UTILITY OF PEER ASSESSMENT METHODS FOR POLICE OFFICER EVALUATION: VALIDITY, RELIABILITY, AND USER REACTION

KEVIN G. LOVE

CENTER FOR EVALUATION RESEARCH
UNIVERSITY OF SOUTH FLORIDA

UTILITY OF PEER ASSESSMENT METHODS FOR POLICE OFFICER EVALUATION: VALIDITY, RELIABILITY, AND USER REACTION

by

Kevin G. Love, Ph.D.

Center for Evaluation Research

University of South Florida

January, 1980

PREFACE

The material in this project was prepared under Grant No. 79-NI-AX-0023 from the Law Enforcement Assistance Administration, U. S. Department of Justice. Researchers undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment. Therefore, points of view or opinions stated in this document do not necessarily represent the official position or policy of the U. S. Department of Justice.

ACKNOWLEDGEMENTS

The author would like to thank Herb Meyer and Ed Levine for their friendship, support, and direction of this research project. Thanks must also go to Frank Sistrunk for his true friendship and personal guidance during the past four years of graduate study. Finally, a special thanks to my wife, Denise, for her unending support, encouragement, and love throughout these years of intense study.

TABLE OF CONTENTS

LIST OF TABLES

EXECUTIVE SUMMARY

The investigation was conducted to test for differences between three ways of using police officers to evaluate the performance of fellow officers, namely peer nomination, peer ranking, and peer rating. For each method of assessment three variables were examined. These were validity (the relation of the peer assessments to police sergeant evaluations and certain objective performance measures), reliability (consistency among fellow officers when evaluating the same officer), and user reaction (given by the officers on each of the peer assessment techniques).

The relationship between assessments made by fellow officers and (a) rankings and ratings made by the sergeants and (b) five objective performance indices was described by simple and multiple correlations. These correlations were tested for significant differences across the three methods. Agreement among fellow officers when evaluating the same officer was estimated and then tested for significant differences among the three peer assessment methods. The notion that the officers had reacted differently towards any one method of peer assessment was tested for by means of multivariate analysis of variance.

Based on previous research in the area, several results were expected. First, it was believed that peer nominations

would relate more closely with the rankings and ratings provided by sergeants and the objective measures. This belief was based on the method used to collect peer nominations. That is, only the high performing police officers were identified. Second, it was felt that peer nomination would also display the greatest consistency among fellow officers evaluating the same officer. Again this was based on this same "method bias." Third, since there has been no previous research which has examined officers' reactions to various types of peer assessment methods, the present study represented an important initial inquiry into this area. Fourth, the friendship among the police officers who served as both peer assessors and assessees was measured. The impact of this friendship upon each method's relationship with the sergeant evaluations was expected to be equal across the three peer assessment methods.

The majority of past research in this area has concerned itself with investigating only one method of peer assessment at a time, usually peer nomination. Moreover, only reliability and validity has been examined. There has not been, however, a systematic comparison of the three techniques to measure the relative strengths and weaknesses of individual methods and the user reaction, especially when used for police officer performance evaluation.

One hundred and twenty-six police officers and 19 police corporals provided peer assessments on all police officer and corporal personnel working in a patrol capacity within a municipal law enforcement agency serving a medium sized southeastern city. In all, 256 police officers and 35 police corporals were

assessed by fellow officers.

A peer assessment instrument was provided to each officer. The assessment booklet requested nominations of fellow officers and rankings and ratings of squad members on each of nine performance dimensions. Whereas rankings and ratings were to be made on only the assessor's fellow squad members, any officer working patrol could be nominated. For each person assessed by any of the three methods, five questions regarding the peer assessor's friendship or social relationship with the officer assessed were answered. The evaluating officer also provided indications as to his or her reactions to using each method by responding to a series of seven statements.

All three peer assessment methods displayed close relationships with rankings and ratings provided by the squad sergeant. However, peer ranking revealed a significantly greater correlation with the sergeant rankings and ratings than either peer nomination or peer rating. High correlations were also obtained between peer assessments and the number of commendations and awards received by the assessee and the number of one-the-job injuries sustained. Moreover, all methods showed great consistency (reliability estimates) among fellow officers evaluating the same officer, with peer ranking showing significantly more consistency than either peer nominations or peer ratings. Reactions of those completing the peer assessments were mildly negative towards all three peer assessment techniques, with no one method being either liked or disliked to a greater extent than the others. The measures of friendship did not have a substantial impact upon the relationship between each peer

assessment method and the rankings and rating provided by the sergeants.

The results suggested that peer assessment is a valid and reliable means of evaluating police officer performance. In particular, peer ranking was seen as having the greatest potential for accurately and consistently discriminating among patrol officers. Questions were raised, however, regarding the proper description and measure of the friendship which exists among fellow officers. It was suggested, in light of past research in the area, that the description and measurement of friendship between the peer assessor and assessee may be the determining factor as to whether the social relationship among officers reveals a significant influence on peer assessments.

CHAPTER I

INTRODUCTION

In a comprehensive review of relevant research, Kane and Lawler (1978) have commented upon the potential usefulness of three peer-assessment techniques for evaluating employee performance. As they pointed out, in spite of much previous research which has documented high degrees of validity and reliability for peer assessment, there has been a reluctance to adopt peer assessment for evaluative purposes within organizations. This reluctance has stemmed primarily from a complacency with more traditional performance evaluation techniques and to a lesser extent from an overall failure to recognize the utility of peer assessments.

Recently, various authors (Borman, 1974; Klimoski & London, 1974; Lawler, 1967) have attempted to justify the use of fellow workers in the evaluative process. Their justification has centered upon the notion that peers can contribute evaluative performance information which is psychometrically superior to and significantly different from that which is obtained through more common performance appraisal techniques. Korman (1968), in reference to peer ratings, concluded that "peer ratings are better predictors of performance than other

psychometric procedures," and "better than most objective tests." Similar conclusions have been advanced by other authors (Lindzey & Byrne, 1969; Miner, 1968).

In an attempt to allay the cautiousness which has guided the implementation of peer assessment in comprehensive performance evaluation systems, Kane and Lawler (1978) outlined several organizational settings which contain factors facilitative to the use of peer assessment. Three optimal organizational settings were described:

(1) where members of peer groups have unique views of important dimensions of each other's performance;

(2) where members of peer groups are actually capable of accurately observing and interpreting the crucial aspects of each other's performance; and

(3) where there is a need to improve the effectiveness of assessments gathered on certain characteristics of peer group members.

The first two settings would seem common among those types of organizations which divide their production capacity into departmental or work-group units. The third factor would be a reaction to mounting pressure to improve the job-relatedness and accuracy of present evaluative methods. Thus, if the above conditions are in evidence, Kane and Lawler (1978) suggest that the use of peer assessment could prove quite beneficial to an organization.

Methods of Peer Assessment

Three basic techniques for measuring the judgments of peers within the evaluative process have come to represent

the field of peer assessment. Unfortunately, many past studies have confused these peer assessment techniques by referring to peer nomination, peer ranking, and peer rating under the single category of "peer or buddy ratings." This inaccuracy has hindered needed research into the respective strengths and weakness of the varied methods of peer assessment through the global prescriptions for the use of "peer ratings." Kane and Lawler (1978) have provided an explanation of each technique. Their definitions will serve as the foundation for the following overview.

Peer Nomination

Each member of a specified group designates a certain number of the group members as being highest on a particular dimension of performance or trait. Occasionally each member is also required to indicate a specified number of group members who are lowest on the same performance dimensions. Common practice has also had the nominees listed in decreasing order of extremeness on each dimension or trait. Those nominated are ranked from highest to lowest, with the group members being required to exclude themselves from those considered when nominations are made.

Peer Ranking

This method of peer assessment requires that each group member rank all other group members from best to worst on each of a specified number of performance dimensions.

Kane and Lawler (1978) have stated that ranking is the most discriminating of all peer assessment methods based on the possibility that each group member could be differentiated from all others by the average rank received (if no ties were allowed). Peer ranking is recommended when the purpose of the assessment is solely to discriminate between all members of a group.

Peer Rating

Under this technique each group member rates every other group member on a specified and defined number of performance dimensions or personal characteristics. In so doing several kinds of rating scales may be used. At present the type of rating scale usually regarded as providing the most detailed performance information is the "behaviorally anchored rating scale" (BARS; Smith & Kendall, 1963). In such scales each interval of the scale is linked to a behavioral incident which exemplifies the specified level of the performance dimension being rated. These incidents serve to connect each level of performance to the realm of on-the-job behavior. Springer (1953) and Borman (1974) have incorporated BARS in their peer rating research. Other authors have used different, perhaps more traditional, types of rating scales, such as the adjective or numerically anchored graphic rating scale (Borg & Hamilton, 1956; Brehm & Festinger, 1957; Cox & Krenboltz, 1958; Fiske & Cox, 1960; Ricciuti, 1955;

Swanson & Johnson, 1975; Trites, 1960; Winch & Anderson, 1967; Wodder & Hall, 1962), forced-distribution method (Kruat, 1975; Roadman, 1964; Schmidt & Johnson, 1973), and forced-choice method (Bartlett, 1959).

Differences in Discriminability Among Methods

The three methods of peer assessment represent three different approaches towards discriminating among organizational members. Whereas the nomination methodology seeks to identify only the best (and occasionally the worst) performers within the subject group, by design it inherently designates the remaining group members, those not nominated, as equal in ability. In so doing, peer nomination may be regarded as basically an incomplete derivative of ranking or rating.

Thus, peer nomination may be seen as the least powerful method of peer assessment when it comes to discriminating between all members of the subject group. Peer rating would rank second in discriminability due to the likelihood of equal ratings assigned to two or more group members. Peer rankings then would be the method of peer assessment which allows the finest discrimination among all group members. This assumes, however, that no ties in assigned ranks would be allowed with all members of the group being ranked. Since these conditions may not be appropriate in all

applications of peer ranking its discriminability may be actually closer to that of peer rating.

Perhaps the key consideration in determining the most effective type of peer assessment with regard to power of measurement is feedback of the evaluations to those assessed. In this regard peer rating allows for the most detailed and discriminating provision of feedback, when compared to peer nominations and peer rankings.

The Validity of Peer Assessments

Validity Coefficients

The validity of peer assessment techniques, as with other evaluation methods, is indicated by the magnitude of the relationship between a peer's evaluation of some trait, behavior, or performance outcome and a defined and consistently measured criterion, usually supervisor judgments. Of course the ideal case of calculating validity is one where a more reliable and objective measure than supervisor judgments of the same trait, behavior, or performance outcome is linked to the peer assessment.

Peer nomination. In their extensive analysis of the literature, Kane and Lawler (1978) summarized the reported validity of peer nominations obtained in a variety of organizational settings. Past research has indicated that peer nominations can be a valid and psychometrically sound method of evaluating current performance and predicting future job

proficiency (see Amir, Kovarsky & Sharon, 1970; Booker & Miller, 1966; Downey, Medland & Yates, 1976; Gunderson & Nelson, 1966; Hollander, 1954b, 1965; Kaufman & Johnson, 1974; Kubany, 1957; Mayfield, 1972; Mayo, 1956; Weitz, 1958; Wherry & Fryer, 1949; Williams & Leavitt, 1947).

Moreover, these authors have revealed certain factors which affect the validity of the nominations and defined them as follows:

(a) The setting--Overall, the validities of the peer nomination process have been shown to be higher in military settings than in civilian settings (Kane & Lawler, 1978). The usual explanation for this finding has been the increased intensity and duration of contact among groups of military personnel as compared to civilians. Moreover, within the military setting the highest validities have been found for objectively measured criteria (i.e., retention, graduation, promotion, etc.) (Downey et al., 1956; Wherry & Fryer, 1949) as opposed to superiors' judgments (Hollander, 1956b, 1965; Kaufman & Johnson, 1974);

(b) The performance dimensions being measured--Consistently it has been shown that peer nominations obtained using characteristics most directly related to the validity criteria have higher validity than when more abstract factors have

been utilized, such as personality traits (Hollander, 1954b, 1956b, 1965; Mayfield, 1972; Smith, 1967; Waters & Waters, 1970);

(c) The predictiveness of nominations to other groups--Both Amir et al. (1970) and Mayfield (1970) indicated that the predictiveness of peer nominations will deteriorate when used within a group other than that which generated the nominations. That is, in a group setting distinct from the original nomination group, the predictiveness of peer nomination scores decreases as a linear function of the dissimilarity between the original and new group; and

(d) The early development of substantial validity--Several studies (Amir et al., 1970; Hollander, 1956b; Wherry & Fryer, 1949) have revealed that the validity of peer nominations for predicting leadership performance develops very early in the life of a group. According to this research, predictiveness reaches a plateau after no more than three weeks. This finding, however, was only apparent when fairly intensive groups were used.

Even though the majority of the peer nomination research has indicated sufficient validity for the peer nomination process, there remain several points of contention.

Much past research has centered upon leadership qualities as predictors of leadership performance. The inherent relationship between the performance predictors and criteria (i.e., a "bootstrapping" effect upon the validity coefficient) in such studies has not been adequately accounted. Those who have used nonleadership-related criteria (especially of a nonrating variety) have found only marginal validities (Kubany, 1957; Mayfield, 1972; Smith, 1967; Waters & Waters, 1970). Thus, it is possible that the validity of the peer nomination method as demonstrated through past research could be largely a function of leadership performance. Future research must employ the peer nomination process with both nonleadership-related subjective evaluative criteria and objective criteria when validity is to be examined.

In addition, Kane and Lawler (1978) described the inherent bias within the methodology of the peer nomination process. This bias has increased the validity coefficients reported over those in evidence with other peer assessment techniques. In contrast with other methods of peer assessment, peer nominations deal only with the extreme high and low performers of a specified group of persons. Thus, the process combines all those remaining group members not nominated into an undifferentiated middle-range category. Validity obtained using only these extreme performers of

a group has considered only a small portion of the group members, taken out of the context of the group as a whole.

Peer ranking. Unfortunately, research dealing with the validity of peer ranking is noticeably absent. Utilizing the few available investigations in this area, only a slight trend regarding validity becomes apparent.

Previous research has indicated that peer rankings display a tendency to correlate more highly with rankings provided by the ratee (self-rankings) and the supervisor than with ratings provided by these same sources (Borg & Hamilton, 1956; Lawler, 1967; Tucker et al., 1967). Considering this finding, Kane and Lawler (1978) suggested that ranking either contributes unique method variance to the overall evaluation process or it may simply be a more valid procedure than ratings.

Another possible explanation not addressed by Kane and Lawler (1978) is based on the greater reliabilities obtained with the ranking method in general. Campbell, Dunnette, Lawler, and Weick (1970) have declared the superiority of rankings in reliably discriminating among assessees. Yet, to date, there has been no systematic study of the validity of peer rankings using sound performance criteria. Thus, the reported findings which indicate a lack of convergence between rankings and ratings cannot be attributed totally to the ranking procedure itself, the scoring of the measure, or even the dimension of performance being evaluated.

Peer rating. Overall, past research seems to indicate that peer ratings are less valid predictors of independent criteria than peer nominations (Borman, 1974; Freeberg, 1969; Gordon & Medland, 1965; Hoffman & Rohrer, 1954, Ricciuti, 1955; Springer, 1953; Swanson & Johnson, 1975; Trites, 1960; Tucker, Cline & Schmitt, 1967; Wiggins, Blackburn & Hackman, 1967). It is interesting to note, however, that the validity coefficients obtained for peer ratings unlike peer nominations in both military and civilian settings have been similar in magnitude.

It has been suggested (Kane & Lawler, 1978) that the lower validity of peer rating is due to the fact that peer ratings utilize the entire proficiency continuum of each performance dimension under consideration rather than only the extremes of the distribution. This may account for the substantial decrease in validity when peer ratings are compared to peer nominations.

The demonstrated strength of peer ratings, on the other hand, has been in its superiority in the specification of feedback for group members. Peer ratings make it possible for group members to better discriminate among the performance of group members. Thus, peer ratings are more sensitive in documenting information about each group member.

Friendship Bias

A substantial amount of past research has been geared at the evaluation of the systematic biasing effects of

friendship or social attraction upon the validity of peer assessment methods, most frequently peer nominations. Nevertheless, little progress has been made in determining the effect of friendship upon the validity of peer rankings and peer ratings and investigating the very composition of the social attraction phenomenon which exists among fellow organization members.

Peer nomination. The major area of concern regarding bias in the nomination process has been the influence of so-called "friendship" between members of the nomination group. Previous research using friendship nominations to measure the impact of social attraction on validity has indicated that even though the friendship and leadership nominations were partially correlated, leadership scores were not a direct function of friendship or another variate of social attraction, "popularity" (Hollander, 1956a; Hollander & Webb, 1955). Unfortunately, the issue of whether leadership scores based on people who nominated their friends as a leader were as valid as those which were based on people who did not do so was not addressed by this research.

In another study Waters and Waters (1970) found that friendship exerted a strong influence on the validity of peer nominations. Their research suggested that actual "antipathy" toward a person affected the validity of peer nomination scores to a lesser extent than did friendliness. Conversely, Doll and Longo (1962) demonstrated that antipathy adversely

affected validity where the nomination system served as an outlet for the expression of negative feelings through negative nominations. They indicated that a large portion of negative nominations was made for reasons unrelated to the performance attributes upon which the nominations were supposedly to be obtained.

In opposition to this finding, Kane and Lawler (1978) have pointed out that Waters and Waters (1970) did not compare the validity of nomination scores based on the number of positive nominations received with those based on the number of negative nominations. In examining their data, Kane and Lawler found that the means of the three groups of assessors (high friendship, average friendship, and low friendship) indicated that negative nominations were a greater factor in the scores generated by the low-friendship group members than those in the other friendship categories. Yet, the scores of the low-friendship assessee group were equally or more valid than those from either of the other assessee groups. This finding has been interpreted in light of the specific criterion against which the scores were validated. In this particular study successful performance of the criterion activity required the ability to avoid the dislike of the other group members. Thus, the criterion scores were correlated with a measure of likeability (i.e., the negative nominations).

Nonetheless, other research has suggested that "likeability" may not interfere with the validity of the nomination scores for other performance criteria. Theodorson (1957) reported that in more cohesive groups, when the bases for person attraction were more closely related to the members' individual contribution to group achievement, negative nominations and friendship bias did not detract from the validity of performance nomination scores. Theodorson stated that negative nominations should be considered with regard to the relevance of likeability to both the criterion measure and group cohesiveness.

In spite of the abundance of past research on this topic, little has been done to investigate the nature of the friendship factor. Much past research has proposed theories to explain why certain conditions produce attraction between persons. Researchers utilizing peer nominations have avoided integrating these theories into their investigations by using only the expedient method of asking people who they like and (in some cases) how much each is liked.

Most of the empirical evidence concerning the development, maintenance, and improvement of attraction between people has been gathered in laboratory settings. Even though its generalizability may be questioned, these findings do bear consideration in the determination of the biasing effects of friendship.

This past research was studied in relation to the concept of social attraction or friendship among fellow workers. In this

regard previous investigations have delineated a number of antecedent conditions which have been found to be linked with liking between persons. These conditions were the basis for the creation of five variables used to measure friendship among fellow workers.

