AFB, TX: Training Systems Division, Air Force Human Resources Laboratory.

Hedge J. W., Lipscomb M. S., & Teachout, M. S. (1988). Work sample testing in the Air Force job performance measurement project. 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). Brooks AFB, TX: Training Systems Division, Air Force Human Resources Laboratory.

Hedge, J. W., & Teachout, M. S. (1986). Job performance measurement: A systematic program of research and development (AFHRL-TP-86-37). Brooks AFB, TX: Training Systems Division, Air Force Human Resources Laboratory.

Hedge, J. W., Teachout, M. S., & Dickinson, T. L. (1987, August). User acceptability as a criterion for choosing performance measures. In M. Secunda (Chair), New paradigms for evaluating performance appraisals. Symposium conducted at the annual meeting of the American Psychological Association, New York, NY.

Huber, V. L., Neale, M. A., & Northcraft, G. B. (1987). Judgment by heuristics: Effects of ratee and rater characteristics and performance standards on performance-related judgments. Organizational Behavior and Human Decision Processes, 40, 149-169.

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. Journal of Vocational Behavior, 29, 340-362.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. Psychological Bulletin, 96, 72-98.

Ilgen, D. R., & Feldman, J. M. (1983). Performance appraisal: A process focus. In B. M. Staw & L. Cummings (Eds.), Research in organizational behavior (Vol. 5). Greenwich, CT: JAI Press.

Jay, R., & Copes, J. (1957). Seniority and criterion measures of job proficiency. Journal of Applied Psychology, 41, 58-60.

Joreskog, K. G., & Sorbom, D. (1988). PRELIS: A preprocessor for LISREL (2nd ed.). Mooresville, IN: Scientific Software, Inc.

Joreskog, K. G., & Sorbom, D. (1989). LISREL 7 user's reference guide. Mooresville, IN: Scientific Software, Inc.

Jurgensen, C. E. (1950). Intercorrelations in merit ratings. Journal of Applied Psychology, 34, 240-243.

Kane, W. D. (1979). A heuristic model of Air Force maintenance performance (Final Report). Cullowee, NC: Western Carolina University, Southeast Center of Electrical Engineering Education.

Kane, W. D. (1981). Task accomplishment in an Air Force maintenance environment (80-0146). Bolling AFB, Washington, DC: Air Force Office of Scientific Research.

Kavanagh, M. J., Borman, W. C., Hedge, J. W., & Gould, R. B. (1986). Job performance measurement classification scheme for validation research in the miitary (AFHRL-TP-85-51). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

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

Kavanagh, M. J., Hedge, J. W., Ree, M., Earles, J., & Debiasi, G. L. (1984, 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.

Kavanagh, M. J., & Taber, T. (1987, August). Employee acceptability and rights in performance appraisal: Directions for research. In M. Secunda (Chair), New paradigms for evaluating performance appraisals. Symposium conducted at the annual meeting of the American Psychological Association, New York, NY.

Klores, M. S. (1966). Rater bias in forced-distribution ratings. Personnel Psychology, 19, 411-421.

Kraiger, K. (1990). Generalizability of performance measures across four Air Force specialties (AFHRL-TP-89-60). Brooks AFB, TX: Training Systems Division, Air Force Human Resources Laboratory.

Lahey, M. A., & Saal, F. E. (1981). Evidence incompatible with a cognitive compatibility theory of rating behavior. Journal of Applied Psychology, 66, 706-715.

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.

Landy, F. J., & Farr, J. L. (1980). Performance rating. Psychological Bulletin, 87, 87-102.

Landy, F. J., & Farr, J. L. (1983). The measurement of work performance: Methods, theory, and applications. New York: Academic Press.

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

Lipscomb, M. S. (1984, August). A task-level domain sampling strategy: A content valid approach. In S. Zedeck (Chair), Walk-Through Performance Testing: An innovative approach to work sample testing. Symposium conducted at the annual meeting of the American Psychological Association, Toronto, Canada.

Lipscomb, M. S., & Dickinson, T. L. (1988). 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). Brooks AFB, TX: Training Systems Division, Air Force Human Resources Laboratory.

Mabe, P. A., & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. Journal of Applied Psychology, 67, 280-296.

MacCallum, R. (1986). Specification searches in covariance structure modeling. Psychological Bulletin, 100, 107-120.

Mandell, M. M. (1956). Supervisory characteristics and ratings: A summary of recent research. Personnel Psychology, 32, 435-440.