Someone who provides is liked, especially when the rewards are greater then expected (Berkowitz & Levy, 1956; Kleiner, 1960; Myers, 1962). Other conditions for liking are similarity (Byrne, 1969; Byrne & Nelson, 1965; Newcomb, 1961), proximity (Gullahorn, 1952; Kipnis, 1957; Zander & Havelin, 1960), self-esteeem (Aronson & Linder, 1965; Deutsch & Solomon, 1959; Jones, Knurek & Regan, 1973), and physical attractiveness (Berscheid, Dion, Walster & Walster, 1971; Cavior & Dokecki, 1971; Dion, Berscheid & Walster, 1972; Murstein, 1972; Sigall & Landy, 1973).

It seems an individual is more prone to like someone who has similar attitudes, beliefs, and values, who lives or works nearby, and who provides positive evaluations that inflate one's self-esteem. Physical attraction is also an important factor in the development of attraction, especially between persons of the opposite sex.

Based on this research five variables (contact on the job, contact off the job, knowledge of person, liking of person, and friendship with person) were constructed to investigate the biasing effects of friendship upon the validity of peer nominations along with the other methods of peer assessment.

Peer ranking. Turk (1961) produced evidence that peers' evaluations of a person's task proficiency were based

on their perception of a person's attractiveness when they felt it was important to do well at the group task. For these people, the rank they assigned a person regarding task performance influenced the rank assigned on likeability or attractiveness, rather than the reverse. It has been suggested that Turk's finding may be a replication of Theodorson's results obtained using peer nominations (Kane & Lawler, 1978). This could be accepted if it could be shown that the level of task identification in a group is isomorphic to the group's cohesiveness. If task identification is not similar to group cohesiveness, these two factors nonetheless interact and limit the conditions under which personal attraction influences the effectiveness of peer assessment.

Yet, within this area of research, however, there has been little comparative investigation of the impact of social attraction (friendship, likeability, etc.) on the validity of peer rankings. A study of this kind would prove quite significant given the present status of research.

Peer rating. Apparently, no previous research has reported the effects of friendship or any other aspect of social attraction on the validity of peer ratings. There seems to be little reason, however, to believe that the impact of such a phenomenon would be any less pronounced with peer ratings than with peer nominations or peer rankings.

Any systematic attempt to evaluate the impact of friendship on the validity of peer ratings would serve to partially fill the void of research in this area.

Reliability of Peer Assessments

Past research has utilized two measures of reliability in reference to peer assessment. These are internal consistency, referring to the amount of agreement among those making the assessments, and test-retest reliability, obtained by comparing measures obtained upon the same people using the same characteristics and procedure at two different points in time. For purposes of the present research, internal consistency reliability was determined.

Peer Nomination

Previous investigations have indicated high levels of internal consistency for peer nominations (Gunderson & Nelson, 1966; Hollander, 1956b; Kubany, 1957; Mayo, 1956; Smith, 1967; Suci, Vallance & Glickman, 1955; Waters & Waters, 1970; Williams & Leavitt, 1947). Similar levels of internal consistency have been found for both civilian and military settings. Nevertheless, as with validity, it has been suggested (Kane & Lawler, 1978) that the high reliability estimates prevalent with the use of peer nominations may be due to the method's identification of only the extreme performers of a group.

Peer Ranking

There has been little research dealing with the reliability of peer rankings. What has been conducted has indicated that peer ranking may exhibit acceptable levels of reliability (Borg & Hamilton, 1956; Hollander, 1954a). This would seem to be consistent with the reliability estimates obtained when ranking has been used by supervisors for performance evaluation purposes (Campbell et al., 1970). A systematic investigation into the reliability of the peer ranking process, as compared to other methods of peer assessment, would provide the much needed empirical foundation for future research.

Peer Rating

Usual internal consistency reliability estimates for peer ratings have been disappointingly low when compared to those obtained for the peer nomination method (Bartlett, 1959; Borman, 1974; Freeberg, 1969; Springer, 1953; Stahl & Steger, 1977). In explanation of this finding, Kane and Lawler (1978) suggested that the internal consistency of peer ratings is affected by the possibilities of disagreements among peers when they must rate everyone in their group. This disagreement is not evident in the peer nomination process where assessors have only to decide upon the most and least proficient group members.

Yet, peer rating has displayed certain advantages which distinguish it from the peer nomination method, in

spite of its low reliability. Peer ratings have been found to be more useful in providing feedback than peer nominations.

Investigations are needed to determine how the reliability of peer ratings may be improved so that the method may be empirically evaluated for possible organizational use.

User Reaction to Peer Assessment

User reaction has generally referred to the degree to which members of a group react positively or negatively to the experience of evaluating fellow group members with peer assessments.

Peer Nomination

Only three previous studies have mentioned the reaction of their respondents. Downey et al. (1976) and Webb (1955) have characterized the reactions as negative. Mayfield (1970) reported no resistance to obtaining the requested nominations, but no enthusiasm was apparent. Only an interpretation of the organizational settings in which these reactions were obtained seems to shed any light on the issue. In the first two studies nominations were collected within the military, whereas Mayfield utilized civilian insurance salespeople.

It must be emphasized that these reports were basically anecdotal and did not constitute a systematic analysis of the user reactions to the peer nomination method. Such research has yet to be reported.

Peer Ranking

To date there has been no published indication of users' reactions to the peer ranking method.

Peer Rating

Only Roadman (1964) has provided any insight into the area of user reaction for peer ratings. In an anecdotal fashion, he reported that managers who provided peer ratings felt that it was a "constructive" and "nonthreatening" experience.

Present Investigation and Hypotheses

Considerations of Validity of Peer Assessment

Based on the reported research dealing with the validity of peer nominations, peer rankings, and peer ratings, the present study furnished a systematic comparative investigation of the relationship of each method to selected judgmental and objective performance measures. The following factors were considered:

(1) The correlation of each peer assessment technique with performance criteria (i.e., supervisor rankings and ratings and objective indices)--Due to the inherent method bias of peer nominations in inflating obtained validity coefficients in past research, it was hypothesized that peer nominations would reveal higher correlations with the

criteria than either peer rankings or peer ratings. However, based on past peer nomination research dealing with leadership performance and leadership-related criteria (i.e., the bootstrapping effect), it was expected that the correlation of peer nominations with nonleadership objective performance measures and supervisor judgments would be somewhat lower;

(2) The impact of "friendship" upon the relationship of each method of peer assessment with supervisor rankings and ratings--It seems quite superficial to express a complex sociometric phenomenon, such as social attraction among fellow workers, as a singular concept. Therefore, the present study incorporated several components of the perceived social attraction among organization members in analyzing the impact of so-called "friendship" on the correlation of each method of peer assessment with supervisor rankings and ratings. It was expected that friendship would have similar effects on all methods of peer assessment--peer nomination, peer ranking, and peer rating. Due to the method of obtaining peer nominations, it was hypothesized that the measure of "friendship" would reveal a greater impact on the validity of peer nominations.

Because little past research had been done in the area, the present study would contribute greatly to this body of knowledge.

Reliability of Peer Assessment Methods

The present study went beyond the suggestions of Kane and Lawler (1978) in calculating improved estimates of inter-rater reliability for each method of peer assessment. These reliability estimates were subsequently compared across the three peer assessment techniques. Based on past research which had documented the impact of the method bias of peer nominations on both validity and reliability, it was expected that peer nomination would prove more reliable than either peer ranking or peer rating.

User Reactions to the Different Techniques of Peer Assessment

With a paucity of previous research for an empirical base, the present study systematically delineated and described the reactions of the individuals who had used all three peer assessment methods. The user reactions were compared among methods to find out which peer assessment technique generated the most and least resistance. With only a few studies remotely addressing this issue, the present research was an important systematic investigation of user reactions to processes of peer evaluation.

CHAPTER II

METHOD

Subjects

One hundred and twenty-six police officers and 19 police corporals served as peer assessors. Seven hundred and sixty-seven peer rankings and ratings were collected on 256 police officers and 35 police corporals (peer assessees). Peer rankings and ratings were made by the peer assessors on all fellow squad members over the 35 patrol squads. One hundred and ninety peer nominations were obtained on the same sample of assessees. Nominations could be made on any police officer or corporal performing a patrol function, regardless of squad.

Thirty-three police sergeants (squad supervisors) provided rankings and ratings of all their respective squad members. Overall, 263 supervisor rankings and ratings were obtained.

All police personnel involved in the study were members of a municipal police department serving a medium-sized city (population approximately 273,000) within the southeastern United States. At the time of the study, all police personnel asked to participate were performing in a patrol capacity.

Additional demographic data on the participants could not be reported in order to maintain the utmost anonymity.

Procedure and Experimental Variables

Preliminary Procedure

Peer assessment. A peer assessment instrument (see Appendix 1) was developed which consisted of three sequential evaluation phases:

(1) nomination, ranking, and rating of eligible fellow officers or corporals on nine performance dimensions;

(2) evaluation of the sociometric relationship or "friendship" between the peer assessor and each assessee; and

(3) reaction of the peer assessors to the use of each method of peer assessment.

The peer assessment instrument was distributed to all police officer and corporal personnel performing in a patrol capacity. Assessment instruments were provided to each squad supervisor (sergeant) for distribution to their respective squad members. A total of 256 police officers and 35 police corporals was given the opportunity to provide evaluative performance data regarding their fellow squad members. One hundred and twenty-six police officers and 19 police corporals completed the peer assessment instrument, a return rate of 50%. These peer assessors provided 190 nominations, 767 rankings, and 767 ratings. The greater number of rankings and ratings reflected the fact that the peer assessors were instructed to rank and rate all of their respective squad

members (6 to 9 individuals) on each of the nine performance dimensions. On the other hand, only three nominees were requested for each performance dimension.

The principal researcher, who was representing the Test Validation Division of the local Civil Service Board, attended roll calls for all shifts for two consecutive days. The project was described in detail, being an integral part of a criterion-related validation study being conducted to test the job relatedness of a newly implemented selection system for police officers. The rationale for the collection of peer assessments was the provision of accurate performance criteria needed for the concurrent validation strategy being used.

Supervisor assessment. A supervisory assessment instrument (see Appendix 2) was developed, which was a duplicate of the peer assessment form but without the nomination, evaluation of friendship, and reaction sections. That is, the squad supervisors were asked to provide only rankings and ratings on their squad personnel. Thirty-six squad supervisors were asked to evaluate their respective squad members.

Compilation of the assessment data. In order to protect the anonymity of the persons being evaluated by both peers and supervisors, each officer and each corporal was assigned a randomly generated identification number for use

as a reference index. Moreover, all assessment instruments were returned directly to the principle researcher, bypassing the usual police agency flow of paperwork.

Participation in the study was voluntary. Every effort was made to assure the participating persons that all performance information obtained would be strictly confidential and used for test validation purposes only.

A two week period was allowed for return of the assessment instruments. After that period of time, failure to return the instrument was interpreted as a decision not to participate in the study. There appeared to be several reasons for not responding: (a) since the peer assessments were requested during the final stage of labor negotiations between city officials and the police union, the assessment instrument was seen as a way for city officials to gather performance information; (b) there was a hesitancy by the police personnel to participate in any research effort due to bad past experiences with studies where controversial findings were never acted upon; and (c) certain officers and corporals were on sick leave, absent, or on vacation when the assessment instruments were distributed.

Objective Performance Indices

Due to the expense of accessing individual personnel files within the police agency, five objective indices of police performance were obtained on a random sample of 104

police officers and 10 police corporals rather than on all peer assessees. These objective performance indices had been shown to be significant predictors of police officer performance in other research studies (Cascio & Valenzi, 1978). They were:

(1) age (in years);

(2) length of service with police department (in months);

(3) number of commendations and awards;

(4) number of on-the-job injuries; and

(5) number of sick days (to the nearest half day).

Additional objective performance indices, which had also been shown in past research to be predictive of police officer performance, could not be obtained due to pending litigation calling into question the legality of releasing sensitive performance-related information to persons other than the top staff of the police agency.

Subjective Nominations, Rankings, and Ratings of Performance

Nominations, rankings, and ratings of squad personnel were gathered from the peer assessors, with supervisors providing only rankings and ratings. All evaluations utilized nine distinct performance dimensions. These dimensions were based upon a task-based job analysis which had been recently completed for the police officer position within the participating police department (Love, 1978). The definition and

development of the performance dimensions was similar to the procedure outlined by Landy, Farr, Saal, and Freytag (1976). The title and definition of each performance dimension was as follows:

(1) job knowledge--use of knowledge of laws, procedures, policies, and techniques related to the law enforcement function (patrol, arrest, testifying, etc.), including the application of prior training.

(2) decision-making--analytic assessment of the situation and taking necessary and appropriate action after consideration of alternative approaches.

(3) dealing with co-workers--ability to work with fellow employees, both sworn and nonuniformed personnel, including accepting and giving constructive criticism, mutual decisionmaking, and taking an equal share of the workload.

(4) use of equipment--skill in the use of firearms, other weapons, and other specialized equipment.

(5) dealing with the public--knowledge and skill in using techniques geared at dealing with the public in a respectful, tactful style.

(6) communication--ability to make oneself understood and understand others in face-to-face situations and to transmit and receive information in both oral and written form.

(7) reliability--dependability in job attendance, effort expenditure, acceptance of responsibility, functioning appropriately under stress, and accuracy in all details of work.

(8) demeanor--personal and professional pride as shown by his or her standards of behavior and physical appearance.

(9) work attitude--interested in serving the public through the fair and objective enforcement and administration of the law, gaining satisfaction from doing his or her job well.

For clarification purposes, in addition to the title and definition of the dimensions, the knowledges, skills, abilities, and personal characteristics which formed the basis of the dimensions were presented to the assessor.

Peer nominations. Nominations of fellow officers or corporals were made for each performance dimension by the peer assessors. Eligible nominees were any police officer or corporal performing patrol functions within either uniform patrol district. A peer assessor was not allowed to nominate him/herself. For each performance dimension the peer assessor was asked to nominate three eligible individuals who, in his or her opinion, performed best on that dimension. Those chosen were then ordered from first to third.

The peer nominations were scored using a continuous scale. A value of three was given for a first-place nomination,

two for a second-place nomination, and one for a third-place nomination. Nomination scores for each individual nominee were summed across assessors, yielding a total nomination value.

Peer and supervisor ranking. Both peer and supervisor assessors were instructed to consider the performance of all squad members, excluding themselves, which would be characterized by the first performance dimension. The assessors then ranked each squad member from one (1) to n (depending on the size of the squad) as to the proficiency of the squad members within the context of the respective dimension. This procedure was repeated for all nine performance dimensions.

Peer and supervisor rating. Ratings were gathered from both peers and squad supervisors for all squad members, excluding the rater, on each performance dimension. A 9-point behaviorally anchored scale (BARS) was utilized for the ratings. Past research using similar BARS revealed interrater reliabilities across performance dimensions in the mid .80s (Cascio & Valenzi, 1978). The development of the BARS for each of the nine performance dimensions followed a procedure which has been described by Landy et al. (1976). Whereas Landy et al. (1976) used 15 police jurisdictions in the generation and construction of behavioral anchors, the present study utilized only a single police agency. Thus,

the BARS were applied in the same organizational setting from which they had been developed.

The rating process required the assessor to consider each squad member's performance on the first performance dimension. A rating for that person from one (1) to nine (9) was given using the BARS. This rating process was repeated for each performance dimension.

Friendship Ratings

After the peer assessors had nominated, ranked, and rated the eligible police personnel on all nine performance dimensions, ratings on five "friendship" variables were obtained. For each person either nominated, ranked, or rated by the peer assessor, responses for the following questions were provided. The available alternative responses followed a 5-point Likert scale format. The five friendship variables with possible responses were:

(1) how much contact do you have with this person OFF THE JOB? (1-none at all; 2-very little; 3-some contact; 4-quite a bit; 5-a great deal);

(2) how much contact do you have with this person ON THE JOB? (1-none at all; 2-very little; 3-some contact; 4-quite a bit; 5-a great deal);

(3) how well do you KNOW this person? (1-not at all; 2-not very well; 3-somewhat; 4-fairly well; 5-extremely well);

(4) how well do you LIKE this person? (1-strongly dislike; 2-dislike; 3-neither like nor dislike; 4-like; 5-strongly like); and

(5) is this person a FRIEND of yours? (1-could never by my friend; 2-not a friend; 3-merely an acquaintance; 4-is a friend; 5-is one of my best friends).

Reactions to the Use of Each Peer Assessment Method

As the final phase of the entire peer assessment procedure, the peer assessor's reactions towards each type of peer assessment were measured. A 5-point Likert rating scale was used with the available responses ranging from strongly disagree (1) to strongly agree (5). The statements were repeated for each method of peer assessment. Ratings were obtained in reaction to the following statements:

(1) this system (nominating/ranking/rating fellow officers and corporals) is a fair way to rate law enforcement personnel;

(2) I like this way of rating people;

(3) with this system people will nominate/rank highly/rate highly only their closest friends.

(4) this system will generate too much competition between officers and corporals who work together;

(5) with this system, most people will nominate/rank highly/rate highly poor performers in order to

lessen the competition when they are compared with better performers;

(6) this system will provide an accurate indication of a person's ability to perform law enforcement work; and

(7) this system should be used as one way of deciding who should be promoted.

CHAPTER III

RESULTS

Measurement Level of Assessment Methods

The level of measurement between the assessment techniques was similar. Upon application of the scoring procedure to the nomination data (as described in the Method section) all assessments--nominations, rankings, and ratings--were of an ordinal level of measurement.

Empirical evidence supports the treatment of ordinal variables as if they conform to interval scales. By assigning scores to form ordinal responses, ranging from highly skewed to equidistant systems, similar point-biserial coefficients, t-tests, and critical ratios were produced (Labovitz, 1967).

Although some small error may occur, this has been shown to be offset by (a) the use of more powerful, more sensitive, better developed, and more interpretable statistics with known sampling error, (b) the ability to retain more information about the characteristics of the data, and (c) greater versatility in statistical manipulation. Labovitz (1970) determined the degree of error of results when treating ordinal variables as if they were interval in analyzing the relation between occupational prestige (ordinal ranking)

and suicide rates. The application of 20 randomly generated scoring systems to the ordinal data showed that the correlations obtained between occupational prestige and suicide rates were interchangeable across the scoring methods. Thus, Labovitz (1970) suggested that a linear scoring system be assigned according to the available evidence on the distance between ranks and all available rank order categories be used, rather than collapsing to a smaller number.

Labovitz (1970) cautioned, however, that the actual scales of the data should be reported and interval level statistics be interpreted with care. Following Labovitz's arguments, multivariate statistics were applied to the ordinal data of the present research with due caution. A linear scoring system was applied to the nominations, rankings, and ratings to obtain equal intervals between adjacent scores.

Intercorrelations Among Performance Dimensions and Among Friendship Variables

Tables 1, 2, 3, 4, 5, and 6 present data pertaining to the intercorrelations of peer assessments and supervisor assessments among the nine performance dimensions and among the five friendship variables. The statistically significant intercorrelations indicated several aspects of the assessment data: (a) the nine performance scales (dimensions) were not seen as independent aspects of police officer performance; (b) a degree of halo error may have been present in all

Table 1

Intercorrelations Among Performance Dimensions for Peer Nominations*

Performance Dimension	Job Knowledge	Decision-making	Dealing with Co-Workers	Equipment	Dealing with Police	Communi-cation	Relia-bility	Demeanor	Work Attitude
Job Knowledge	——	.91	.83	.86	.79	.89	.87	.85	.88
Decision-making	.91	——	.80	.85	.76	.87	.87	.86	.86
Dealing with Co-Workers	.83	.80	——	.74	.79	.81	.82	.81	.83
Use of Equipment	.86	.84	.74	——	.75	.79	.84	.79	.79
Dealing with Police	.79	.76	.79	.75	——	.77	.83	.77	.73
Communi-cation	.89	.86	.81	.79	.77	——	.86	.84	.86
Reliability	.87	.87	.82	.81	.84	.86	——	.85	.81
Demeanor	.85	.86	.81	.80	.77	.84	.85	——	.88
Work Attitude	.88	.86	.83	.79	.73	.86	.81	.88	——

n = 190 peer nominations--sample composed of all persons nominated at least once

* all correlations significant beyond the .01 level.

Table 2

Intercorrelations Among Performance Dimensions for Peer Rankings*

Performance Dimension	Job Knowledge	Decision-making	Dealing with Co-workers	Equipment	Dealing with Police	Communi-cation	Relia-bility	Demeanor	Work Attitude
Job Knowledge	—	.91	.80	.86	.80	.84	.82	.81	.79
Decision-making	.91	—	.84	.88	.85	.87	.85	.85	.82
Dealing with Co-workers	.80	.84	—	.83	.85	.82	.84	.83	.85
Use of Equipment	.86	.89	.83	—	.85	.88	.85	.85	.85
Dealing with Police	.80	.85	.84	.85	—	.85	.83	.84	.81
Communi-cation	.84	.87	.82	.88	.85	—	.87	.89	.86
Reliability	.82	.85	.84	.85	.83	.87	—	.88	.85
Demeanor	.81	.84	.83	.85	.84	.89	.88	—	.88
Work Attitude	.79	.81	.85	.85	.81	.86	.85	.88	—

n = 767 peer rankings

* all correlations significant beyond the .01 level

Table 3

Intercorrelations Among Performance Dimensions for Peer Ratings*

Performance Dimension	Job Knowledge	Decision-making	Dealing with Co-workers	Equipment	Dealing with Police	Communi-cation	Relia-bility	Demeanor	Work Attitude
Job Knowledge	—	.79	.64	.71	.68	.71	.67	.64	.66
Decision-making	.79	—	.71	.81	.73	.77	.74	.73	.70
Dealing with Co-workers	.64	.71	—	.72	.72	.70	.74	.67	.70
Use of Equipment	.70	.81	.73	—	.78	.81	.75	.75	.73
Dealing with Police	.68	.73	.72	.78	—	.79	.76	.75	.66
Communi-cation	.71	.77	.70	.81	.79	—	.80	.78	.71
Reliability	.67	.74	.74	.75	.76	.80	—	.77	.73
Demeanor	.64	.73	.67	.75	.75	.78	.77	—	.75
Work Attitude	.66	.70	.70	.73	.66	.71	.73	.75	—

n = 767 peer ratings

* all correlations significant beyond the .01 level.

Table 4

Intercorrelations Among Performance Dimensions for Supervisor Rankings*

Performance Dimension	Job Knowledge	Decision-making	Dealing with Co-workers	Equipment	Dealing with Police	Communi-cation	Relia-bility	Demeanor	Work Attitude
Job Knowledge	——	.97	.85	.89	.79	.91	.83	.86	.86
Decision-making	.97	——	.86	.89	.79	.90	.83	.86	.84
Dealing with Co-workers	.85	.86	——	.83	.83	.86	.84	.82	.81
Use of Equipment	.89	.89	.83	——	.76	.85	.80	.84	.81
Dealing with Police	.79	.79	.83	.76	——	.82	.85	.76	.79
Communi-cation	.91	.90	.86	.85	.82	——	.85	.85	.88
Reliability	.83	.83	.84	.80	.85	.85	——	.80	.83
Demeanor	.86	.86	.82	.84	.76	.86	.80	——	.83
Work Attitude	.86	.84	.81	.81	.79	.88	.83	.83	——

n = 263 supervisor rankings

* all correlations significant beyond the .01 level

Table 5

Intercorrelations Among Performance Dimensions for Supervisor Ratings

Performance Dimension	Job Knowledge	Decision-making	Dealing with Co-workers	Equipment	Dealing with Police	Communi-cation	Relia-bility	Demeanor	Work Attitude
Job Knowledge	——	.87	.68	.66	.67	.80	.72	.66	.70
Decision-making	.87	——	.67	.70	.70	.80	.70	.69	.66
Dealing with Co-workers	.68	.67	——	.56	.69	.69	.71	.69	.70
Use of Equipment	.66	.70	.56	——	.68	.69	.62	.60	.53
Dealing with Police	.67	.70	.69	.68	——	.71	.77	.64	.65
Communi-cation	.80	.80	.69	.69	.71	——	.74	.71	.72
Reliability	.72	.70	.71	.62	.77	.74	——	.70	.76
Demeanor	.66	.69	.69	.60	.64	.71	.70	——	.71
Work Attitude	.70	.66	.70	.53	.65	.72	.76	.71	——

n = 263 supervisor ratings

* all correlations significant beyond the .01 level

Table 6

Intercorrelations Among Variables for Friendship Ratings*

Variable	Contact Off the Job	Contact On the Job	Know Person	Like Person	Is Person Friend
Contact off the Job	——	.32	.41	.35	.41
Contact on the Job	.32	——	.48	.28	.31
Know Person	.41	.48	——	.32	.46
Like Person	.35	.28	.31	——	.54
Is Person Friend	.41	.31	.46	.54	——

n = 957 friendship ratings

* all correlations significant beyond the .01 level

assessments across the nine dimensions; and (c) all performance dimensions were actually measuring some singular construct of police officer performance. The friendship ratings revealed significant intercorrelations among the five variables. Thus, all the friendship variables were measuring the same construct of social attraction.

Assessments using all evaluation methods (peer nominations, peer rankings, peer ratings, supervisor rankings, and supervisor ratings) across the nine performance dimensions were factor analyzed to explore the possibility of reducing the data into a smaller number of performance dimensions. The friendship ratings were also factor analyzed to detect the patterning of responses among the five variables and to possibly reduce the data to ratings on some overall construct of friendship. The factor analyses also provided a secondary measure of halo rating error across the nine performance dimensions.

A principle-factoring solution (with iterations) was used to factor analyze the correlation matrix for each method of assessment and the friendship ratings (Nie, Hull, Jenkins, Steinbrenner & Bent, 1975). In all, six factor analyses were performed.

Due to the significant intercorrelations between the nine performance dimensions using each assessment method and the significant intercorrelations between the friendship

ratings obtained on each variable, an oblique rotation of the factor matrix was used in all of the factor analyses.

With the principle-factoring solution the main diagonal elements of the correlation matrix were replaced with communality estimates (squared multiple correlations between each variable and the rest of the variables in the matrix). Thus, inferred factors were produced in an attempt to describe the underlying regularity within the data. With the iterations, the number of factors to be extracted from the original correlation matrix was determined, the communality estimates were assigned as the main diagonal elements, the same number of factors extracted from the reduced matrix, and the variance accounted for by these factors inserted as the new communality estimates. The process continued until the differences between two successive communality estimates were negligible.

All factor analyses showed similar results. For each method of assessment a single unrotated factor best described the linear relationship within the data. Since only a single factor was initially extracted, rotation of the factor matrix was not warranted. A similar result was found for the friendship ratings.

Each of the nine performance dimensions was seen to load equally, but not highly, on one general performance factor. Each of the five friendship variables also loaded equally, but not highly, on a single overall social attraction factor.

Intercorrelations Among Peer Assessments, Friendship Ratings, and Supervisor Assessments

Based on the significant intercorrelations among the nine performance dimensions for both the peer assessments and supervisor rankings and ratings, a composite measure for each method of assessment was calculated. A similar composite was constructed across the five friendship variables for identical reasons. These composite measures were then intercorrelated. The results are shown in Table 7.

As is apparent from the table, the composite measures of peer nomination, peer ranking, and peer rating were significantly correlated. Furthermore, the composite measure of the friendship ratings was significantly related to each method of peer assessment in a negative direction.

It must be noted, as presented earlier, that the peer nominations were scored so that a higher nomination score represented a higher assessment. For the peer rankings, conversely, a lower numerical ranking (i.e., a "1") indicated a higher evaluation. The same was true for the peer rating scale. Thus, the negative direction of the relationship among peer nominations and the other two methods was to be expected, making consideration of only the magnitude of the correlation important.

The composite measures of supervisor rankings and rating were also significantly intercorrelated.

The intercorrelations between the composite indices of peer assessment and friendship ratings with the supervisor assessments will not be discussed here as more in-depth analyses follow.

Table 7

Intercorrelations Among Composite Peer Assessments, Composite Friendship Ratings, and Composite Supervisor Rankings and Ratings

Composite Assessment Variable	Peer Nomination	Peer Ranking	Peer Rating	Friendship Rating	Supervisor Ranking	Supervisor Rating
Peer Nomination	——	-.51**	-.20*	-.32**	-.56**	-.51**
Peer Ranking	-.51**	——	.61**	-.30**	.59**	.52**
Peer Rating	-.20*	.61**	——	-.33**	.40**	.43**
Friendship Rating	-.32**	-.30**	-.33**	——	-.11	-.22*
Supervisor Ranking	-.56**	.59**	.40**	-.11	——	.69**
Supervisor Rating	-.51**	.52**	.43**	-.22*	.69**	——

$\underline{n}_{\text{(peer nomination total scores)}} = 190$

$\underline{n}_{\text{(peer rankings and ratings)}} = 767$

* $\underline{p}<.05$

** $\underline{p}<.01$

Correlation Between Peer Assessments and Supervisor Rankings and Rating

Based on the results of the factor analyses, composite measures across performance dimensions were formulated for peer nominations, peer rankings, peer ratings, supervisor rankings, and supervisor ratings. Each composite consisted of the sum of nominations, rankings, or ratings across the nine performance dimensions divided by nine. Thus, the assessments across performance dimensions were equally weighted. Before this was done it was determined that the variability of each type of assessment was approximately equal. A similar composite was calculated for the friendship ratings across the five variables, with five as the appropriate divisor.

Using these overall measures, multiple linear regression equations were constructed to gauge the correlation of each method of peer assessment and the friendship ratings with supervisor rankings and ratings.

Since the patrol squads differed in size, ranging from 6 to 9 members, the magnitude of the ranks assigned to respective squad members differed among squads. That is, a rank of three in a squad with six members meant something different than a rank of three in a squad of nine members. This problem, however, did not affect the calculation of the linear regression equation because the peer rankings assigned an individual were correlated with supervisor rankings and

ratings on a within-squad basis. Thus, the number of ranks which could be assigned to an individual squad member was consistent between the peer rankings and supervisor ranks, differing only by one due to self-exclusion of the peer assessor.

For each multiple regression analysis a simultaneous entering of all predictor variables was employed. All predictors (peer assessments and friendship ratings) were entered into the calculation of the linear equation at the same time without regard to any preconceived ordering. Analysis of variance was used to determine the statistical significance of the resultant regression equation.

Similar multiple regression analyses were performed deleting the friendship ratings as a predictor of supervisor rankings and ratings. The difference in the two $\underline{R}^2$'s, the squared part correlation, indicated the absolute increment of $\underline{R}^2$ due to the addition of the friendship ratings to the equation already containing the peer assessments. That is, the effect of the friendship ratings was partialed out of the relationship between the peer assessments and the criterion.

Moreover, peer nominations, peer rankings, and peer ratings (including appropriate friendship ratings) respectively, were divided into a screening sample and a calibration sample (hold-out group). This split was necessary for an empirical cross-validation of the multiple $\underline{R}^2$'s to measure

the amount of shrinkage in the statistic when the regression equations were applied to another independent sample (Lord & Novick, 1968, p. 285). The calibration sample consisted of a random selection of approximately 15% of the total sample for each peer assessment method. Although much controversy surrounds the issue of the proper size of a calibration sample, substantial consensus has indicated with a large n 15% of the total sample is sufficient for a proper determination of shrinkage in the multiple $\underline{R}^2$.

Shrinkage of the $\underline{R}^2$ was calculated by applying the standardized linear regression weights (i.e., beta weights) obtained using the screening sample for each group of assessments to the raw scores of the calibration sample. The resultant estimated criterion value was then correlated (Pearson Product-Moment correlation) with the actual criterion raw score (supervisor ranking or rating). This simple correlation was analogous to the $\underline{R}$. The difference between the squared simple correlation of the calibration sample and the $\underline{R}^2$ of the screening sample provided an estimate of the amount of shrinkage.

Correlation Between Peer Nominations and Supervisor Rankings

With the simultaneous multiple regression analysis on the screening sample (n = 160 total nomination scores) peer nominations (X_1) and "friendship" ratings (X_2) significantly predicted supervisor rankings (Y); $\underline{R}^2 = .32$; $\underline{R} = .56$;

$F(2,157) = 36.72$, $p < .01$. The squared part correlation proved that the "friendship" ratings accounted for only a minute amount of unique variance in predicting the supervisor rankings, $r^2_{y(1.2)} = .00012$.

Application of the standardized raw score weights yielded by the linear regression equation to the raw scores of the calibration sample (n = 30 total nomination scores) revealed a very small amount of shrinkage for the R^2 value (shrinkage = .08).

Correlation Between Peer Nominations and Supervisor Ratings

The multiple regression analysis with a simultaneous solution on the screening sample yielded (n = 160 total nominations) $R^2 = .27$; $R = .52$; $F(2,157) = 29.14$, $p < .01$, which indicated a significant prediction of supervisor ratings (Y) by peer nominations (X_1) and friendship ratings (X_2). The friendship ratings did not account for a significant amount of unique variance in the regression equation, $r^2_{y(1.2)} = .013$.

Cross-validation of the standardized raw score weights produced by the regression equation of the screening sample to the calibration sample (n = 30 total nomination scores) indicated some shrinkage in the R^2 (shrinkage = .19).

Correlation Between Peer Rankings and Supervisor Rankings

Peer rankings (X_1) and friendship ratings (X_2) significantly predicted supervisor rankings, $R^2 = .36$; $R = .60$;

$F(2,618) = 170.66$, $p < .01$. The regression equation was calculated on the screening sample of 621 peer rankings. The "friendship" ratings added virtually no predictive power, accounting for little unique variance in the regression equation. The squared part correlation, amount of unique variance accounted for by friendship, was $r^2_{y(1.2)} = .0044$. Shrinkage of the R^2 revealed through the empirical cross-validation using the calibration sample (n = 146 peer rankings) was quite small (shrinkage = .01).

Correlation Between Peer Rankings and Supervisor Ratings

A significant multiple correlation was found with peer rankings (X_1) and friendship ratings (X_2) predicting to supervisor ratings (Y) in the screening sample (N = 621 peer ratings), $R^2 = .28$; $R = .53$; $F(2,618) = 120.10$, $p < .01$. Friendship ratings again accounted for only an insignificant amount of unique variance in the prediction of supervisor ratings, $r^2_{y(1.2)} = .0044$.

Applying the standardized regression weights to the calibration sample of peer ratings (n = 146 peer ratings) little shrinkage in the R^2 was apparent (shrinkage = .05).

Correlation Between Peer Ratings and Supervisor Rankings

With a dependent measure of supervisor ranking (Y), the peer ratings (X_1) and friendship ratings (X_2) produced a

significant multiple R; $R^2 = .16$; $R = .40$; $F(2,618) = 57.28$, $p < .01$. Using the screening sample of 621 peer ratings, with the effects of friendship partialed out of the predictiveness of the peer ratings, little additional unique variance was left unaccounted for, $r^2_{y(1.2)} = .0003$.

The peer ratings in the calibration sample (n = 146 peer ratings) were weighted by the standardized regression coefficients produced by the linear regression equation of the screening sample. Only an insignificant amount of shrinkage in the R^2 statistic was observed (shrinkage = .10).

Correlation Between Peer Ratings and Supervisor Ratings

The linear regression analysis on the screening sample of 621 peer ratings yielded $R^2 = .19$; $R = .44$; $F(2,618) = 72.12$, $p < .01$, a significant prediction of supervisor ratings (Y) by peer ratings (X_1) and friendship ratings (X_2). However, the friendship ratings accounted for almost no unique variance, that was not accounted for by the peer ratings, in predicting the supervisor ratings, $r^2_{y(1.2)} = .007$.

Cross-validation of the standardized raw score weights produced by the linear regression equation calculated upon the screening sample to the raw scores of the calibration sample (n = 146 peer ratings) indicated little shrinkage of the multiple R^2 (shrinkage = .07).

Differences in the Relationship Between Peer Assessment Methods and Supervisor Rankings and Ratings

The determination of significant differences between the multiple R's was made using the following statistic, which is distributed approximately as t (Clement, 1979):

$$t = \frac{R_1 - R_2}{\sqrt{(SE_1)^2 + (SE_2)^2 - 2(SE_1)(SE_2)(f_{(1,2)})}}$$

The t-test provides a robust estimate of significant difference and is not significantly affected by violations of its underlying assumptions (Boneau, 1960). Thus, it provided a conservative estimate of where significant differences among R's occurred.

Comparisons of multiple R's were made for the three methods of peer assessment, obtained using both supervisor ranking and rating as performance criteria (see Table 8). A significant difference was found between the multiple R's of the peer rating method and peer ranking method. This significant difference was obtained for the multiple R's calculated with both supervisor rankings and ratings. As Table 8 presents, comparisons between the remaining multiple R's did not reveal significant differences.

Relationship of Peer Assessments to Objective Performance Indices

A test of the relationship between each method of peer assessment and each of the five objective performance indices (i.e., age, length of service, number of commendations and awards, number of on-the-job injuries, and number of sick days) was made through the calculation of Pearson product-

Table 8

Comparison of Multiple R's for All Methods of Peer Assessment

Predictors	Criterion	Multiple R	t
Peer Nomination/ Friendship Rating	Supervisor Rating	.52	2.00
Peer Rating/ Friendship Rating	Supervisor Rating	.44	
Peer Nomination/ Friendship Rating	Supervisor Rating	.52	.30
Peer Ranking/ Friendship Rating	Supervisor Rating	.53	
Peer Rating/ Friendship Rating	Supervisor Rating	.44	4.50*
Peer Ranking Friendship Rating	Supervisor Rating	.53	
Peer Nomination/ Friendship Rating	Supervisor Ranking	.56	4.21
Peer Rating/ Friendship Rating	Supervisor Ranking	.40	
Peer Nomination/ Friendship Rating	Supervisor Ranking	.56	1.33
Peer Ranking/ Friendship Rating	Supervisor Ranking	.60	
Peer Rating/ Friendship Rating	Supervisor Ranking	.40	10.00*
Peer Ranking/ Friendship Rating	Supervisor Ranking	.60	

* $p < .05$, two-tailed test

moment correlation coefficients.

The objective performance measures were available through only the personnel files of the police agency. Thus, due to concerns regarding the expense of accessing these files, objective performance indices were not obtained on all peer assessees.

A random sample of 74 total nomination scores, one score per nominee, were shown to be significantly related to the number of commendations and awards received by an officer or corporal in a positive manner, $\underline{r} = .24$, $\underline{p}<.05$.

One hundred and eight peer assessees were randomly selected on which a total of 333 peer rankings and ratings were then correlated with the respective objective performance measures for each individual. A significant negative relationship was revealed, $\underline{r}_{(ranking)}=-.14$, $\underline{p}<.01$; $\underline{r}_{(rating)}=-.12$, $\underline{p}<.05$.