McDaniel, M. A., Schmidt, F. L., & Hunter, J. E. (1988). Job experience correlates of performance. Journal of Applied Psychology, 73, 327-330.

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

Mitchell, T. R., & Kalb, L. S. (1982). Effects of job experience on supervisor attributions for a subordinate's poor performance. Journal of Applied Psychology, 67, 181-188.

Mount, M. K., & Thompson, D. E. (1987). Cognitive categorization and quality of performance ratings. Journal of Applied Psychology, 72, 240-246.

Mullins, C. J., & Force, R. C. (1962). Rater accuracy as a generalized ability. Journal of Applied Psychology, 46, 191-193.

Murphy, K. R., & Balzer, W. K. (1986). Systematic distortions in memory-based behavior ratings and performance evaluations: Consequences for rating accuracy. Journal of Applied Psychology, 71, 39-44.

O'Connor, E. J., Eulberg, J. R., Peters, L. H., & Watson, T. W. (1984). Situational contraints in the Air Force: Identification, measurement, and impact on work outcomes (AFHRL-TP-84-10). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

O'Connor, E. J., Peters, L. H., Pooyan, A., Weekley, J., Frank, B., & Erenkrantz, B. (1984). Situational constraint effects on performance, affective reactions and turnover: A field replication and extension. Journal of Applied Psychology, 69, 663-672.

Palmer, P., Hartke, D. D., Ree, M. J., Welsh, J. R., & Valentine, L. D. (1986, March). Armed Services Vocational Aptitude Battery (ASVAB): Alternate forms reliability (Forms 8, 9, 10, and 11) (AFHRL-TP-87-48). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

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

Peters, L. H., O'Connor, E. J., & Rudolf, C. J. (1980). The behavioral and affective consequences of performance-relevant situational variables. Organizational Behavior and Human Performance, 25, 79-96.

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

Sauser, Jr., W. I., & Pond, S. B. (1981). Effects of rater training and participation on cognitive complexity: An exploration of Schneier's cognitive reinterpretation. Personnel Psychology, 34, 563-577.

Schmidt, F. L., Hunter, J. E., & Outerbridge, A. N. (1986). Impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. Journal of Applied Psychology, 71, 432-439.

Schneider, D. E., & Bayroff, A. G. (1953). The relationship between rater characteristics and validity of ratings. Journal of Applied Psychology, 37, 278-280.

Schneier, C. E. (1977). Operational utility and psychometric characteristics of behavioral expectation scales: A cognitive reinterpretation. Journal of Applied Psychology, 62, 541-548.

Schneier, C. E., & Beusse, W. E. (1980). The impact of sex and time in grade on management rating in the public sector: Prospects for the Civil Service Reform Act. Proceedings of the Academy of Management, 40, 329-333.

Schwab, D. P., & Heneman, H. G. (1978). Age stereotyping in performance appraisal. Journal of Applied Psychology, 63, 573-578.

Smithers, J. W., & Reilly, R. R. (1987). True intercorrelation among job components, time delay in rating, and rater intelligence as determinants of accuracy in performance ratings. Organizational Behavior and Human Decision Processes, 40, 369-391.

Sulsky, L. M., & Balzer, W. K. (1988). Meaning and measurement of performance rating accuracy: Some methodological and theoretical concerns. Journal of Applied Psychology, 73, 497-506.

Svetlik, B., Prien, E., & Barrett, G. (1964). Relationships between job difficulty, employee's attitude toward his job, and supervisory ratings of the employee effectiveness. Journal of Applied Psychology, 48, 320-324.

Taft, R. (1971). The ability to judge people. In W. W. Ronan & E. P. Prien (Eds.), Perspectives on the measurement of human performance. New York: Appleton-Century-Crofts.

United States Department of Defense (1984). Armed Services Vocational Aptitude Battery (ASVAB): Test manual. (DoD Publication No. 1304.12AA), Washington, DC: United States Military Processing Command.

Vance, R. J., Coovert, M. D., MacCallum, R. C., & Hedge, J. W. (1989). Construct models of task performance. Journal of Applied Psychology, 74, 447-455.

Widaman, K. F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. Applied Psychological Measurement, 9, 1-26.

Wherry, R. J., & Bartlett, C. J. (1982). The control of bias in ratings: A theory of ratings. Personnel Psychology, 35, 521-551.