In addition, peer rankings and peer ratings were shown to be significantly related to the number of on-the-job injuries in a negative direction, $\underline{r}_{(ranking)}=-.25$, $\underline{p}<.01$; $\underline{r}_{(rating)}=-.14$, $\underline{p}<.01$.

However, since a low numerical ranking or rating represented a high assessment, the direction of these correlation coefficients reflected only the difference in scaling between the peer rankings and ratings and the cumulative number of commendations and awards and number of on-the-job injuries.

Differences in the Predictive Ability of Peer Assessment Methods with Objective Performance Indices

The simple product-moment correlations which were calculated to measure the relationship between the three types of peer assessments and the objective performance indices were comparatively tested for significant differences. A test of significance for the difference between nonindependent *r*'s was used (Hotelling, 1940). The practical formula for calculating the *t* statistic is presented by Edwards (1966).

Comparison of the significant simple *r*'s yielded one significant difference. The relationship between peer rankings and number of injuries on-the-job was significantly greater than the similar relationship involving peer ratings, $t(330)=2.09$, $p<.05$, two-tailed test.

Due to the differences in scaling between the peer nominations and the peer rankings and ratings (i.e., higher peer nomination score equaled higher evaluation, lower ranking and rating equaled higher evaluation), the directionality of the relationships were disregarded in testing for significant differences. Thus, the overall magnitude of the correlation coefficients were compared.

Differences in the Predictive Ability of Peer Assessment Methods and Friendship Ratings

Another means of testing for differences among the three peer assessment techniques in their relationship with supervisor rankings and ratings was employed. Composite scores for each method of peer assessment and the friendship ratings

were determined across the nine performance dimensions and five friendship variables, respectively. These were correlated with a similar composite of supervisor ranking and rating via multiple linear regression analyses. That is, for assessees who had been nominated, ranked, rated (on performance), and evaluated as to their friendship with the assessor, a linear regression equation was developed to reveal the ability of each assessment to predict corresponding supervisor rankings and ratings. Since not all assessees had been assessed using all three peer assessment methods, the sample upon which these regression analyses were based was reduced.

Using 184 assessees, the linear combination of peer nomination, peer ranking, peer rating, and friendship rating scores significantly predicted supervisor rankings, $R^2 = .54$, $R = .73$, $F(4,179) = 51.81$, $p<.01$; and supervisor ratings, $R^2 = .37$, $R = .61$, $F(4,179) = 26.83$, $p<.01$.

Differences in the predictiveness of each method of peer assessment and the friendship ratings was determined through examination of the significance of the standardized regression weights of the regression equation using appropriate F-ratios. With the dependent variable of supervisor ranking the standardized regression weight for peer nomiations, $F(1,179) = 4.13$, $p<.05$, and peer rankings, $F(1,179) = 102.89$, $p<.01$, were significant. Predicting to supervisor ratings only peer ratings were found to be significantly weighted in the linear equation, $F(1,179) = 46.64$, $p<.01$.

Since peer ratings, unlike peer rankings, were not shown to be a significant predictor of the supervisor assessments, the

earlier finding showing the significantly greater relationship between peer rankings and the supervisor assessments, as compared to the same relationship for peer ratings, was reinforced.

Reliability of the Peer Assessment Methods

In determining the reliability of each method of peer assessment it was necessary that the reliability estimate be comparable across the three techniques. The issue of comparability of any estimate of reliability across methods was of concern due to the varying size of individual patrol squads and the nature of the performance data generated by the three methods--nominations, rankings, and ratings.

In order to make sure that the reliability estimates calculated for each method of peer assessment were similar and ultimately comparable, the following procedure was utilized. Three of the nine performance dimensions were randomly selected on which the estimates of reliability would be made. This was appropriate due to the significant intercorrelations indicating a strong relationship among the dimensions. The dimensions randomly chosen were "dealing with the public," "reliability," and "demeanor."

So that the final estimates of reliability could be compared and tested for significant differences in magnitude, three independent and random samplings of assessor pairs within squads were made. In total, 29 squads were selected in which a pair of peer assessors was randomly determined. This yielded an 80% sampling of squads.

Reliability of Peer Nominations

Past research has calculated the reliability of peer nominations through an estimation of the internal consistency or test-retest reliability using only those people who have been nominated within the single experimental sample. With this approach to estimating reliability, a method bias inherently inflates the resultant coefficient through the consideration of only those people who have been nominated--supposedly the top performers of the group. In effect the majority of the sample, those not nominated, remain undifferentiated with regard to performance. An attempt was made in the present study to overcome this past oversight.

The peer nominations were obtained through the random sampling of squads with subsequent comparison of nominees between two randomly selected peer assessors within each squad. Interrater reliability was the method of estimation employed. If a person was nominated by one member of the assessor pair, but not by the other, a score of zero was assigned for the non-nomination. Thus, the reliability estimate for the peer nomination method was based on a random sample of all possible assessees, not just those who had been identified as top performers via the nomination process.

For five of the randomly selected squads there was no agreement between the assessors as to the peer nominations,

resulting in the absence of variation. Thus, calculation of an interrater reliability estimate was impossible. Exclusion of these squads reduced the random sample to 24 squads or approximately 67% of the total number of patrol squads.

For each pair of raters, within each squad, interrater reliability estimates were calculated on each of the three selected performance dimensions. The reliability coefficients were converted into Fischer Z scores and averaged across squads for each performance dimension. These mean Z scores were then averaged across performance dimensions and converted back to a coefficient of correlation representing the interrater reliability for the peer nomination method. The interrater reliability obtained for peer nominations using 70 pairs of nomination scores was significant, $r = .48$, $p < .01$.

The corrected interrater reliability was calculated using the Spearman-Brown prophecy formula. The corrected estimate was based on an average of 10 assessor pairs per squad. This average held for all three peer assessment methods. In the case of peer nominations, the corrected interrater reliability estimate was significant, $r_{(corrected)} = .90$, $p < .01$.

Reliability of Peer Ranking

The interrater reliability for the method of peer ranking was estimated through a similar procedure as previously

outlined for the peer nomination method. It must be remembered that with the ranking method the peer assessors within each squad were allowed to evaluate only their fellow squad members. Thus, interrater reliability estimates for the peer rankings provided by the random pair of assessors were calculated on an intrasquad basis and averaged across the entire random sample of 29 squads. Using Spearman's rank-order correlation formula, the interrater reliability of the peer ranking method, using 172 pairs of rankings, was significant, $r = .62$, $p < .01$. The corrected interrater reliability estimate was also significant, $r_{(corrected)} = .94$, $p < .01$.

Reliability of Peer Ratings

Using the same methodology as outlined for the peer nomination and peer ranking methods, interrater reliability was calculated between a random pair of assessors on a randomly selected sample of 29 squads. The product-moment correlation coefficient represented the interrater reliability of the peer rating method, using 172 pairs of ratings, and was significant, $r = .23$, $p < .01$, as was the corrected estimate, $r_{(corrected)} = .75$, $p < .01$.

Differences in Reliability Between Peer Assessment Methods

To determine whether the reliability estimate of any method of peer assessment was significantly greater or less

than the others, a test of homogeneity of the three uncorrected reliability estimates (r) was performed. The uncorrected reliability estimates were used in a comparative analysis to test for significant differences in magnitude. This was done to guard against masking true differences through application of the Spearman-Brown prophecy formula. This calculation was designed to test the null hypothesis that the three uncorrected reliability estimates were homogeneous, or all estimates of the same population value. Using a technique described by Edwards (1966), it was determined that the three uncorrected reliability estimates were found not to be homogeneous, $X^2 = 21.48$, $p < .01$.

Due to the heterogeneity, the three comparisons of the uncorrected reliability coefficients were tested for significant differences. Fischer's Z transformation converted to a normal deviate of z (standardized score) was used to test for a significance of the difference between the methods' reliability estimates.

The peer rating method was found to have significantly less interrater reliability than both the peer nomination ($z = 2.04$, $p < .05$) and peer ranking ($z = 4.64$, $p < .01$) methods. The uncorrected interrater reliabilities of the peer nomination and peer ranking methods were not significantly different.

User Reactions to the Peer Assessment Methods

Reactions of the peer assessors towards each of the evaluation methods were obtained as the final section of the peer assessment instrument package. A total of 143 respondents provided ratings as to their agreement/disagreement with each of the seven statements.

Table 8 presents the mean response to each statement for each assessment method. Most statements generated mild to strong disagreement. Disagreement with a statement, however, did not necessarily indicate a negative reaction to the method of peer assessment. The trend towards disagreement on items four and five revealed a slight tendency to believe that none of the methods would generate too much competition among peers and that people would not nominate, rank highly, or rate highly poor performers to make themselves look good in the final analysis. Moreover, the mean response to item three suggested neither agreement nor disagreement with the statement. That is, the respondents were not sure whether only close friends would be nominated, ranked highly, or rated highly. Thus, the trend towards negative reactions did not hold for item three.

In order to test whether the reactions to any one peer assessment method were significantly different from the reactions to the others, a multivariate analysis of variance (MANOVA), as packaged by Finn (1968), was performed. MANOVA

Table 9

Mean and Standard Deviation of Responses to User Reaction Statements by Peer Assessment Method*

Statement	Peer Assessment Method					
	Peer Nomination		Peer Ranking		Peer Rating	
	Mean	SD	Mean	SD	Mean	SD
(1) This system (nominating/ranking/ rating fellow Officers or Corporals) is a fair way to rate law enforcement personnel.	2.2	1.3	2.3	1.3	2.3	1.2
(2) I like this way of rating people.	2.0	1.1	2.1	1.2	2.1	1.1
(3) With this system people will nominate/rank/rate highly only their closest friends.	3.4	1.1	3.2	1.2	3.1	1.2
(4) This system will generate too much competition between Officers and Corporals who work together.	2.4	.9	2.4	.9	2.4	.9
(5) With this system, people will nominate/rank/rate highly poor performers to lessen the competition when they are compared to the better performers.	2.3	1.0	2.2	1.0	2.3	1.0
(6) This system will provide an accurate indication of a person's ability to perform law enforcement work.	2.1	1.1	2.2	1.2	2.2	1.2
(7) This system should be used as one way of deciding who should be promoted.	2.1	1.2	2.2	1.2	2.2	1.3

* Response range on a 5-point Likert scale--1 (strongly disagree) to 5 (strongly agree).

was chosen due to the obvious relationships between responses across the seven statements. The Finn (1968) computer analysis yielded multivariate, univariate, and step-down (covariate) F-ratios, with the appropriate degrees of freedom.

A significant multivariate F-ratio was not obtained for the reaction data. Thus, across the seven statements, responses regarding any one of the peer assessment methods were not significantly different from those pertaining to the other two.

Spector (1977) has described the utility of inspecting the univariate and step-down F-ratios when a significant multivariate F has been found. In the present study these statistics were computed in spite of a nonsignificant multivariate F-ratio. However, no significant univariate F-ratios were obtained.

CHAPTER IV

DISCUSSION

The Generality of Results

Organizational Setting

The organizational environment of the study could be described as paramilitary, typical of most municipal police agencies. In such an environment certain military traditions survive (e.g., deference to rank, explicit organizational chain of command, etc.). The paramilitary setting may have had an influence on validity due to the close on-the-job contact between patrol squad members. Kane and Lawler (1978) have argued that the commonly reported high validities of the peer nomination method may have been an artifact of the militaristic setting in which the performance data were gathered. They suggest in such organizations more on-the-job contact is experienced among members, thus improving the validity of the peer nominations.

In the present study this influence may have surfaced in all of the peer assessment methods. Since a patrol squad basically functions as an autonomous unit, responsible for the security of a specified geographic area, all assistance needed by a patrol officer at any time is provided by

fellow squad members. Therefore, frequent and unique opportunities exist for fellow squad members to observe significant on-the-job behaviors of their peers.

In the opinion of this author, however, the degree of on-the-job contact displayed by police officers is generalizable to other nonpolice organizations. Similar amounts of contact most probably exist in organizations where workers are divided into small departments or production work units. Thus, the findings of the present study are applicable to a variety of production and service organizations.

Moreover, in such a setting the influence of the chain of command upon a police officer's day-to-day performance is substantial. Therefore, the issue of whether the supervisor may influence the peer assessments, as reported in the present study, becomes a consideration. There exists a remote possibility in any organization wherein a supervisor may attempt to convince subordinates to inflate their assessments of fellow workers in an attempt to glorify the work group as a whole. As there was no actual test for this bias in the present study, its possibility remains a consideration.

Predictive vs Concurrent Study Design

Most past research conducted in a military setting utilizing peer assessments attempted to predict future performance--32 predictive designs, 5 concurrent designs (Kane & Lawler, 1978). That is, peer assessments taken early in the life of the assessee group were correlated with some later measure of success (i.e., admission to officer candidate

school (OCS), promotion, superior ratings, etc.). Thus, the correlations in these studies reflected the relationship between predictor performance measures and a distant performance criterion.

As the present study employed a concurrent design within a paramilitary setting, the correlations between the peer assessments and the supervisor rankings and ratings were expected to be somewhat greater in magnitude than those reported in predictive investigations conducted in a similar setting (see Kane & Lawler, 1978, pp. 559, 571, 580). This expectation was based on the fact that the peer assessments and supervisor judgments were all collected within a short period of time. It was anticipated that since these measures were obtained during ashort time span, they would correlate more highly than peer assessments and criteria which were gathered at two separate and distant times. This indeed was the case.

Yet, the validities of peer nomination scores have been reported to hold up for much longer periods than most other psychometric procedures, such as scores on a typical performance test. Hollander (1965) commented that significant validity of nomination scores was still evident after three years. Amir et al. (1970) and Mayfield (1972) revealed a similar lack of decrease in predictiveness for peer nominations. It is interesting to note that the findings of these authors are in direct contrast with the usual phenomenon whereby the predictiveness of psychometric devices decay for

criteria measured farther in the future (Alvares & Hulin, 1973).

The ability of peer nominations to not decay in predictiveness may apply to both peer rankings and ratings. If so, one would expect the predictiveness of the peer assessments inferred via the concurrent design of the present study to not significantly decrease, even if the criteria had been measured at a more distant time. Even though true predictiveness is not possible with a concurrent design, based on past research the inference that a peer assessment is predicting future performance via a concurrent study seems appropriate.

Administrative vs. Research Set in the Collection of Peer Assessments

As Kane and Lawler (1978) present, only a few studies in the past have indicated that assessors were told that their assessments would be used for administrative decisions (e.g., promotions, pay, and hiring decisions). The majority of the studies have assured assessors that their peer evaluations were to be used strictly for research purposes.

Whereas the present study collected the peer assessment data as part of a criterion-related test validation study, the context of the present investigation falls into the research set. Hollander (1957) has shown, however, that the rationale for the collection of peer nominations does not have a significant impact on the reliability of the method. That is, whether peer nominations are to be used

for administrative decisions or strictly research should not significantly affect their reliability. Yet, this factor still presents a constraint on the generalizability of the present research in that previous investigations have not dealt with the effects of the experimental rationale (research orientation) on either peer rankings or peer ratings. Thus, overall conclusions regarding differential effectiveness of peer assessments under these two conditions of administration must be made with caution, not only with regard to reliability but to other psychometric properties as well.

The Problem of Criteria

It must be noted that the use of supervisor judgments as criterion in determining the strength of the relationship between peer assessments and supervisor rankings and ratings (i.e., validity) places some limimtations on the accuracy of the estimate. Even though supervisors' judgments seem to be quite prevalent in the determination of validity for peer assessment methods, supervisor rankings and ratings are subject to the usual types of assessment errors (i.e., halo error, leniency error, central tendency error, etc.).

Reliance on more objective measures of performance as criteria (i.e., promotion, awards and commendations, etc.) also has certain drawbacks. The major problem with such criteria is the inability to disregard all influences on these measures which are beyond the control of the assessee. That is, promotion of an employee to a new position may have been a function of reorganization rather than good performance.

The answer to the delineation of the proper criterion for measuring validity seems to lie in a comprehensive measure of considering both individual and organizational performance which utilizes both judgmental and objective information. Campbell et al. (1970) have noted that:

> A person should be evaluated through his or her impact on the organization's continued functioning through the optimal acquisition and utilization of internal and external resources. (p. 125)

The basis of this conception becomes one of fulfilling previously stated and defined performance objectives. There seems to be no reason why some type of goal-setting system measuring both subjective and objective standards of performance would not be applicable in the case of police officer evaluation, or for most workers for that matter.

Use of the Assessment Instruments

Intercorrelation Among the Performance Dimensions

A finding which bears consideration was the high intercorrelations revealed for both peer assessments and supervisor assessments, among the nine performance dimensions. Two explanations may be offered in regard to this matter.

First, the peer assessors and supervisor assessors may have exhibited a high degree of halo in their assessments. That is, a person was given a high assessment on all

nine performance dimensions based on his or her high performance in only one of the areas. The usual way of decreasing halo is through assessor training. The peer assessors and supervisor assessors in the present study had not received training in the use of the performance appraisal instrument which incorporated several distinct dimensions of performance. If familiarity existed at all, it was with the evaluation process as a whole gained through previous experiences using a more global performance appraisal instrument. Most likely assessor training in the proper use and interpretation of the multi-dimensional performance appraisal instrument would have reduced the halo which was evidenced in all assessments.

Secondly, the assessors may have believed that all the people who actually "make it" as a patrol officer are well seasoned, top notch personnel. That is, the assessors saw only eminently qualified personnel serving around them, each assessee being a good performer on each of the nine performance areas.

Overall, it is interesting to note that the supervisors were not any more capable of accurately discriminating among behaviors classified under the nine performance categories than were the peer assessors. In the very least, the peer assessors were no worse than their supervisors in utilizing the appraisal instrument.

Integration of Peer Assessments as an Organizational System

Another factor for consideration concerns speculation as to what would happen to the peer assessments after they had been actually implemented in a real organizational setting. It is possible that over time peer assessments may be subjected to biases which have been shown to affect supervisor evaluations. Such biases as political motivations for evaluations, assessor errors, and self-serving motivations for evaluations could affect the high degree of validity and reliability revealed through the present research. It would seem unlikely that peer assessments could escape the influence of these omnipresent factors of evaluation spurred via the informal organization.

Future research must examine the actual use of peer assessment for organizational purposes. This research must also trace the validity and reliability of the peer assessment measures over time as they become an integral part of the organizational environment.

Results of the Factor Analyses

Factor analyses on all assessment methods (both peer and supervisor) indicated that the judgments made across the nine performance dimensions were actually tapping some single construct--most likely overall police officer performance. That is, an effective officer or corporal was not one who excelled in every area of performance (as portrayed by the nine performance dimensions), but one who "got the job

done" in general. Even though getting the job done may entail many combinations of differing levels of performance across the nine performance areas, these were not reflected in the actual assessments.

The factor analysis of the friendship ratings indcated that each of the five variables was measuring a single factor, most logically some construct of social attraction in evidence between peer assessor and assessee. It was unfortunate that only a single aspect of this complex phenomenon had been tapped. Initially it seemed that the five variables were independent and equal parts of what the peer assessors saw as friendship between fellow officers and corporals. But in reality they all were measuring some composite facet of attraction.

Therefore, the present assessment instruments were not capable of eliciting detailed performance and social attraction data. As was presented earlier, the performance dimensions were based on an in-depth task-based job analysis of the police officer position (Love, 1978). From this present research it would seem that the development of a peer or supervisor assessment instrument from sound job analysis data does not guarantee its utility for evaluating performance. Perhaps, other factors play more crucial roles in the proper functioning of an appraisal instrument, such as: the motivation of the assessor to assess accurately (Guion, 1965), the amount of training in using the appraisal instrument and proper assessment procedures, the assessor's

conception of the position to be evaluated, and the assessor's perception of the proper social relationship between fellow workers.

Differential Relationships Between the Peer Assessments and Supervisor Rankings and Ratings

All the methods of peer assessment--peer nomination, peer ranking, and peer rating--revealed significant correlation with supervisor rankings and ratings, which were inferential measures of the validity of each method. Moreover, for all three peer assessment techniques, higher correlations were observed with supervisor rankings and ratings than with the objective indicators of performance.

Consistent with past research (Borman, 1974; Freeberg, 1969; Gordon & Medland, 1965; Hoffman & Rohrer, 1964; Ricciuti, 1955; Springer, 1953; Swanson & Johnson, 1975; Trites, 1960; Tucker, Cline & Schmitt, 1967; Wiggins, Blackburn & Hackman, 1967), both peer nominations and peer rankings showed greater validity (higher correlations) than peer ratings. A significant difference in validity was observed between peer rankings and peer ratings.

The greater inferred validity of the peer ranking method may have been prompted by its high degree of reliability. The reliability of the peer ranks was much greater than that of the other meathods of peer assessment. Since reliability sets the upper bound for the estimate of validity (correlation coefficient), the two results are consistent.

It must be noted in regard to the inferred validity

CONTINUED

1 OF 2

of each method of peer assessment that the present research did not utilize explicit leadership-related performance assessments or criterion. Past research which has reported the validity of peer nominations (Hollander, 1954a, 1954b) has centered on the calculation of correlations using leadership predictors and military-based leadership/promotion criterion. Even though this had been most prevalent with peer nomination research, with the advent of the present comparative study of peer assessment methods, caution must be exercised when the validity coefficients of past studies are compared with those found in the present research.

While leadership is an important factor in the job of police officer, on the basis of the task-based job analysis, a leadership performance dimension was not warranted. Instead, leadership components were represented in many of the dimensions such as communication, decision-making, and demeanor (see Appendix 1). Thus, while caution must be advised, the correlation coefficients of the present research which represented the validity of each method of peer assessment may indeed be constrasted with those studies which have incorporated leadership-related measures.

The Influence of Friendship on the Relationship Between Peer Assessments and Supervisor Rankings and Ratings

One of the most noteworthy findings of the present investigation was the absence of any apparent biasing effects due to the social attraction or friendship between peer assessor and assessee. Based on past research, it was

hypothesized that friendship would affect peer nominations to a greater extent than the other peer assessment methods. This hypothesis implicitly stated that friendship would indeed have some sort of biasing effect on the inferred validity of the peer assessments. This expected finding was noticeably absent in the present research. The second part of the stated hypothesis, moreover, indicated that friendship should have similar biasing effects across the three peer assessment methods. This was indeed the case. Friendship had no effect on the correlation between any of the peer assessments with either supervisor rankings or ratings.

Furthermore, it was found that the friendship ratings did not explain any unique variance in the relationship between the peer assessments and the supervisory criterion. Thus, it would seem that the friendship ratings were measuring a construct which was irrelevant to the peer assessments. That is, being a close or distant friend of the peer assessor was not a factor which distorted the accuracy of the peer assessor's judgments in evaluating a fellow officer's or corporal's performance, as compared to the supervisor's judgments.

Peer assessments have long been scorned due to a strong belief that they are extremely prone to error induced by the social attraction between the peer assessor and the peer assessee. The present study seemed to suggest that this was not the case with the sample studied here. This is not

to say that a certain facet of social attraction, other than what was measured through the friendship ratings, may have had a significant effect on peer assessments. Yet, in this study, one's contact off the job or on the job with the assessee, knowledge of the person, liking of the person, and whether the person was considered as a friend did not significantly influence the accuracy of the peer evaluation process, as compared to supervisor judgments. In light of the close contact among squad members, which would imply close friendships, this finding is quite important.

There was a difference, however, in the way friendship was measured in the present study and as compared with past research. (Hollander, 1955; 1956). The present investigation measured friendship using responses to five questions, each scored on a 5-point continuous scale. Past research has utlilized a method whereby the peer assessor (usually a peer nominator) simply listed those people in the assessee group who were his or her friends. This technique, in the view of this author, implies that social attraction between fellow workers is an all or none phenomenon. The technique discounts the reality that people have all "degrees" of friends, from mere acquaintances to intimate friends. With many types of friends, a continuous measure would seem more appropriate.

Disregarding the measure of friendship at hand, the present study failed to support the notion that peer assessments can be easily influenced by several social attraction

variables. Whether it was the method of ascertaining the degree of friendship between assessor and assessee in the present study that led to this intriguing finding cannot be determined at this time. What is significant, however, is the attempt to describe in detail the basic constructs or what we call friendship among fellow employees.

The Relationship Between Peer Assessments and Number of Commendations and Awards and Work-Related Injuries

Commendations and Awards

Peer nominations revealed a statistically significant and positive relationship with the number of commendations and awards an officer or corporal had received. In light of the scale used for collecting the peer rankings and ratings, with one (1) being a high assessment, it can be stated that a simlar positive relationship was revealed for peer rankings and ratings.

In order to adequately consider the implications of these findings one must describe the impact of receiving a commendation or award from the police department upon one's reputation within the organization. As with most police organizations, commendations and awards were received as a result of single incidents of distinguished performance. Along with the commendation or award comes deserving praise through departmental announcements and media coverage. Bearing this in mind, the relationships with the peer assessments become more easily interpreted.

The significant relationship between peer nominations, peer rankings, and peer ratings and the number of commendations and awards received seemed to indicate that there was some connection between a person's reputation within the police agency and the frequency of assigned nominations and the magnitude of the rankings and ratings.

All peer assessments, therefore, were substantially influenced by the number of awards and commendations received by an officer or corporal. The peer assessments did not predict whether an award or commendation would be received by an individual; instead, they seemed to be a direct function of the assessee's reputation.

The impact of one's reputation on the peer assessments was in constrast with the absence of any effect of friendship (measured via the friendship ratings) on the relationship between peer assessments and supervisor rankings and ratings. If peer assessments were indeed related to a measure of popularity within the police organization, it was unclear why this social-attraction construct had no effect on the prediction of supposedly stable supervisor evaluations.

It seemed that two distinct types of social attraction had been identified in relation to peer assessments. Past research which has investigated the influence of friendship, most notably that of Hollander (1955, 1956) and Waters and Waters (1970), has utilized a reputation-based concept of friendship in determining its impact on the validity of peer nominations. Both researchers had subjects identify their friends within the assessee group and then

examined the subsequent nomination scores accorded those individuals. It was likely that the determination of friends within the assessee group was influenced by their respective reputation. Past sociological research has shown that it is possible to gain a positive evaluation in a group through association with a highly regarded other. Perhaps this motive played a part in the selection of "friends" in these studies.

The present investigation, on the other hand, utilized ratings of various aspects of the social relationship between fellow workers. These friendship ratings did not directly tap the reputation of the peer assessee. Unfortunately, due to the small amount of research documenting the effect of social attraction on the validity of peer assessments, no definitive statement regarding the best method of ascertaining friendship can be made. Future investigations are needed, not only to clearly delineate the constructs of the phenomenon called friendship between fellow workers, but to compare and contrast friendship with one's reputation.

Work-Related Injuries

Bearing in mind the scales used for the ranking and rating data, the negative relationship found between peer rankings and ratings and on-the-job injuries indicated that those officers and corporals ranked and rated highly as to their performance by their fellow squad members experienced more injuries. It would seem that if the job of patrol officer is performed at a high level of proficiency, the officer or

corporal is more apt to find themself in situations where injuries occur. Moreover, the good performer would most probably put forth that extra physical effort in certain aspects of the job (i.e., subduing a suspect, pushing a disabled automobile, etc.), thus providing more opportunities for injuries to occur.

Differential Reliability of Peer Assessment

Consistent with past research (Bartlett, 1959; Borman, 1974; Campbell et al., 1970; Freeberg, 1969; Springer, 1953; Stahl & Steger, 1977), the reliability of peer ratings was poor when compared to that of peer nomination and peer ranking. Kane and Lawler (1978) have suggested that peer ratings are generally less reliable because of the possibilities of more disagreements in assigned ratings among peer raters.

Moreover, certain errors were more probable in the peer rating process which were not of concern with peer nominations and peer rankings (Cummings & Schwab, 1973). Since peer rating was a method which required an assessor to compare an assessee to some absolute standard of performance it was subject to additional interindividual errors in assessment (i.e., leniency, strictness, and central tendency). These errors may have been repeated across assessees, as equal ratings could have been given to any number of squad members.

Peer ranking required all squad members to be assigned a rank, thus forcing discriminations. As no ties

were allowed, all squad members were distinguished from one another by their assigned ranks. Peer ranking, therefore, was not subject to the interindividual constant errors which may have affected the peer ratings.

The peer nomination method, on the other hand, was also not affected by these additional errors. It forced the assessor to identify, and rank order only three high performers within the total assessee group, across all squads. No ties or repetitions were allowed. Thus, peer nomination became, in effect, an incomplete ranking procedure.

The interindividual constant errors in evidence with peer ratings were not apparent in the peer nomination and peer ranking methods. This fact would seem to account, in part, for the reduced reliability of the peer rating technique in the present study.

An issue which was only partially addressed by the present study concerns the determination of reliability estimates for each method of peer assessment. Kane and Lawler (1978) strongly advocated the use of coefficient alpha in the calculation of internal consistency estimates for peer assessment. The advantages of coefficient alpha over the split-half reliability estimate are minimal at best. The real issue is whether either estimate is indeed the appropriate means of determining reliability for all methods of peer assessment, most notably for peer nomination.

Since the peer nomination process does not discriminate between all members of the assessee group, identifying

only the best (or worst) performers, split-half or coefficient alpha estimation of reliability can be applied to only those persons who have received at least one nomination. This reduction of the total sample of possible assessees through a selection process, on which the subsequent reliability estimate is based, would seem to inflate the coefficient. Moreover, this determination of reliability does not take into account the fact that by not nominating an eligible assessee, the assessor is implicitly assigning a nomination score (most probably zero) to that individual. Therefore, all members of the assessee group have been assigned scores by all of the assessors and should rightfully be considered in estimation of reliability for the method.

Through the calculation of interrater reliability using a randomly selected group of assessor pairs, the present study attempted to circumvent this inherent difficulty of the peer nomination process. Unfortunately, the issue of the proper estimation of reliability for peer nominations has not been solved. In many instances a reliability estimate could not be calculated due to total disagreement among the assessors. Thus, the random sample was reduced in size making the reliability coefficient less stable. More research is needed to determine the psychometric idiosyncracies of the reliability of peer nominations.

Reactions to the Methods of Peer Assessment

Overall, the peer assessors expressed mildly negative reactions to all three peer assessment techniques--peer

nomination, peer ranking, and peer rating. They indicated that peer assessment was not fair, accurate, liked by them, nor should it be used in promotion decisions. Moreover, peer assessments would not generate too much competition between fellow workers and would not be used by the peer assessor to present a fellow worker in a bad light to lessen competition. The peer assessors, however, were not sure whether close friends would be assessed most highly.

As was presented earlier, not all persons asked to participate in the study did so. Thus, there may have been a bias evident within the reactions provided by the nonrandom sample of police personnel. It was possible that those persons who did not respond did so out of extreme dislike for the peer assessment methodology. If so, these persons most probably would have supplied even harsher, more negative reactions.

CHAPTER V

CONCLUSIONS

Proper Application of Peer Assessment Methodologies

It was apparent in the present study that all three methods of peer assessment--peer nomination, peer ranking, and peer rating--revealed substantial degrees of inferred validity and reliability when used for evaluation of law enforcement officers. The magnitude of the relationship between each method of peer assessment and supervisor rankings and ratings, along with certain objective performance indices, was greater than most similar correlation coefficients reported in past studies (see Kane & Lawler, 1978). Such was the case for the interrater reliability estimates as well.

In any study utilizing a method of peer assessment, regardless of the setting, the inherent nature of the evaluative technique cannot be disregarded. That is, each method of peer assessment seems to be suited to different assessment needs.

Peer nomination, for example, is able to identify persons with extreme levels of performance on the appropriate

dimensions or traits. It fails, however, to allow any differentiation among the majority of assessees. The larger the group upon which nominations are allowed, the greater the number of assessees who will be indistinguishable from one another and clustered into a single group. The present study was an example of this happening. As such, peer nomination would seem to be most appropriate when only promotional decisions are to be made. That is, if only the very top performers within the overall assessee group need to be recognized, peer nomination would be the choice method. Peer nomination may also find an appropriate use in the early identification of outstanding managerial or administrative talent. Previous research has shown that early nominations retain their accuracy over time, up to three years. During this era where cost considerations make it crucial to identify top performers early in their careers in order to provide subsequent training and development, peer nominations may be one answer for doing so accurately and reliably.

Even though there is a lack of past research dealing with the peer ranking method, the present study presented a sound relationship with supervisor rankings and ratings and adequate reliability. Rankings on the whole discriminate quite well among all members of a group, thus creating and documenting an entire range of performance among group members. The fact that rankings force discriminations among assessees who may be actually close in performance proficiency should

not overshadow the accuracy and reliability of peer rankings reported in the present study.

Peer rating, as with rating techniques in general, seems to be the most widely applicable method of peer assessment. That is, it is appropriately used in many types of organizational settings to fulfill many assessment needs. The present investigation yielded empirical support showing the close relationship between peer ratings and supervisor rankings and ratings, upon which the validity of the method was inferred. Substantial reliability for the peer rating method was also reported. As a rating method, peer rating is quite conducive to providing feedback, which may be the major fact which distinguishes it from both peer nominations and peer rankings.

Social Attraction Among Fellow Workers

In the present investigation, friendship did not prove to be a serious problem in affecting the inferred validity of any of the three peer assessment techniques using supervisor assessments as criteria. Friendship was shown to have almost no effect on the relationship between each peer assessment and supervisor judgments. However, the peer assessment scores may have been influenced by a certain aspect of the social attraction between fellow officers. This conclusion is based on the close relationship between peer assessment scores and the number of commendations and awards received by an officer or corporal. One's reputation within the police agency, a function of receiving awards and

commendations, seemed to affect the peer assessments made for that individual.

Throughout the present study it was noted that several aspects of friendship were most likely in evidence within the workplace. Those measured using a continuous scale were contact on and off the job, liking of assessee, knowledge of person, and actual friendship. Yet, another aspect of social attraction was apparent in the significant effect commendations and awards received had upon the magnitude of the peer assessment scores. This was termed "reputation within the police agency." This factor of social attraction obviously had not been tapped by the five variables of friendship collected within the peer assessment instrument. Thus, the whole matter of what constructs of social attraction lend themselves to the relationship among fellow workers must be dealt with in future research. This issue is not only one which faces the impact of one's reputation on peer assessments, but may also affect in some way all types of organizational evaluation.

Furthermore, it was apparent that the friendship ratings were not discriminating among all degrees of social attraction. Future investigations must incorporate a method such as ranking for gathering social attraction data. If rankings could be obtained on specified friendship-related variables one would be assured that an entire range of social attraction among the assessees would be identified.

Use and Improvement of Peer Assessments

The peer assessment methods were not unique in requiring the cooperation of the peer assessors to be effective. In spite of the documented negative reactions to all methods, the peer assessments were accurate and reliable measures of performance. In this regard, the present study portrayed the first systematic investigation into user reactions to peer assessments. Even though the overall reactions were negative, this must not stop further development and refinement of peer assessment technology. As it stands, peer assessment represents a novel, basically untried performance appraisal system. As such, much of the negative reactions evident in the present research could be traced to unfamiliarity with this approach towards performance evaluation. Perhaps with continued use, peer assessment would be regarded along with supervisor rating as a bona fide means of assessing employee performance.

Finally, peer assessment should optimally be viewed as only one part of a comprehensive performance appraisal system. Such a system should encompass views of performance from all levels of the organization, including supervisor, peer, and self assessments. With this in mind, the issue then becomes one of which aspects of performance should be measured via the peer assessment component. The present study did not fully address this problem as the same dimensions of performance were used for both peer and supervisor evaluation. Ideally, peer assessment would utilize only those

aspects of performance which are uniquely observed by fellow workers. The same would be true for both supervisor and self appraisals. Thus, each portion of the overall evaluation system would measure unique aspects of an employee's performance. Upon combination of all these assessments, the most accurate and detailed picture of a person's past and present performance would be obtained.

Overall, it would seem that the time has come to consider peer assessment as a useful tool within a comprehensive organizational performance appraisal system. Only through detailed accounts of such use will acceptance of peer assessment be improved. Research seems to indicate the potential usefulness of these three different methods of peer assessment. It is time this potential be realized in a real organizational sense.

LIST OF REFERENCES

Alvares, K. M., & Hulin, C. L. An experimental evaluation of a temporal decay in the prediction of performance. Organizational Behavior and Human Performance, 1973, 9, 169-185.

Amir, Y., Kovarsky, Y., & Sharon, S. Peer nominations as a predictor of multistage promotions in a ramified organization. Journal of Applied Psychology, 1970, 54, 462-469.

Aronson, E., & Linder, D. Gain and loss of esteem as determinants of inter-personal attraction. Journal of Experimental Social Psychology, 1965, 1, 156-171.

Bartlett, C. J. The relationships between self-ratings and peer ratings on a leadership behavior scale. Personnel Psychology, 1959, 12, 237-246.

Berkowitz, L., & Levy, B. I. Pride in group performance and group-task motivation. Journal of Abnormal and Social Psychology, 1956, 53, 300-306.

Berscheid, E., Dion, K., Walster, E., & Walster, G. W. Physical attractiveness and dating choice: A test of the matching hypothesis. Journal of Experimental Social Psychology, 1971, 7, 173-189.

Boneau, C. A. The effects of violations of assumptions underlying the 't' test. Psychological Bulletin, 1960, 57, 49-64.

Booker, G. S., & Miller, R. W. A closer look at peer ratings. Personnel, 1966, 43, 42-47.

Borg, W. R., & Hamilton, E. R. Comparison between a performance test and criteria of instructor effectiveness. Psychological Reports, 1956, 2, 111-116.

Borman, W. C. The rating of individuals in organizations: An alternate approach. Organizational Behavior and Human Performance, 1974, 12, 105-124.

Brehm, J., & Festinger, L. Pressures toward uniformity of performance in groups. Human Relations, 1957, 10, 85-91.

Byrne, D. Attitudes and attraction. In L. Berkowitz (Ed.), Advances in experimental social psychology. Vol. 4. New York: Academic Press, 1969.

Bryne, D., & Nelson, B. A. Attraction as a linear function of proportion of positive reinforcements. Journal of Personality and Social Psychology, 1965, 1, 659-663.

Campbell, J. P., Dunnette, M. D., Lawler, E. E., & Weick, K. E. Managerial behavior, performance, and effectiveness. New York: McGraw-Hill, 1970.

Cascio, W. F., & Valenzi, E. R. Relations among criteria of police performance. Journal of Applied Psychology, 1978, 63, 22-28.