Wyer, R. S., & Srull, T. K. (1981). Category accessibility: Some theoretical and empirical issues concerning the processing of social stimulus information. In E. Higgins, C. Herman, & M. Zanna (Eds.), Social cognition: The Ontario symposium (Vol. 1). Hillsdale, NJ: Erlbaum.

Zedeck, S., & Baker, H. T. (1972). Nursing performance as measured by behavioral expectation scales: A multitrait-multirater analysis. Organizational Behavior and Human Performance, 7, 457-466.

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

APPENDIX A:

Examples of hands-on work sample tests for three Air Force specialties

Air Traffic Control Operator Test

Phase I Tower and Radar Hands-On Task 278

Objective: To evaluate the incumbent's ability to issue weather advisories.

Estimated Time: 2.5M Start: Finish: Time Req:

Time Limit: 3M #Times Performed: Last Performed:

Tools and Equipment: Current weather sequence and SIGMET Delta Three. Reference: FAAH 7110.65D, Chapter 2, Section 6, paragraph 2-101 and CDC Volume 1, Chapter 4, paragraph 4-2.

Background Information: The first exercise here is strictly an example of the issuing of weather information. By definition, a weather advisory is an announcement of a significant change in the current weather conditions, and subject matter experts' input indicated that this task is not often performed. However, the response to this task statement on the Occuaptional Survey showed a very high proportion of the respondents performing this task. Therefore, two exercises were developed which represent the strict definition of the weather advisory, and another which deals with current weather conditions.

Instructions: Administer in a quiet place.

SAY TO THE INCUMBENT

THESE EXERCISES WERE DEVELOPED IN REFERENCE TO CHAPTER 2, SECTION 6, PARAGRAPH 2-101 OF THE FAAH 7110.65D AND CHAPTER 4, PARAGRAPH 4-2 FROM THE CDC.

READ THE FOLLOWING WEATHER SEQUENCE AS YOU SHOULD TO ANY AIRCRAFT UNDER YOUR CONTROL. INDICATE TO ME WHEN YOU ARE READY TO BEGIN.

Performed or Answered Correctly	Yes	No
Did the incumbent read the weather as follows:		
1. Bergstrom record observation at 1158 Zulu (UTC)?	___	___
2. Measured ceiling nine hundred broken?	___	___

Phase I Tower and Radar — Hands-On Task 278

Performed or Answered Correctly	Yes	No
3. Visibility one-half?	___	___
4. Runway 21 R-V-V one quarter, Fog?	___	___
5. Temperature six zero?	___	___
6. Dewpoint six zero?	___	___
7. Wind 330 at 4?	___	___
8. Altimeter two niner niner seven?	___	___
9. Tower visibility north 1?	___	___

DURING YOUR SHIFT, SIGMET DELTA THREE HAS BEEN RECEIVED WHICH AFFECTS YOUR AIRSPACE. ISSUE THE SIGMET TO AIRCRAFT UNDER YOUR CONTROL.

Did the incumbent say:

10. "Attention all aircraft?"	___	___
11. "SIGMET Delta Three?"	___	___
12. "From Myton to Tuba City to Milford severe turbulence and severe clear icing below one zero thousand feet expected to continue beyond two three zero zero Coordinated Universal Time (UTC)"?	___	___

STOP TIME: _____

Avionic Communications Specialist Test

Phase I Hands-On Task 232

Objective: To evaluate the incumbent's ability to replace radio frequency coaxial connectors.

Estimated Time: 25M Start: Finish: Time Req:

Time Limit: 30M #Times Performed: Last Performed:

Tools and Equipment: T.O. 1-1A-14; RG 58 RF Cable, Connector (crimp type) Solder, Flux, Soldering Iron, Knife, Scissors, Cable Cutter, Eye Protection, Scribe, Crimping Tool.

Background: This general avionics maintenance task evaluates soldering and general techniques.

Configuration: This task could be used for actual production if a need for built-up cables exists. If not, two connectors will be required and can be recycled with a helper desoldering the salvage connector. All tools and materials should be placed on the bench with the length of cable and spare connector.

Instructions to Administrator: Administer at the soldering station. The task starts with the removal of the installed connector and includes the removal and replacement actions.

SAY TO THE INCUMBENT

THIS TASK IS THE REMOVAL AND REPLACEMENT OF THE COAXIAL CONNECTOR ON THE CABLE. YOU WILL BE EVALUATED AGAINST THE PROCEDURES IN T.O. 1-1A-14 AND GENERAL MAINTENANCE PROCEDURES.