Cavior, N., & Dokecki, P. R. Physical attractiveness self-concept: A test of Mead's hypothesis. Paper presented at the 79th Annual Meeting of the American Psychological Association, Washington, D. C., 1971.

Clement, D. E. Personal communication, June 15, 1979.

Cox, J. A., & Krumboltz, J. D. Racial bias in peer ratings of basic airmen. Sociometry, 1958, 21, 292-299.

Cummings, L. L., & Schwab, D. P. Performance in organizations: Determinants and appraisal. Glenview, Illinois: Scott, Foresman and Co., 1973.

Deutsch, M., & Solomon, L. Reactions to evaluations by others as influenced by self-evaluations. Sociometry, 1959, 22, 93-112.

Dion, K., Berscheid, E., & Walster, E. What is beautiful is good. Journal of Personality and Social Psychology, 1972, 24, 285-290.

Doll, R. E., & Longo, A. E. Improving the predictive effectiveness of peer ratings. Personnel Psychology, 1962, 15, 215-220.

Downey, R. G., Medland, F. F., & Yates, L. G. Evaluation of a peer rating system for predicting subsequent promotion of senior military officers. Journal of Applied Psychology, 1976, 61, 206-209.

Edwards, A. L. Experimental design in psychological research. New York: Holt, Rinehart and Winston, 1966.

Finn, J. D. Multivariance: Univariance and multivariate analysis of variance, covariance, and regression; Version IV. Ann Arbor, Mich.: National Education Resources, 1968.

Fiske, D. W., & Cox, J. A. The consistency of rating by peers. Journal of Applied Psychology, 1960, 44, 11-17.

Freeberg, N. E. Relevance of rater-ratee acquaintance in the validity and reliability of ratings. Journal of Applied Psychology, 1969, 53, 518-524.

Gordon, L. V. Estimating the reliability of peer ratings. Educational and Psychological Measurement, 1969, 29, 305-313.

Gordon, L. V., & Medland, F. F. The cross-group stability of peer ratings of leadership potential. Personnel Psychology, 1965, 18, 173-178.

Guion, R. M. Personnel testing. New York: McGraw-Hill, 1965.

Gullahorn, J. T. Distance and friendship as factors in the gross interaction matrix. Sociometry, 1952, 15, 123-134.

Gunderson, E. K. E., & Nelson, P. D. Criterion measures for extremely isolated groups. Personnel Psychology, 1966, 19, 67-80.

Hackman, J. R., & Suttle, L. Improving life at work. Santa Monica, Calif.: Goodyear, 1977.

Hoffman, E. L., & Rohrer, J. H. An objective peer evaluation scale: Construction and validity. Educational and Psychological Measurement, 1954, 14, 332-341.

Hollander, E. P. Buddy ratings: Military research and industrial implications. Personnel Psychology, 1954, 7, 385-393. (a)

Hollander, E. P. Peer nominations on leadership as a predictor of the pass-fail criterion in naval air training. Journal of Applied Psychology, 1954, 38, 150-153. (b)

Hollander, E. P. The friendship factor in peer nominations. Personnel Psychology, 1956, 9, 435-447. (a)

Hollander, E. P. Interpersonal exposure time as a determinant of the predictive utility of peer ratings. Psychological Reports, 1956, 2, 445-448. (b)

Hollander, E. P. The reliability of peer nominations under various conditions of administration. Journal of Applied Psychology, 1957, 41, 85-90.

Hollander, E. P. Validity of peer nominations in predicting a distant performance criterion. Journal of Applied Psychology, 1965, 49, 434-438.

Hollander, E. P., & Webb, W. B. Leadership, followership, and friendship: An analysis of peer nominations. Journal of Abnormal and Social Psychology, 1955, 50, 163-167.

Hotelling, H. The selection of variates for use in prediction, with some comments on the general problem of nuisance parameters. Annual Review of Mathematics and Statistics, 1940, 11, 271-283.

Jones, S. C., Knurek, D. A., & Regan, D. T. Variables affecting reactions to social acceptance and rejection. Journal of Social Psychology, 1973, 90, 269-284.

Kane, J. S., & Lawler, E. E. Methods of peer assessment. Psychological Bulletin, 1978, 85, 555-586.

Kaufman, G. G., & Johnson, J. C. Scaling peer ratings: An examination of the differential validities of positive and negative nominations. Journal of Applied Psychology, 1974, 59, 302-306.

Kavanagh, M. J., MacKinney, A. C., & Wollins, L. Issues in managerial performance: Multitrait-multimethod analyses of ratings. Psychological Bulletin, 1971, 75, 34-49.

Keaveny, T. J., & McGann, A. F. A comparison of behavioral expectation scales and graphic rating scales. Journal of Applied Psychology, 1975, 60, 695-703.

Kipnis, D. Interaction between members of bomber crews as a determinant of sociometric choice. Human Relations, 1957, 10, 263-270.

Kirk, R. E. Experimental design: Procedures for the behavioral sciences. Belmont, Calif.: Brooks/Cole, 1968.

Kleiner, R. The effects of threat reduction upon interpersonal attraction. Journal of Personality, 1960, 28, 145-155.

Klimoski, R. J., & London, M. Role of the rater in performance appraisal. Journal of Applied Psychology, 1974, 59, 445-451.

Korman, A. K. The prediction of managerial performance: A review. Personnel Psychology, 1968, 21, 295-322.

Kraut, A. I. Prediction of managerial success by peer and training-staff ratings. Journal of Applied Psychology, 1975, 60, 14-19.

Kubany, A. J. Use of sociometric peer nominations in medical education research. Journal of Applied Psychology, 1957, 41, 389-394.

Labovitz, S. Some observations on measurement and statistics. Social Forces, 1967, 46, 151-160.

Labovitz, S. The assignment of numbers of rank order categories. American Sociological Review, 1970, 35, 515-524.

Landy, F. J., Farr, J. L., Saal, F. E., & Freytag, W. R. Behaviorally anchored scales for rating the performance of police officers. Journal of Applied Psychology, 1976, 61, 750-758.

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

Lindzey, G., & Byrne, D. Measurement of social choice and interpersonal attractiveness. In G. Lindzey and E. Aronsen (Eds.), Handbook of social psychology (Vol. 2). Reading, Mass.: Addison-Wesley, 1969.

Lord, F. M., & Novick, M. R. Statistical theories of mental test scores. Reading, Mass.: Addison-Wesley, 1968.

Love, K. G. Validation of selection procedures for police classes: Officer, Sergeant, Lieutenant, and Captain Phase I: Job analysis. Tampa, Florida: City of Tampa, Civil Service Board, Test Validation Division, 1978.

Mayfield, E. C. Management selection: Buddy nominations revisited. Personnel Psychology, 1970, 23, 377-391.

Mayfield, E. C. Value of peer nominations in predicting life insurance sales performance. Journal of Applied Psychology, 1972, 56, 319-323.

Mayo, G. D. Peer ratings and halo. Educational and Psychological Measurement, 1956, 16, 317-323.

Miner, J. B. The early identification of managerial talent. The Personnel and Guidance Journal, 1968, 46, 586-591.

Murstein, B. I. Physical attractiveness and marital choice. Journal of Personality and Social Psychology, 1972, 22, 8-12.

Myers, A. Team competition, success, and the adjustment of group members. Journal of Abnormal and Social Psychology, 1962, 65, 325-332.

Newcomb, T. M. The acquaintance process. New York: Holt, Rinehart, & Winston, 1961.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., & Bent, D. H. Statistical package for the social sciences. New York: McGraw-Hill, 1975.

Ricciuti, H. N. Ratings of leadership potential at the United States Naval Academy and subsequent officer performance. Journal of Applied Psychology, 1955, 39, 194-199.

Roadman, H. E. An industrial use of peer ratings. Journal of Applied Psychology, 1964, 48, 211-214.

Schmidt, F. L., & Johnson, R. H. Effect of race on peer ratings in an industrial situation. Journal of Applied Psychology, 1973, 57, 237-241.

Sigall, H., & Landy, D. Radiating beauty: Effects of having a physically attractive partner on person perception. Journal of Personality and Social Psychology, 1973, 28, 218-225.

Smith, G. M. Usefulness of peer ratings of personality in educational research. Educational and Psychological Measurement, 1967, 27, 967-984.

Smith, P. C., & Kendall, L. M. Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. Journal of Applied Psychology, 1963, 47, 149-155.

Spector, P. E. What to do with significant multivariate effects in MANOVA. Journal of Applied Psychology, 1977, 62, 158-163.

Springer, D. Ratings of candidates for promotion by co-workers and supervisors. Journal of Applied Psychology, 1953, 37, 347-351.

Stahl, M. J., & Steger, J. A. Measuring innovation and productivity--A peer rating approach. Research Management, 1977, 20, 35-38.

Suci, G. J., Vallance, T. R., & Glickman, A. S. A study of the effects of "likingness" and level of objectivity on peer rating reliabilities. Educational and Psychological Measurement, 1955, 16, 147-152.

Swanson, R. G., & Johnson, D. A. Relation between peer perception of leader behavior and instructor-pilot performance. Journal of Applied Psychology, 1975, 60, 198-200.

Theodorson, G. A. The relationship between leadership and popularity roles in small groups. American Sociological Review, 1957, 18, 58-67.

Trites, E. K. Adaptability measures as predictors of performance ratings. Journal of Applied Psychology, 1960, 44, 349-353.

Tucker, M. F., Cline, V. B., & Schmidt, J. R. Prediction of creativity and other performance measures from biographical information among pharmaceutical scientists. Journal of Applied Psychology, 1967, 51, 131-138.

Turk, H. Instrumental and expressive ratings reconsidered. Sociometry, 1961, 24, 76-81.

Waters, L. K., & Waters, C. W. Peer nominations as predictors of short-term sales performance. Journal of Applied Psychology, 1970, 54, 42-44.

Webb, W. B. The problem of obtaining negative nominations in peer ratings. Personnel Psychology, 1955, 8, 61-64.

Weitz, J. Selecting supervisors with peer nominations. Personnel Psychology, 1958, 11, 25-36.

Wherry, R. J., & Fryer, D. H. Buddy ratings: Popularity contest or leadership criteria? Personnel Psychology, 1949, 2, 147-159.

Wiggins, N., Blackburn, M., & Hackman, J. R. Prediction of first-year graduate success in psychology: Peer ratings. Journal of Educational Research, 1969, 63, 81-85.

Williams, S. B., & Leavitt, H. J. Group opinion as a predictor of military leadership. Journal of Consulting Psychology, 1947, 11, 283-291.

Willingham, W. W. Estimating the internal consistency of mutual peer nominations. Psychological Reports, 1959, 5, 163-167.

Winch, R. F., & Anderson, R. B. W. Two problems involved in the use of peer rating scales and some observations on Kendall's coefficient of concordance. Sociometry, 1967, 30, 316-322.

Wodder, N. C., & Hall, W. E. An analysis of peer ratings. Personnel Guidance Journal, 1962, 40, 606-609.

Zander, A., & Havelin, A. Social comparison and interpersonal attraction. Human Relations, 1960, 13, 21-32.

APPENDICES

APPENDIX 1

Peer Assessment Instrument

INSTRUCTIONS

****A**** PART A -- NOMINATION OF OFFICERS/CORPORALS IN DISTRICTS I & II ****A****

IN PART A ON THE FOLLOWING PAGE, FOR THE FIRST AREA OF PERFORMANCE, READ WHAT JOB KNOWLEDGE IS, BOTH THE DEFINITION AND ALL THE ELEMENTS (KNOWLEDGES, SKILLS, ABILITIES, AND PERSONAL CHARACTERISTICS).

THEN, LIST IN DECREASING ORDER THE ID NUMBERS (FOUND IN PART F) OF THREE (3) FELLOW OFFICERS OR CORPORALS PRESENTLY WORKING WITHIN EITHER DISTRICT I OR II WHOM YOU THINK DISPL THE BEST AMOUNT OF JOB KNOWLEDGE ON THE JOB. (IF SOMEONE HAS NO ASSIGNED ID NUMBER, USE THEIR NAME, DEPARTMENTAL RANK, AND SQUAD NUMBER)

IN OTHER WORDS, PLACE THE ID NUMBER OF THE OFFICER OR CORPORAL YOU THINK DISPLAYS THE MOST JOB KNOWLEDGE BY THE NUMBER ONE (1), THEN PLACE THE ID NUMBER OF THE OFFICER OR CORPOR YOU THINK DISPLAYS THE SECOND MOST JOB KNOWLEDGE BY THE NUMBER TWO (2), AND THEN PLACE THE ID NUMBER OF THE OFFICER OR CORPORAL YOU THINK DISPLAYS THE THIRD MOST JOB KNOWLEDGE BY THE NUMBER THREE (3).

REMEMBER, WHEN NOMINATING FELLOW OFFICERS OR CORPORALS, YOU SHOULD THINK OF THEIR PERFORMANC ON ONLY THE SINGLE AREA OF PERFORMANCE YOU ARE CURRENTLY WORKING ON. DO NOT NOMINATE PEOPLE BASED ON THEIR OVERALL PERFORMANCE AS POLICE OFFICERS.

DO NOT NOMINATE YOURSELF AND LEAVE NO SPACES BLANK.

AFTER YOU HAVE LISTED THE ID NUMBERS OF THREE OFFICERS OR CORPORALS FOR JOB KNOWLEDGE, GO ON TO PART B.

****B**** PART B -- RANKING OF FELLOW SQUAD MEMBERS ****B****

NOW THAT YOU ARE FAMILIAR WITH THE PERFORMANCE AREA OF JOB KNOWLEDGE, THINK OF ALL THE OFFICERS AND CORPORALS IN YOUR SQUAD. BY COMPARING THEM WITH ONE ANOTHER, RANK THEM FROM FIRST (1ST) TO LAST IN PART B. EXCLUDE YOURSELF FROM THE RANKING. USE ONLY THE ID NUMBER OF THE PERSON YOU ARE RANKING (FOUND IN PART F). (IF SOMEONE HAS NO ASSIGNED ID NUMBER, USE THEIR NAME, DEPARTMENTAL RANK, AND SQUAD NUMBER)

REMEMBER, RANK YOUR SQUAD MEMBERS ON HOW WELL THEY DISPLAY ALL THE ELEMENTS OF JOB KNOWLEDGE. THIS RANKING WILL ONLY APPLY TO JOB KNOWLEDGE, NOT TO ALL THE AREAS OF PERFORMANCE. THERE CAN BE NO TIES FOR ANY RANK.

****C**** PART C -- RATING OF FELLOW SQUAD MEMBERS ****C****

AFTER ALL SQUAD MEMBERS HAVE BEEN RANKED ON JOB KNOWLEDGE, YOU ARE READY TO RATE THESE SAME PEOPLE.

READ THE BEHAVIORAL EXAMPLES WHICH APPEAR BELOW THE RATING SCALE ON THE PAGE. THESE WILL GIV YOU AN IDEA OF THE KINDS OF JOB KNOWLEDGE BEHAVIOR WHICH ARE CONSIDERED "HIGH," "AVERAGE," AN "LOW."

NOW, TAKE THE OFFICER OR CORPORAL YOU RANKED AS FIRST (1ST) ON JOB KNOWLEDGE. RATE THAT PERSON AS TO HOW WELL HE OR SHE DISPLAYS ALL ELEMENTS OF JOB KNOWLEDGE ON THE JOB. TO RATE A PERSON USE THE 9-POINT SCALE WHICH APPEARS IN PART C. USE ANY WHOLE OR HALF NUMBER ON THE SCALE FROM 1 TO 9 (SUCH AS, 1, 1.5, 2, 2.5, ETC.).

REPEAT THE RATING PROCEDURE FOR EACH SQUAD MEMBER YOU RANKED IN PART B.

INSTRUCTIONS CONTINUED ON NEXT PAGE

INSTRUCTIONS (CONT.)

AFTER COMPLETING PARTS A, B, AND C FOR JOB KNOWLEDGE, REPEAT THIS ENTIRE PROCEDURE (PARTS A, B, AND C) FOR THE REMAINING EIGHT AREAS OF PERFORMANCE

AFTER YOU HAVE FINISHED ALL PARTS OF THE NINE PERFORMANCE AREAS, JOB KNOWLEDGE THROUGH WORK ATTITUDE, GO ON TO PARTS D AND E.

JOB KNOWLEDGE -- USE OF KNOWLEDGE OF LAWS, PRACEDURES, POLICIES, AND TECHNIQUES RELATED TO THE LAW ENFORCEMENT FUNCTION (PATROL, ARREST, TESTIFYING, ETC.), INCLUDING THE APPLICATION OF PRIOR TRAINING.

This performance area is composed of these elements:

KNOWLEDGE OF:
- laws and procedures
- patrol procedures
- stakeout procedures
- investigation procedures
- field interrogation procedures
- police codes
- procedures for responding to calls
- procedures for identifying and preserving evidence
- court procedures
- log recording procedures
- exact booking procedures
- job jargon
- purpose of roll call
- arrest techniques
- reporting and documentation requirements
- reporting and documentation procedures

SKILL IN:
- observation techniques
- concealment and disguise
- checking residences and buildings
- handling pertinent criminal evidence
- preserving pertinent criminal evidence
- search and seizure techniques
- use of physical force
- various self-defense tactics
- making arrests using minimum force

ABILITY TO:
- use necessary force up t extreme of taking an life
- see acutely
- do extended observation
- physically intervene an maintain/restore ord

****A**** NOMINATION OF OFFICERS/CORPORALS IN DISTRICTS I & II: (1)_____ (2)_____ (3)_____ (use ID NO.)

FOR FELLOW SQUAD MEMBERS ONLY

****B**** **RANKING** ****B****

FIRST	SECOND	THIRD	FOURTH	FIFTH	SIXTH	SEVENTH	EIGHTH	NINTH
_____	_____	_____	_____	_____	_____	_____	_____	_____

(use assigned ID NO. of squad members in spaces above)

****C**** **RATING** ****C****

_____	_____	_____	_____	_____	_____	_____	_____	_____

(use rating of 1 - 9 in these spaces)
(rating should be made on the squad member directly above each rating space)

1 2 3 4 5 6 7 8 9
HIGH AVERAGE LOW

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "HIGH" ON JOB KNOWLEDGE

Lets supervisor know what happens in court so that procedures may be altered and the whole department may profit from his/her experience.

Conducts an external search (frisk) under suspicious circumstances with knowledge of laws permitting such action.

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "AVERAGE" ON JOB KNOWLEDGE

Applies the precise law to a case, avoiding ambiguous or wrong charges.

Reacts appropriately to the department's normal presenta of changes in procedural or legal matters.

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "LOW" ON JOB KNOWLEDGE

Has to be frequently reminded of the elements of the job and personal conduct.

Fails to document location (by photo, identification mar etc.) or recognize evidence before moving it.

DECISION-MAKING -- ANALYTIC ASSESSMENT OF THE SITUATION AND TAKING NECESSARY AND APPROPRIATE ACTION AFTER CONSIDERATION OF ALTERNATIVE APPROACHES.

This performance area is composed of these elements:

SKILL IN:	ABILITY TO:	PERSONAL CHARACTERISTICS:
discriminating between messages that are significant and insignificant	find and follow directions	creativity
analyzing a situation, circumstance, or incident	pursue a logical line of inquiry	risk-taking
identifying criminal evidence	organize thoughts and materials	decisiveness
reconstructing traffic accidents	accurately assess situations	courage
detecting the activities and intent of individuals	determine probable cause	
	solve problems	
	reach logical conclusions	

****A**** NOMINATION OF OFFICERS/CORPORALS IN DISTRICTS I & II: (1)_____ (2)_____ (3)_____
(use ID NO.)

FOR FELLOW SQUAD MEMBERS ONLY

****B**** **RANKING** ****B****

FIRST	SECOND	THIRD	FOURTH	FIFTH	SIXTH	SEVENTH	EIGHTH	NINTH
_____	_____	_____	_____	_____	_____	_____	_____	_____

(use assigned ID NO. of squad members in spaces above)

****C**** **RATING** ****C****

_____ _____ _____ _____ _____ _____ _____ _____ _____
(use rating of 1 - 9 in these spaces)
(rating should be made on the squad member directly above each rating space)

1 2 3 4 5 6 7 8 9
HIGH AVERAGE LOW

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "HIGH" ON DECISION-MAKING

Makes correct/proper decisions of which he is highly confident, even in difficult situations.

Upon arrival at the scene of a large fire, sees the need for additional help and calls for it to be dispatched to specific locations.

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "AVERAGE" ON DECISION-MAKING

Waits for assistance when it is appropriate in order to handle a physical confrontation.

Occasionally has to confer with other officers concern the proper action to take in a given situation.

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "LOW" ON DECISION-MAKING

Continues in high speed pursuit without considering the consequences, even though they may be severe.

Approaches a car stopped for a traffic violation witho appropriate safety precautions.

DEALING WITH CO-WORKERS -- ABILITY TO WORK WITH FELLOW OFFICERS, BOTH SWORN AND NONUNIFORMED PERSONNEL, INCLUDING ACCEPTING AND GIVING CONSTRUCTIVE CRITICISM, MUTUAL DECISION-MAKING, AND TAKING AN EQUAL SHARE OF THE WORK LOAD.

This performance area is composed of these elements:

ABILITY TO:
- work cooperatively with other personnel
- interact effectively with all types of people

PERSONAL CHARACTERISTICS:
- flexibility
- sensitivity to organizational goals (political sensitivity)
- respect for authority
- loyalty
- emotional control
- impartiality

****A**** NOMINATION OF OFFICERS/CORPORALS IN DISTRICTS I & II: (1)_____ (2)_____ (3)_____
(use ID NO.)

FOR FELLOW SQUAD MEMBERS ONLY

****B**** **RANKING** ****B****

FIRST	SECOND	THIRD	FOURTH	FIFTH	SIXTH	SEVENTH	EIGHTH	NINTH
_____	_____	_____	_____	_____	_____	_____	_____	_____

(use assigned ID NO. of squad members in spaces above)

****C**** **RATING** ****C****

_____ _____ _____ _____ _____ _____ _____ _____ _____

(use rating of 1 - 9 in these spaces)
(rating should be made on the squad member directly above each rating space)

1 2 3 4 5 6 7 8 9
HIGH AVERAGE LOW

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "HIGH" ON DEALING WITH CO-WORKERS

Volunteers to assist a fellow officer who has a very heavy work load.

Works willingly in assisting an officer who is having trouble adjusting to various duties.

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "AVERAGE" ON DEALING WITH CO-WORKERS

Handles his or her own share of assigned duties within his or her squad.

Gets along well with his or her co-workers and avoids annoying habits.

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "LOW" ON DEALING WITH CO-WORKERS

Thwarts squad productivity by failing to share information relevant to police hazards and suspects.

Antagonizes his supervisors and fellow officers by starting rumors about them.

USE OF EQUIPMENT -- SKILL IN THE USE OF FIREARMS, OTHER WEAPONS, HANDCUFFS, RADIO, FIRST AID, VEHICLES, AND OTHER SPECIALIZED EQUIPMENT.

This performance area is composed of these elements:

SKILL IN:

- driving a patrol car
- the use of radios, binoculars, and other specialized equipment
- use of equipment to collect evidence
- the use of weapons
- the use of traffic control equipment

****A**** NOMINATION OF OFFICERS/COMPANIES IN DISTRICTS I & II: (1)_____ (2)_____ (3)_____
(use ID NO.)

FOR FELLOW SQUAD MEMBERS ONLY

****B**** **RANKING** ****B****

FIRST	SECOND	THIRD	FOURTH	FIFTH	SIXTH	SEVENTH	EIGHTH	NINTH
_____	_____	_____	_____	_____	_____	_____	_____	_____

(use assigned ID NO. of squad members in spaces above)

****C**** **RATING** ****C****

_____	_____	_____	_____	_____	_____	_____	_____	_____

(use rating of 1 - 9 in these spaces)
(rating should be made on the squad member directly above each rating space)

1 2 3 4 5 6 7 8 9
HIGH AVERAGE LOW

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "HIGH" ON USE OF EQUIPMENT

Uses nightstick as a "come-along" in such a manner so as to minimize injury to the suspect.

Takes appropriate breaks in the use of the radio when broadcasting lengthy messages/information.

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "AVERAGE" ON USE OF EQUIPMENT

Uses vehicle to appropriately protect incident scene.

Uses proper handcuffing procedure/technique so as to secure suspect, minimize discomfort, etc.

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "LOW" ON USE OF EQUIPMENT

Damages his/her patrol car by driving it in areas which are not suited for vehicular traffic.

Cannot meet departmental standards ("qualify") with weapon.

DEALING WITH THE PUBLIC -- KNOWLEDGE AND SKILL IN USING TECHNIQUES GEARED AT DEALING WITH THE PUBLIC IN A RESPECTFUL, TACTFUL STYLE.

This performance area is composed of these elements:

KNOWLEDGE OF:	SKILL IN:	PERSONAL CHARACTERISTICS:
the basic tenets of human behavior	interviewing and questioning people	sensitivity
impact of self on others	counseling individuals or groups	open-mindedness
counseling techniques		
geographic area		
community social service resources		

****A**** NOMINATION OF OFFICERS/CORPORALS IN DISTRICTS I & II: (1)_____ (2)_____ (3)_____
(use ID NO.)

FOR FELLOW SQUAD MEMBERS ONLY

****B**** **RANKING** ****B****

FIRST	SECOND	THIRD	FOURTH	FIFTH	SIXTH	SEVENTH	EIGHTH	NINTH
_____	_____	_____	_____	_____	_____	_____	_____	_____

(use assigned ID NO. of squad members in spaces above)

****C**** **RATING** ****C****

_____	_____	_____	_____	_____	_____	_____	_____	_____

(use rating of 1 - 9 in these spaces)
(rating should be made on the squad member directly above each rating space)

1 2 3 4 5 6 7 8 9
HIGH AVERAGE LOW

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "HIGH" ON DEALING WITH THE PUBLIC

Makes him/herself knowledgeable of the type and location of businesses in assigned area.

Does not let improper reaction of the public dictate his/her behavior.

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "AVERAGE" ON DEALING WITH THE PUBLIC

Is patient with a citizen he or she has contact with.

Tells a traffic violator specifically why he or she has been stopped to avoid an argumentative situation

EXAMPLES OF BEHAVIORS OF OFFICE$S WHO ARE USUALLY RATED "LOW" ON DEALING WITH THE PUBLIC

Creates a bad impression with citizens he or she comes in contact with by being abrupt and insensitive.

Unintentionally creates an uncomfortable situation for a victim of crime by making unnecessary comments or telling "war stories" within earshot.

COMMUNICATION -- ABILITY TO MAKE ONESELF UNDERSTOOD AND UNDERSTAND OTHERS IN FACE TO FACE SITUATIONS AND TO TRANSMIT AND RECEIVE INFORMATION IN BOTH ORAL AND WRITTEN FORM.

This performance area is composed of these elements:

SKILL IN:
- completing reports accurately

ABILITY TO:
- follow simple and complex oral instructions
- communicate in writing
- communicate verbally (orally)
- follow simple and complex written instructions
- express moderately complex ideas
- comprehend and remember writ materials
- comprehend questions
- speak in public
- explain procedures verbally (orally)
- explain charges
- recall and record work activ

****A**** NOMINATION OF OFFICERS/CORPORALS IN DISTRICTS I & II: (1)_____ (2)_____ (3)_____
(use ID NO.)

FOR FELLOW SQUAD MEMBERS ONLY

****B**** **RANKING** ****B****

FIRST	SECOND	THIRD	FOURTH	FIFTH	SIXTH	SEVENTH	EIGHTH	NINTH
_____	_____	_____	_____	_____	_____	_____	_____	_____

(use assigned ID NO. of squad members in spaces above)

****C**** **RATING** ****C****

_____ _____ _____ _____ _____ _____ _____ _____ _____

(use rating of 1 - 9 in these spaces)
(rating should be made on the squad member directly above each rating space)

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "HIGH" ON COMMUNICATION

- Speaks clearly and concisely with well-thought-out ideas (when testifying in court, speaking to fellow officers, etc.)
- Has ability to communicate effectively at the appropriat educational level.

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "AVERAGE" ON COMMUNICATION

- Uses acceptable sentence structure and grammar in written reports, following a logical order of presentation.
- Recognizes and notes informative items necessary for proper job performance.

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "LOW" ON COMMUNICATION

- Includes incorrect grammar and incomplete sentences in his or her written reports, making them extremely difficult to understand.
- Creates confusion by lack of effort and ability to communicate clearly and concisely (in explaining charges, procedures, etc.).

RELIABILITY -- DEPENDABILITY IN JOB ATTENDANCE, EFFORT EXPENDITURE, ACCEPTANCE OF RESPONSIBILITY, FUNCTIONING APPROPRIATELY UNDER STRESS, AND ACCURACY IN ALL DETAILS OF WORK.

This performance area is composed of these elements:

ABILITY TO:
- control own emotions
- maintain composure in the face of antagonistic questions

PERSONAL CHARACTERISTICS:
- stress tolerance
- resilience
- perseverence
- dependability

****A**** NOMINATION OF OFFICERS/CORPORALS IN DISTRICTS I & II: (1)_____ (2)_____ (3)_____
(use ID NO.)

FOR FELLOW SQUAD MEMBERS ONLY

****B**** **RANKING** ****B****

FIRST	SECOND	THIRD	FOURTH	FIFTH	SIXTH	SEVENTH	EIGHTH	NINTH
_____	_____	_____	_____	_____	_____	_____	_____	_____

(use assigned ID NO. of squad members in spaces above)

****C**** **RATING** ****C****

_____	_____	_____	_____	_____	_____	_____	_____	_____

(use rating of 1 - 9 in these spaces)
(rating should be made on the squad member directly above each rating space)

1 2 3 4 5 6 7 8 9
HIGH AVERAGE LOW

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "HIGH" ON RELIABILITY

Remains in assigned area, treating public he or she comes in contact with in a courteous fashion while dealing with the hazards of police work.

Recognizes police hazards in his or her assigned area and takes repetitive action until it is resolved.

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "AVERAGE" ON RELIABILITY

Immediately applies minimum amount of physical assistance/force in order to subdue a fighting suspect.

Reliably responds to all assigned calls without compl or hesitation.

EXAMPLES OB BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "LOW" ON RELIABILITY

Permits his or her effectiveness to deteriorate as a result of losing control of his or her emotions.

Stays "out of service" to avoid getting a late call.

DEMEANOR -- PERSONAL AND PROFESSIONAL PRIDE AS SHOWN BY HIS OR HER STANDARDS OF BEHAVIOR AND PHYSICAL APPEARANCE.

This performance area is composed of these elements:

PERSONAL CHARACTERISTICS:
- impact
- personal appearance
- integrity
- self-confidence

****A**** NOMINATION OF OFFICERS/CORPORALS IN DISTRICTS I & II: (1)_____ (2)_____ (3)_____
(use ID NO.)

FOR FELLOW SQUAD MEMBERS ONLY

****B**** **RANKING** ****B****

FIRST	SECOND	THIRD	FOURTH	FIFTH	SIXTH	SEVENTH	EIGHTH	NINTH
_____	_____	_____	_____	_____	_____	_____	_____	_____

(use assigned ID NO. of squad members in spaces above)

****C**** **RATING** ****C****

_____ _____ _____ _____ _____ _____ _____ _____ _____

(use rating of 1 - 9 in these spaces)
(rating should be made on the squad member directly above each rating space)

1 2 3 4 5 6 7 8 9
HIGH AVERAGE LOW

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "HIGH" ON DEMEANOR

Insures that his or her patrol car, equipment and personal appearance meet high standards without reminders.

Isn't affected by peer or public pressure to compromise his or her integrity.

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "AVERAGE" ON DEMEANOR

Takes appropriate action on police calls for service in absence of direct supervision.

Requires occasional reminders to maintain proper standards of appearance and conduct.

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "LOW" ON DEMEANOR

Is loud and boisterous, and on frequent occasions is disruptive due to inappropriate comments.

Uses his or her position for personal gain.

WORK ATTITUDE -- INTERESTED IN SERVING THE PUBLIC THROUGH THE FAIR AND OBJECTIVE ENFORCEMENT AND ADMINISTRATION OF THE LAW, GAINING SATISFACTION FROM DOING HIS OR HER JOB WELL.

This performance area is composed of these elements:

ABILITY TO:
- be objective
- exercise discretion

PERSONAL CHARACTERISTICS:
- motivation for work
- energy
- career ambition
- desire for self-improvement
- independence
- initiative

****A**** NOMINATION OF OFFICERS/CORPORALS IN DISTRICTS I & II: (1)_____ (2)_____ (3)_____
(use ID NO.)

FOR FELLOW SQUAD MEMBERS ONLY

****B**** **RANKING** ****B****

FIRST	SECOND	THIRD	FOURTH	FIFTH	SIXTH	SEVENTH	EIGHTH	NINTH
_____	_____	_____	_____	_____	_____	_____	_____	_____

(use assigned ID NO. of squad members in spaces above)

****C**** **RATING** ****C****

_____ _____ _____ _____ _____ _____ _____ _____ _____

(use rating of 1 - 9 in these spaces)
(rating should be made on the squad member directly above each rating space)

1 2 3 4 5 6 7 8 9
HIGH AVERAGE LOW

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "HIGH" ON WORK ATTITUDE

Seeks additional education and training from sources available to him/her (educational institutions, within the department, etc.) in order to improve his or her efficiency and effectiveness.

Stays abreast of crime trends in his or her area of responsibility and initiates corrective/preventive actions.

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "AVERAGE" ON WORK ATTITUDE

Makes an effort to perform his or her duties within departmental rules and regulations in the best interest of the public.

Generally establishes priority of activities during his or her normal tour of duty.

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "LOW" ON WORK ATTITUDE

Lets outside interests and activities interfere with his or her production on the job.

Only goes through the motions of the job and blames others' success on luck of politics.

****PART D****

HOW WELL YOU KNOW THOSE NOMINATED AND/OR YOUR FELLOW SQUAD MEMBERS

PLACE I.D. NO. OF PERSON IN LEFT-HAND COLUMN. ANSWER EACH OF THE FIVE QUESTIONS BY PLACING THE NUMBER WHICH REPRESENTS YOUR CHOICE OF ANSWERS IN THE BOX UNDER THE QUESTION AND ACROSS FROM THE PERSON'S I.D. NO.

ANSWER ALL FIVE QUESTIONS FOR EACH PERSON YOU HAVE NOMINATED OR RANKED/RATED.

IDENTIFICATION NUMBER	How much contact do you have with this person OFF THE JOB? 1-none 2-very little; 3-some contact; 4-quite a bit; 5-a great deal.	How much contact do you have with this person ON THE JOB? 1-none 2-very little; 3-some contact; 4-quite a bit; 5-a great deal.	How well do you KNOW this person? 1-not at all; 2-not very well; 3-somewhat; 4-fairly well; 5-extremely well.	How well do you LIKE this person? 1-strongly dislike; 2-dislike; 3-neither like nor dislike; 4-like; 5-strongly like.	Is this person a FRIEND of yours? 1-could never be my friend; 2-not a friend; 3-merely an acquaintance 4-is a friend; 5-is one of my best friends.
_____					
_____					
_____					
_____					
_____					
_____					
_____					
_____					
_____					
_____					
_____					
_____					
_____					
_____		12			

(CONTINUED ON NEXT PAGE)

HOW WELL YOU KNOW THOSE NOMINATED AND/OR YOUR FELLOW SQUAD MEMBERS

PLACE I.D. NO. OF PERSON IN LEFT-HAND COLUMN. ANSWER EACH OF THE FIVE QUESTIONS BY PLACING THE NUMBER WHICH REPRESENTS YOUR CHOICE OF ANSWERS IN THE BOX UNDER THE QUESTION AND ACROSS FROM THE PERSON'S I.D. NO.

ANSWER ALL FIVE QUESTIONS FOR EACH PERSON YOU HAVE NOMINATED OR RANKED/RATED.

IDENTIFICATION NUMBER	How much contact do you have with this person OFF THE JOB? 1-none 2-very little; contact; 3-some at all; 4-quite a bit; 5-a great deal.	How much contact do you have with this person ON THE JOB? 1-none 2-very little; contact; 3-some at all; 4-quite a bit; 5-a great deal.	How well do you KNOW this person? 1-not at all; 2-not very well; 3-somewhat; 4-fairly well; 5-extremely well.	How well do you LIKE this person? 1-strongly dislike; 2-dislike; 3-neither like nor dislike; 4-like; 5-strongly like.	Is this person a FRIEND of yours? 1-could never be my friend; 2-not a friend; 3-merely an acquaintance 4-is a friend; 5-is one of my best friends.
_____					
_____					
_____					
_____					
_____					
_____					
_____					
_____					
_____					
_____					
_____					
_____					
_____					

(GO TO PART II) 15

****PART E****

OPINION SECTION

ANSWER THE FOLLOWING QUESTIONS BY PLACING AN "X" IN THE APPROPRIATE BOX TO THE RIGHT OF THE QUESTION

	STRONGLY DISAGREE	DISAGREE	NEITHER DISAGREE OR AGREE	AGREE	AGREE
OPINIONS REGARDING THE NOMINATION PROCEDURE		(place "X" in box)			
(1) This system (nominating fellow Officers or Corporals) is a fair way to rate law enforcement personnel.					
(2) I like this way of rating people.					
(3) With this system people will nominate only their closest friends.					
(4) This system will generate too much competition between Officers and Corporals who work together.					
(5) With this system, most people will nominate poor performers in order to lessen the competition when they are compared to the better performers.					
(6) This system will provide an accurate indication of a person's ability to perform law enforcement work.					
(7) This system should be used as one way of deciding who should be promoted.					
OPINIONS REGARDING THE RANKING PROCEDURE					
(1) This system (ranking fellow Officers or Corporals) is a fair way to rate law enforcement personnel.					
(2) I like this way of rating people.					
(3) With this system people will rank highly only their closest friends.					
(4) This system will generate too much competition between Officers and Corporals who work together.					
(5) With this system, most people will rank highly poor performers in order to lessen the competition when they are compared to the better performers.					
(6) This system will provide an accurate indication of a person's ability to perform law enforcement work.					
(7) This system should be used as one way of deciding who should be promoted.					
OPINIONS REGARDING THE RATING PROCEDURE					
(1) This system (rating fellow Officers or Corporals) is a fair way to rate law enforcement personnel.					
(2) I like this way of rating people.					
(3) With this system people will rate highly only their closest friends.					
(4) This system will generate too much competition between Officers and Corporals who work together.					
(5) With this system, most people will rate highly poor performers in order to lessen the competition when they are compared to the better performers.					
(6) This system will provide an accurate indication of a person's ability to perform law enforcement work.					
(7) This system should be used as one way of deciding who should be promoted.					