Performed or Answered Correctly	Yes	No
Did the incumbent:		
1. Ensure that the cable was cut cleanly and squarely?	___	___
2. Place the ferrule on the cable?	___	___

Phase I Hands-On Task 232

	Performed or Answered Correctly	Yes	No
3.	Strip the outer jacket of the cable by carefully cutting around the circumference with a sharp knife then making a lengthwise cut and peeling off the jacket ensuring that the shield is not nicked or cut?	___	___
4.	Flare the braid?	___	___
5.	Ensure that the dielectric and shield braids were not damaged?	___	___
6.	Remove the dielectric by cutting around the circumference with a sharp knife, not quite through to the center conductor, and pulling the dielectric straight out?	___	___
7.	Ensure that eye protection was provided prior to soldering?	___	___
8.	Tin the center conductor with the soldering iron?	___	___
9.	Tin the inside of the contact pin with 60/40 solder?	___	___
10.	Solder the pin to the center conductor?	___	___
11.	Ensure that the pin is butted flush to the dielectric?	___	___
12.	Crimp the connector with the proper crimping tool?	___	___
13.	Ensure that center contact pin is properly positioned in the connector?	___	___

STOP TIME: _____

Information Systems Radio Operator Test

Phase I — Hands-On Task 258

Objective: To evaluate the incumbent's ability to transmit or receive messages using High Frequency (HF) equipment.

Estimated Time: 5M Start: Finish: Time Req:

Time Limit: 7M #Times Performed: Last Performed:

Tools and Equipment: ACP 125, messages in incumbent manual.

Background Information: N/A

Instructions:

Administer in a quiet area at the equipment. The incumbent may use ACP 125 as a reference. Hand the incumbent the messages to be transmitted.

SAY TO THE INCUMBENT

YOU HAVE JUST BEEN GIVEN TWO MESSAGES PREPARED IN HF FORMAT. WITHOUT ACTUALLY GOING OUT ON FREQUENCY, READ THESE MESSAGES TO ME AS YOU WOULD TRANSMIT THEM ON THE RADIO, STARTING WITH MESSAGE A. YOU WILL BE EVALUATED WITH RESPECT TO THE PROCEDURES OUTLINED IN ACP 125. YOU MAY USE ACP 125 AS A REFERENCE IF NEEDED.

Performed or Answered Correctly	Yes	No
Did the incumbent say:		
1. BUTLER 01 this is ALBROOK. I have one priority for your station, OVER?	___	___
EVALUATOR SAYS:		
ALBROOK this is BUTLER 01 GO AHEAD		
Did the incumbent say:		
2. This is ALBROOK number zero one?	___	___
3. Priority, time 242000 ZULU July, 85?	___	___

Phase I — Hands-On Task 258

Performed or Answered Correctly	Yes	No
Did the incumbent say:		
4. From HQ MAC?	___	___
5. To BUTLER 01?	___	___
6. Groups 06?	___	___
7. Break?	___	___
8. Request update status on engine trouble?	___	___
9. Break, OVER?	___	___
FOR MESSAGE B:		
10. AGA4TN this is AGA4KE I have one routine for your station, OVER?	___	___
EVALUATOR SAYS:		
AGA4KE this is AGA4TN GO AHEAD		
Did the incumbent say:		
11. This is AGA4KE number zero four?	___	___
12. Routine, time 241 858 ZULU July, 85?	___	___
13. From AGA4KE?	___	___
14. To AGA4TN?	___	___
15. Groups 10?	___	___
16. Break?	___	___
17. Test equipment you requested will be sent 25 July, 85?	___	___
18. Break, OVER?	___	___

STOP TIME: _____

APPENDIX B:

Task Rating Forms for Three Air Force Specialties

Task Rating Form for Air Traffic Control Operator

INSTRUCTIONS

The purpose of this rating form is to rate an airman's proficiency at performing a number of Air Traffic Control tasks. Proficiency refers to how skilled an airman is at performing various tasks on the job. Remember, we are concerned with level of ability to perform these tasks, excluding interpersonal factors (willingness to work, cooperating with others) or situational factors (weather conditions or equipment outages).

As you rate each task, ask yourself, "At what level of proficiency could the airman perform this particular task"? Place the number which corresponds to the appropriate rating in the blank preceding each item. Please provide a rating for each task on the following pages, even if the airman does not perform the task frequently.

The five levels that will be used on this rating form are listed below:

5 Always exceeds the acceptable level of proficiency
4 Frequently exceeds the acceptable level of proficiency
3 Meets the acceptable level of proficiency
2 Occasionally meets the acceptable level of proficiency
1 Never meets the acceptable level of proficiency

PLEASE RATE ALL TASKS

TASK RATINGS

---

5 Always exceeds the acceptable level of proficiency
4 Frequently exceeds the acceptable level of proficiency
3 Meets the acceptable level of proficiency
2 Occasionally meets the acceptable level of proficiency
1 Never meets the acceptable level of proficiency

---

1. _____ Annotate flight progress strip. (FAA Form 7230-8).

2. _____ Issue or transmit enroute clearances using FAA procedures.

3. _____ Perform interfacility communications.

4. _____ Issue weather advisories.

5. _____ Relay information from flight information publications (FLIP).

6. _____ Relay information from runway visual range (RVR) readings.

7. _____ Authorize special visual flight rules (SVFR) operations.

8. _____ Issue Detailed IFR holding instructions.

9. _____ Provide radar surveillance approaches.

10. _____ Perform radar handoffs.

11. _____ Control aircraft using light gun signals.

12. _____ Sequence departing aircraft.

13. _____ Sequence landing aircraft.

14. _____ Approve or disapprove clearances for aircraft or vehicle operation in ILS critical area.

15. _____ Issue taxiing instructions.

16. _____ Issue bird flight advisories.

Task Rating Form for Avionic Communications Specialists

INSTRUCTIONS

The purpose of this rating form is to rate an airman's proficiency at performing a number of Avionic Communications Maintenance tasks. Proficiency refers to how skilled an airman is at performing various tasks on the job. Remember, we are concerned with level of ability to perform these tasks, excluding interpersonal factors (willingness to work, cooperating with others) or situational factors (lack of tools, parts, or equipment).

As you rate each task, ask yourself, "At what level of proficiency could the airman perform this particular task"? Place the number which corresponds to the appropriate rating in the blank preceding each item. Please provide a rating for each task on the following pages, even if the airman does not perform the task frequently.

The five levels that will be used on this rating form are listed below:

5 Always exceeds the acceptable level of proficiency
4 Frequently exceeds the acceptable level of proficiency
3 Meets the acceptable level of proficiency
2 Occasionally meets the acceptable level of proficiency
1 Never meets the acceptable level of proficiency

PLEASE RATE ALL TASKS

TASK RATINGS

---

5 Always exceeds the acceptable level of proficiency
4 Frequently exceeds the acceptable level of proficiency
3 Meets the acceptable level of proficiency
2 Occasionally meets the acceptable level of proficiency
1 Never meets the acceptable level of proficiency

---

1. _____ Makes entries on Maintenance Data Collection Record forms (AFTO Form 349).

2. _____ Researches or identifies parts using illustrated parts breakdown (IPB).

3. _____ Removes or replaces radio frequency (RF) coaxial connectors.

4. _____ Removes or replaces UHF receiver-transmitter components.

5. _____ Sets-up UHF system peculiar test equipment.

6. _____ Sets-up flightline maintenance stands.

7. _____ Traces circuits or signals using wiring diagrams or schematics.

8. _____ Safety wires or bonds system components.

9. _____ Bench checks UHF receiver-transmitters.

10. _____ Aligns Very High Frequency (VHF) AM receiver-transmitters.

11. _____ Bench checks VHF AM receiver-transmitters.

12. _____ Sets-up HF system peculiar test equipment.

13. _____ Isolates malfunctions in HF coupler controls.

14. _____ Isolates malfunctions in high frequency (HF) receiver-transmitters.

15. _____ Sets-up HF system peculiar test equipment.

16. _____ Bench checks HF control units.

TASK RATINGS

---

5 Always exceeds the acceptable level of proficiency
4 Frequently exceeds the acceptable level of proficiency
3 Meets the acceptable level of proficiency
2 Occasionally meets the acceptable level of proficiency
1 Never meets the acceptable level of proficiency

---

17. _____ Removes or replaces HF receiver-transmitter subassemblies.

18. _____ Removes or replaces multiple wire plugs.

19. _____ Removes or replaces avionic system wiring or cables.

20. _____ Removes or replaces UHF receiver-transmitter subassemblies.

21. _____ Bench checks UHF control units.