IF YOU WISH TO MAKE ADDITIONAL COMMENTS, PLEASE USE THE BACK OF THIS BOOKLET.

AFTER YOU HAVE COMPLETED PART E, PLACE ALL MATERIALS INTO THE ENVELOPE PROVIDED, GIVE IT TO YOUR SQUAD SUPERVISOR. HE WILL SEND IT DIRECTLY TO KEVIN LOVE, TEST VALIDATION DIVISION, CIVIL SERVICE, #50, THROUGH THE INTER-DEPARTMENTAL MAIL. PLEASE RETURN ALL MATERIALS.

APPENDIX 2

Supervisor Assessment Instrument

INSTRUCTIONS

****A**** PART A -- DESCRIPTION OF AREA OF PERFORMANCE ****A****

IN PART A ON THE FOLLOWING PAGE, FOR THE FIRST AREA OF PERFORMANCE, READ WHAT JOB KNOWLEDGE IS, BOTH THE DEFINITION AND ALL THE ELEMENTS (KNOWLEDGES, SKILLS, ABILITIES, AND PERSONAL CHARACTERISTICS).

****B**** PART B -- RANKING OF SQUAD MEMBERS ****B****

NOW THAT YOU ARE FAMILIAR WITH THE PERFORMANCE AREA OF JOB KNOWLEDGE, THINK OF ALL THE OFFICERS AND CORPORALS IN YOUR SQUAD. BY COMPARING THEM WITH ONE ANOTHER, RANK THEM FROM FIRST (1ST) TO LAST IN PART B. EXCLUDE YOURSELF FROM THE RANKING. USE ONLY THE ID NUMBER OF THE PERSON YOU ARE RANKING (FOUND IN PART F). (IF SOMEONE HAS NO ASSIGNED ID NUMBER, USE THEIR RACE, DEPENDENTAL RANK, AND SQUAD NUMBER)

REMEMBER, RANK YOUR SQUAD MEMBERS ON HOW WELL THEY DISPLAY ALL THE ELEMENTS OF JOB KNOWLEDGE. THIS RANKING WILL ONLY APPLY TO JOB KNOWLEDGE, NOT TO ALL THE AREAS OF PERFORMANCE. THERE CAN BE NO TIES FOR ANY RANK.

****C**** PART C -- RATING OF SQUAD MEMBERS ****C****

AFTER ALL SQUAD MEMBERS HAVE BEEN RANKED ON JOB KNOWLEDGE, YOU ARE READY TO RATE THESE SAME PEOPLE.

READ THE BEHAVIORAL EXAMPLES WHICH APPEAR BELOW THE RATING SCALE ON THE PAGE. THESE WILL GIVE YOU AN IDEA OF THE KINDS OF JOB KNOWLEDGE BEHAVIOR WHICH ARE CONSIDERED "HIGH," "AVERAGE," AND "LOW."

NOW, TAKE THE OFFICER OR CORPORAL YOU RANKED AS FIRST (1ST) ON JOB KNOWLEDGE. RATE THAT PERSON AS TO HOW WELL HE OR SHE DISPLAYS ALL ELEMENTS OF JOB KNOWLEDGE ON THE JOB. TO RATE A PERSON USE THE 9-POINT SCALE WHICH APPEARS IN PART C. USE ANY WHOLE OR HALF NUMBER ON THE SCALE FROM 1 TO 9 (SUCH AS, 1, 1.5, 2, 2.5, ETC.).

REPEAT THE RATING PROCEDURE FOR EACH SQUAD MEMBER YOU RANKED IN PART B.

AFTER COMPLETING PARTS A, B, AND C FOR JOB KNOWLEDGE, REPEAT THIS ENTIRE PROCEDURE (PARTS A, B, AND C) FOR THE REMAINING EIGHT AREAS OF PERFORMANCE.

JOB KNOWLEDGE -- USE OF KNOWLEDGE OF LAWS, PROCEDURES, POLICIES, AND TECHNIQUES RELATED TO THE LAW ENFORCEMENT FUNCTION (PATROL, ARREST, TESTIFYING, ETC.), INCLUDING THE APPLICATION OF PRIOR TRAINING.

This performance area is composed of these elements:

KNOWLEDGE OF:
- laws and procedures
- patrol procedures
- stakeout procedures
- investigation procedures
- field interrogation procedures
- police codes
- procedures for responding to calls
- procedures for identifying and preserving evidence
- court procedures
- log recording procedures
- exact booking procedures
- job jargon
- purpose of roll call
- arrest techniques
- reporting and documentation requirements
- reporting and documentation procedures

SKILL IN:
- observation techniques
- concealment and disguise
- checking residences and buildings
- handling pertinent criminal evidence
- preserving pertinent criminal evidence
- search and seizure techniques
- use of physical force
- various self-defense tactics
- making arrests using minimum force

ABILITY TO:
- use necessary force up to extreme of taking ano life
- see acutely
- do extended observation
- physically intervene an maintain/restore orde

****B****

****B**** **RANKING** ****B****

FIRST	SECOND	THIRD	FOURTH	FIFTH	SIXTH	SEVENTH	EIGHTH	NINTH
______	______	______	______	______	______	______	______	______

(use assigned ID NO. of squad members in spaces above)

****C**** **RATING** ****C****

______ ______ ______ ______ ______ ______ ______ ______ ______

(use rating of 1 - 9 in these spaces)
(rating should be made on the squad member directly above each rating space)

1 — 2 — 3 — 4 — 5 — 6 — 7 — 8 — 9
HIGH — AVERAGE — LOW

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "HIGH" ON JOB KNOWLEDGE

Lets supervisor know what happens in court so that procedures may be altered and the whole department may profit from his/her experience.

Conducts an external search (frisk) under suspicious circumstances with knowledge of laws permitting such action.

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "AVERAGE" ON JOB KNOWLEDGE

Applies the precise law to a case, avoiding ambiguous or wrong charges.

Reacts appropriately to the department's normal process of changes in procedural or legal matters.

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "LOW" ON JOB KNOWLEDGE

Has to be frequently reminded of the elements of the job and personal conduct.

Fails to document location (by photo, identification mar etc.) or reconnoitre evidence before moving it.

DECISION-MAKING -- ANALYTIC ASSESSMENT OF THE SITUATION AND TAKING NECESSARY AND APPROPRIATE ACTION AFTER CONSIDERATION OF ALTERNATIVE APPROACHES.

This performance area is composed of these elements:

SKILL IN:
- discriminating between messages that are significant and insignificant
- analyzing a situation, circumstance, or incident
- identifying criminal evidence
- reconstructing traffic accidents
- detecting the activities and intent of individuals

ABILITY TO:
- find and follow directions
- pursue a logical line of inquiry
- organize thoughts and materials
- accurately assess situations
- determine probable cause
- solve problems
- reach logical conclusions

PERSONAL CHARACTERISTICS:
- creativity
- risk-taking
- decisiveness
- courage

****A****

****B**** **RANKING** ****

FIRST	SECOND	THIRD	FOURTH	FIFTH	SIXTH	SEVENTH	EIGHTH	NINTH
____	____	____	____	____	____	____	____	____

(use assigned ID NO. of squad members in spaces above)

****C**** **RATING** ****

____: ____ ____ ____ ____ ____ ____ ____ ____

(use rating of 1 - 9 in these spaces)
(rating should be made on the squad member directly above each rating space)

1 2 3 4 5 6 7 8 9
HIGH AVERAGE LOW

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "HIGH" ON DECISION-MAKING

Makes correct/proper decisions of which he is highly confident, even in difficult situations.

Upon arrival at the scene of a large fire, sees th need for additional help and calls for it to be dispatched to specific locations.

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "AVERAGE" ON DECISION-MAKING

Waits for assistance when it is appropriate in order to handle a physical confrontation.

Occasionally has to confer with other officers co the proper action to take in a given situation.

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "LOW" ON DECISION-MAKING

Continues in high speed pursuit without considering the consequences, even though they may be severe.

Approaches a car stopped for a traffic violation appropriate safety precautions.

DEALING WITH CO-WORKERS -- ABILITY TO WORK WITH FELLOW EMPLOYEES, BOTH SWORN AND NONUNIFORMED PERSONNEL, INCLUDING ACCEPTING AND GIVING CONSTRUCTIVE CRITICISM, MUTUAL DECISION-MAKING, AND TAKING AN EQUAL SHARE OF THE WORK LOAD.

This performance area is composed of these elements:

ABILITY TO:
- work cooperatively with other personnel
- interact effectively with all types of people

PERSONAL CHARACTERISTICS:
- flexibility
- sensitivity to organizational goals (political sensitivity)
- respect for authority
- loyalty
- emotional control
- impartiality

****A****

****B**** **RANKING** ****B****

FIRST	SECOND	THIRD	FOURTH	FIFTH	SIXTH	SEVENTH	EIGHTH	NINTH

(use assigned ID NO. of squad members in spaces above)

****C**** **RATING** ****C****

(use rating of 1 - 9 in these spaces)
(rating should be made on the squad member directly above each rating space)

1 2 3 4 5 6 7 8 9
HIGH AVERAGE LOW

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "HIGH" ON DEALING WITH CO-WORKERS

Volunteers to assist a fellow officer who has a very heavy work load.

Works willingly in assisting an officer who is having trouble adjusting to various duties.

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "AVERAGE" ON DEALING WITH CO-WORKERS

Handles his or her own share of assigned duties within his or her squad.

Gets along well with his or her co-workers and avoids annoying habits.

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "LOW" ON DEALING WITH CO-WORKERS

Thwarts squad productivity by failing to share information relevant to police hazards and suspects.

Antagonizes his supervisors and fellow officers by starting rumors about them.

USE OF EQUIPMENT -- SKILL IN THE USE OF FIREARMS, OTHER WEAPONS, HANDCUFFS, RADIO, FIRST AID, VEHICLES, AND OTHER SPECIALIZED EQUIPMENT.

This performance area is composed of these elements:

SKILL IN:

- driving a patrol car
- the use of radios, binoculars, and other specialized equipment
- use of equipment to collect evidence
- the use of weapons
- the use of traffic control equipment

****A****

****B**** **RANKING** ****B****

FIRST	SECOND	THIRD	FOURTH	FIFTH	SIXTH	SEVENTH	EIGHTH	NINTH
____	____	____	____	____	____	____	____	____

(use assigned ID NO. of squad members in spaces above)

****C**** **RATING** ****C****

____	____	____	____	____	____	____	____	____

(use rating of 1 - 9 in these spaces)
(rating should be made on the squad member directly above each rating space)

1 2 3 4 5 6 7 8 9
HIGH AVERAGE LOW

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "HIGH" ON USE OF EQUIPMENT

Uses nightstick as a "come-along" in such a manner so as to minimize injury to the suspect.

Takes appropriate breaks in the use of the radio wh broadcasting lengthy messages/information.

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "AVERAGE" ON USE OF EQUIPMENT

Uses vehicle to appropriately protect incident scene.

Uses proper handcuffing procedure/technique so as t secure suspect, minimize discomfort, etc.

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "LOW" ON USE OF EQUIPMENT

Damages his/her patrol car by driving it in areas which are not suited for vehicular traffic.

Cannot meet departmental standards ("qualify") wit weapon.

DEALING WITH THE PUBLIC -- KNOWLEDGE AND SKILL IN USING TECHNIQUES GEARED AT DEALING WITH THE PUBLIC IN A RESPECTFUL, TACTFUL STYLE.

This performance area is composed of these elements:

KNOWLEDGE OF:
- the basic tenets of human behavior
- impact of self on others
- counseling techniques
- geographic area
- community social service resources

SKILL IN:
- interviewing and questioning people
- counseling individuals or groups

PERSONAL CHARACTERISTICS:
- sensitivity
- open-mindedness

****B****

RANKING

****B****

FIRST	SECOND	THIRD	FOURTH	FIFTH	SIXTH	SEVENTH	EIGHTH	NINTH
____	____	____	____	____	____	____	____	____

(use assigned ID NO. of squad members in spaces above)

****C****

RATING

****C****

____	____	____	____	____	____	____	____	____

(use rating of 1 - 9 in these spaces)
(rating should be made on the squad member directly above each rating space)

1 2 3 4 5 6 7 8 9

HIGH AVERAGE LOW

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "HIGH" ON DEALING WITH THE PUBLIC

- Makes him/herself knowledgeable of the type and location of businesses in assigned area.
- Does not let improper reaction of the public dictate his/her behavior.

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "AVERAGE" ON DEALING WITH THE PUBLIC

- Is patient with a citizen he or she has contact with.
- Tells a traffic violator specifically why he or she has been stopped to avoid an argumentative situat

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "LOW" ON DEALING WITH THE PUBLIC

- Creates a bad impression with citizens he or she comes in contact with by being abrupt and insensitive.
- Unintentionally creates an uncomfortable situation f a victim of crime by making unnecessary comments telling "war stories" within earshot.

COMMUNICATION -- ABILITY TO MAKE ONESELF UNDERSTOOD AND UNDERSTAND OTHERS IN FACE TO FACE SITUATIONS AND TO TIMELY AND ACCURATE INTERPRETATION IN BOTH ORAL AND WRITTEN FORM.

This performance area is composed of these elements:

SKILL IN:
- completing reports accurately

ABILITY TO:
- follow simple and complex oral instructions
- communicate in writing
- communicate verbally (orally)
- follow simple and complex written instructions
- express moderately complex ideas
- comprehend and remember v... materials
- comprehend questions
- speak in public
- explain procedures verbal... (orally)
- explain charges
- recall and record work act...

****X****

******** RANKING ********

FIRST	SECOND	THIRD	FOURTH	FIFTH	SIXTH	SEVENTH	EIGHTH	NINTH

(use assigned ID NO. of squad members in spaces above)

****C**** **RATING** ****C****

(use rating of 1 - 9 in these spaces)
(rating should be made on the squad member directly above each rating space)

1 2 3 4 5 6 7 8 9
HIGH AVERAGE LOW

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "HIGH" ON COMMUNICATION

- Speaks clearly and concisely with well-thought-out ideas (when testifying in court, speaking to fellow officers, etc.)
- Has ability to communicate effectively at the appropri... educational level.

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "AVERAGE" ON COMMUNICATION

- Uses acceptable sentence structure and grammar in written reports, following a logical order of presentation.
- Recognizes and notes informative items necessary for proper job performance.

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "LOW" ON COMMUNICATION

- Includes incorrect grammar and incomplete sentences in his or her written reports, making them extremely difficult to understand.
- Creates confusion by lack of effort and ability to communicate clearly and concisely (in explaining charges, procedures, etc.).

RELIABILITY -- DEPENDABILITY IN JOB ATTENDANCE, EFFORT EXPENDITURE, ACCEPTANCE OF RESPONSIBILITY, FUNCTION APPROPRIATELY UNDER STRESS, AND ACCURACY IN ALL DETAILS OF WORK.

This performance area is composed of these elements:

ABILITY TO:
- control own emotions
- maintain composure in the face of antagonistic questions

PERSONAL CHARACTERISTICS:
- stress tolerance
- resilience
- perseverence
- dependability

****A****

****B**** **RANKING** ****B****

FIRST	SECOND	THIRD	FOURTH	FIFTH	SIXTH	SEVENTH	EIGHTH	NINTH
_____	_____	_____	_____	_____	_____	_____	_____	_____

(use assigned ID NO. of squad members in spaces above)

****C**** **RATING** ****C****

_____ _____ _____ _____ _____ _____ _____ _____ _____

(use rating of 1 - 9 in these spaces)
(rating should be made on the squad member directly above each rating space).

1 2 3 4 5 6 7 8 9
HIGH AVERAGE LOW

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "HIGH" ON RELIABILITY

Remains in assigned area, treating public he or she comes in contact with in a courteous fashion while dealing with the hazards of police work.

Recognizes police hazards in his or her assigned area and takes repetitive action until it is resolved.

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "AVERAGE" ON RELIABILITY

Immediately applies minimum amount of physical assistance/force in order to subdue a fighting suspect.

Reliably responds to all assigned calls without compl or hesitation.

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "LOW" ON RELIABILITY

Permits his or her effectiveness to deteriorate as a result of losing control of his or her emotions.

Stays "out of service" to avoid getting a late call.

DEMEANOR -- PERSONAL AND PROFESSIONAL PRIDE AS SHOWN BY HIS OR HER STANDARDS OF BEHAVIOR AND PHYSICAL APPEARANCE.

This performance area is composed of these elements:

PERSONAL CHARACTERISTICS:
- impact
- personal appearance
- integrity
- self-confidence

****B****

****B****

RANKING

FIRST	SECOND	THIRD	FOURTH	FIFTH	SIXTH	SEVENTH	EIGHTH	NINTH
____	____	____	____	____	____	____	____	____

(use assigned ID NO. of squad members in spaces above)

****C****

RATING

****C****

____ ____ ____ ____ ____ ____ ____ ____ ____

(use rating of 1 - 9 in these spaces)
(rating should be made on the squad member directly above each rating space)

1 2 3 4 5 6 7 8 9
HIGH AVERAGE LOW

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "HIGH" ON DEMEANOR

Insures that his or her patrol car, equipment and personal appearance meet high standards without reminders.

Isn't affected by peer or public pressure to compromise his or her integrity.

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "AVERAGE" ON DEMEANOR

Takes appropriate action on police calls for service in absence of direct supervision.

Requires occasional reminders to maintain proper standard of appearance and conduct.

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "LOW" ON DEMEANOR

Is loud and boisterous, and on frequent occasions is disruptive due to inappropriate comments.

Uses his or her position for personal gain.

WORK ATTITUDE -- INTERESTED IN SERVING THE PUBLIC THROUGH THE FAIR AND OBJECTIVE ENFORCEMENT AND ADMINISTRATION OF THE LAW, GAINING SATISFACTION FROM DOING HIS OR HER JOB WELL.

This performance area is composed of these elements:

ABILITY TO:
- be objective
- exercise discretion

PERSONAL CHARACTERISTICS:
- motivation for work
- energy
- career ambition
- desire for self-improvement
- independence
- initiative

****A****

****B**** **RANKING** ****B****

FIRST	SECOND	THIRD	FOURTH	FIFTH	SIXTH	SEVENTH	EIGHTH	NINTH
____	____	____	____	____	____	____	____	____

(use assigned ID NO. of squad members in spaces above)

****C**** **RATING** ****C****

____ ____ ____ ____ ____ ____ ____ ____ ____

(use rating of 1 - 9 in these spaces)
(rating should be made on the squad member directly above each rating space)

1 2 3 4 5 6 7 8 9
HIGH AVERAGE LOW

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "HIGH" ON WORK ATTITUDE

Seeks additional education and training from sources available to him/her (educational institutions, within the department, etc.) in order to improve his or her efficiency and effectiveness.

Stays abreast of crime trends in his or her area of responsibility and initiates corrective/preventive actions.

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "AVERAGE" ON WORK ATTITUDE

Makes an effort to perform his or her duties within departmental rules and regulations in the best interest of the public.

Generally establishes priority of activities during his or her normal tour of duty.

EXAMPLES OF BEHAVIORS OF OFFICERS WHO ARE USUALLY RATED "LOW" ON WORK ATTITUDE

Lets outside interests and activities interfere with his or her production on the job.

Only goes through the motions of the job and blames others' success on luck of politics.

END