Task Rating Form for Information Systems Radio Operator

INSTRUCTIONS

The purpose of this rating form is to rate an airman's proficiency at performing a number of Information Systems Radio Operator tasks. Proficiency refers to how skilled an airman is at performing various tasks on the job. Remember, we are concerned with level of ability to perform these tasks, excluding interpersonal factors (willingness to work, cooperating with others) or situational factors (lack of tools or parts, weather conditions).

As you rate each task, ask yourself, "At what level of proficiency could the airman perform this particular task"? Place the number which corresponds to the appropriate rating in the blank preceding each item. Please provide a rating for each task on the following pages, even if the airman does not perform the task frequently.

The five levels that will be used on this rating form are listed below:

5 Always exceeds the acceptable level of proficiency
4 Frequently exceeds the acceptable level of proficiency
3 Meets the acceptable level of proficiency
2 Occasionally meets the acceptable level of proficiency
1 Never meets the acceptable level of proficiency

PLEASE RATE ALL TASKS

TASK RATINGS

---

5 Always exceeds the acceptable level of proficiency
4 Frequently exceeds the acceptable level of proficiency
3 Meets the acceptable level of proficiency
2 Occasionally meets the acceptable level of proficiency
1 Never meets the acceptable level of proficiency

---

1. _____ Transmits or receives messages using HF equipment.

2. _____ Authenticates stations or message traffic using the PELE authentication system.

3. _____ Maintains position logs.

4. _____ Identifies incoming calls using call sign list.

5. _____ Maintains current call sign lists.

6. _____ Logs incoming or outgoing messages.

7. _____ Maintains phone patch records.

8. _____ Makes time checks (also known as time hacks).

9. _____ Configures scope control consoles for operation.

10. _____ Sets up duplex operations.

11. _____ Tunes or changes receiver frequencies by means of remote control.

12. _____ Tunes or changes transmitter frequencies by means of remote control.

13. _____ Operates rotating antenna equipment.

14. _____ Checks operation of radio recording equipment.

15. _____ Set up duplex operations on the scope control console.

16. _____ Checks operation of radio recording equipment.

TASK RATINGS

---

5 Always exceeds the acceptable level of proficiency
4 Frequently exceeds the acceptable level of proficiency
3 Meets the acceptable level of proficiency
2 Occasionally meets the acceptable level of proficiency
1 Never meets the acceptable level of proficiency

---

17. _____ Configures the Scope Signal III Console for operation.

18. _____ Operates the MEP-0 26A auxiliary generator.

19. _____ Relays communications traffic between fixed stations and mobile stations.

20. _____ Prepares messages using HF voice format.

21. _____ Receives international Morse Code (IMC).

APPENDIX C:

General Background Questionnaire Items

Measuring Perceived Performance Constraints

GENERAL BACKGROUND QUESTIONNAIRE

For the following questions, use the scale provided below to respond to each statement.

1 = Strongly Disagree
2 = Disagree
3 = Neither Agree or Disagree
4 = Agree
5 = Strongly Agree

1. ___ The technical manuals and other written materials that I use in my job are clear and undearstandable.

2. ___ The technical manuals and other written materials that I use in my job are available when I need them.

3. ___ I am able to use my skills and talents in my job.

APPENDIX D:

Rating Form Questionnaire Items Measuring Motivation to Rate Accurately, Trust in the Appraisal Process, and Rating Form Acceptability

RATING FORM QUESTIONNAIRE

In the following questions we are interested in your beliefs about the usefulness of the rating forms you just completed. Please respond to each statement using the scale provided below.

1 = Not at all
2 = To a small extent
3 = To a moderate extent
4 = To a great extent
5 = To a very great extent

1. ___ Did you care how accurate your ratings were?

2. ___ Did you feel it was important to make accurate ratings?

3. ___ Are you satisfied that you made the most accurate ratings you could?

4. ___ Based on your experience in this project, how important is it to you to make any performance ratings you do as accurate as you can?

5. ___ Do you feel other persons involved really tried to follow the rules in completing their ratings?

6. ___ Do you feel other persons involved really cared about making accurate ratings?

7. ___ Do you feel other persons were comfortable giving low ratings to themselves or others?

8. ___ Do the rating forms evaluate job proficiency fairly?

9. ___ Are the rating forms easy to use and understandable as a means of determining job proficiency?

10. ___ Would you be able to tell the difference between good and poor performers by looking at the ratings they were given?

11. ___ If someone were to look at the ratings on the form, would they be able to get a true picture of the performance level of the person being rated?

12. ___ Overall, are the rating forms acceptable to you as a way to determine job proficiency?