1981

The Development of Two Performance Appraisal Systems for Firefighters

Robert L. Holmgren
Loyola University Chicago

THE DEVELOPMENT OF TWO PERFORMANCE APPRAISAL SYSTEMS FOR FIREFIGHTERS

by

Robert L. Holmgren

A Thesis Submitted to the Faculty of the Graduate School of Loyola University of Chicago in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

MAY

1981

ACKNOWLEDGMENTS

Funding for this research project was provided by IPA grant number 79IL04 and the City of Chicago. The author is grateful for the help provided in meeting the research goals by Robert Joyce, Gloria Zarate, Dr. Homer Johnson and Michael McDaniel. George Hartwein provided greatly appreciated assistance in the reliability data analyses. Dr. John Edwards provided guidance and constructive suggestions during all phases of the project.

Finally, the assistance and efforts of Donald Stensland, Executive Assistant to the Commissioner of the Chicago Fire Department, and Battalion Chief David Lozeau are greatly appreciated.

VITA

Robert Lloyd Holmgren is the son of Lloyd Raymond Holmgren and Elsie Victoria (Bergstrom) Holmgren. He was born in Milwaukee, Wisconsin, on July 24, 1953.

His elementary, Junior High School and Senior High School education was obtained in the Milwaukee, Wisconsin public schools. He was graduated from John Marshall Junior/ Senior High School in 1971.

Bob attended the University of Wisconsin-Stevens Point campus from August 1971 to May 1976. He graduated Cum Laude with a Bachelor of Science with majors in Psychology and Philosophy. He went on to attend The University of Minnesota at Minneapolis from September 1976 through June 1977.

In August 1977, Bob became a graduate student in the Applied Social Psychology program at Loyola University of Chicago. He worked as a Faculty Assistant for two years. In May of 1979 he began his internship at the City of Chicago, Department of Personnel. Bob is also teaching part-time at Loyola in the Psychology Department.

TABLE OF CONTENTS

TABLE OF CONTENTS (cont'd.)

TABLE OF CONTENTS (cont'd.)

LIST OF TABLES

LIST OF FIGURES

CONTENTS OF APPENDICES

CONTENTS OF APPENDICES (cont'd.)

Firefighter Performance Appraisal

Abstract

A behaviorally anchored rating scale (BARS) and a mixed standard rating scale (MSS) were developed to evaluate the performance of firefighters in a large midwestern city. Both formats were developed from the same pool of items. Items were generated using the retranslation and scaling technique of Flanagan (1954) involving two hundred twenty firefighters and officers. The psychometric and psychological qualities of the scale were assessed via a validity study in which officers rated the performance of the firefighters under their command. Validation criterion measures included performance of a brief practical test, a rank order listing on a promotional examination, education level, promotion related experience, and a rank order seniority listing. Reliability, leniency, halo, sensitivity and subjective rater preference analyses were also conducted. Approximately two hundred and fifty firefighters and officers participated in the validation study. Results showed the MSS was judged to be superior in reliability, sensitivity, criterion rated validity

and subjective rater preference. Captain raters were judged to be slightly superior to lieutenant raters. Neither format is recommended for Department-wide use without some suggested revisions.

Chapter 1

OVERVIEW OF THE STUDY

Purpose

The purpose of this research was to develop and evaluate two different formats of a performance evaluation instrument for firefighters in a large midwestern city. The better of the scales developed will be recommended for future department wide use. The scale development procedure follows closely that of Saal (1979). A behaviorally anchored rating scale (BARS) and a mixed standard rating scale (MSS) were constructed. The BARS and MSS scales were developed from the same pool of items and attempt to measure the same nine traits and characteristics of proper firefighter job performance. The traits and characteristics of proper firefighter job performance were developed over the course of four job analysis sessions with fire department personnel. Scale items were written examples of these traits and characteristics and were generated by an additional group of firefighters and officers.

Scale Quality Criteria

The major focus in this project was to attempt to create a reliable and valid rating scale. This scale should be useful to the department for promotional and job improvement feedback purposes. The two primary considerations in assessing the overall quality of the scales were their psychometric and psychological properties. The psychological considerations center around the rater's and ratee's response to the scale's impact upon their lives. The scale was constructed so that the items were meaningful to the raters and ratees. If the items are not meaningful to the raters, then accurate performance judgments cannot be made. If the ratees do not understand the items, then performance cannot improve on the basis of this feedback. A firefighter cannot change a behavior if he does not know which actions to change or which behaviors to replace them with.

Item understandability also points out an additional psychological factor in scale usefulness. The rater's and the ratee's morale is very important to successful scale implementation. Ratees who are evaluated on a scale that they do not think is the most accurate available or was not designed to their liking could balk at its usage and refuse to cooperate with scale implementation. Raters who perceive that the scale is of low quality probably will not give accurate ratings. They might decide to keep peace with their employees by giving all of them the same evaluation score. Finally, since two performance rating scale formats are being developed, the rater may have a preference for one version or the other. These psychological factors are all very important. In order to maximize the possibility of creating a psychologically proper atmosphere for scale development, the rating scales developed in this research incorporated the potential raters and ratees in scale development. The psychometric factors to be discussed below are nearly all attempts to measure these psychological considerations.

Reliability, validity and sensitivity are the three main psychometric considerations used in deciding scale quality. Reliability, according to Anastasi (1976), concerns the computation of the error of measurement of individual scores. Reliability concerns "The consistency of scores obtained by the same persons when reexamined with the same test on different occasions, or with different sets of equivalent items, or under other variable conditions" (p.103). An assessment of a scale's reliability will tell how much of a person's score reflects true variance and how much is due to error variance. The forms of reliability assessed here will be the alpha reliability coefficient, halo error and leniency error.

The reliability coefficient (cf.Cronbach, 1970) tells how well scores on a single administration of a test represent the total possible number, or universe, of scores. Specifically, Cronbach's alpha is a measure of the internal consistency of a test. If the value of alpha is high, then the test contains items that all relate to and measure the same thing. Since all items measure the same thing, it makes sense to add up their scores into a total score. If the value of alpha is low, then the scale items are not consistent and, in fact, measure different things. If they measure different things, then it makes no sense to add up the scale values to compute a final score. The coefficient alpha indicates the amount of observed score variance that is not error variance. This coefficient depends on the spread of scores and the number of items making up a person's score.

Halo error is a constant error (rather than a random one) in which the rater's scores are unduly reflective of a single positive or negative characteristic of the ratee. The rater tends to give the ratees scores on all test items that reflect his favorable or unfavorable bias toward the ratee. Halo errors can be minimized by utilizing objective, behavioral scale items rather than personality type items. Halo effect is in part reflected in a scale by high intercorrelations between scale items and small standard deviations of scores across performance dimensions of a ratee (cf.Saal, 1979).

Leniency errors occur when raters are unwilling to make negative or less than favorable ratings. The most direct measure of leniency is the skewness of the scale score distribution. Large negative values of skew reflect leniency (assuming no ceiling effects). With leniency, the distribution of scores is piled up at the high scale values. Leniency can also be assessed by a t-test between the actual scale total score and the scale midpoint. A significant positive *t* score indicates leniency.

Validity concerns what a test actually measures and the accuracy of these measures (Anastasi, 1976). A performance appraisal rating scale that is valid will accurately measure a person's job related performance. Different types of job duties can be focused on to create a valid rating instrument. The major concern in performance ratings that are used for promotional purposes is criterion validity. The performance appraisal instrument will be criterion valid if evaluation scores can significantly predict proper firefighter performance. These criterion measures will be collected at the time of the performance ratings. Criterion measures will reflect the firefighter's promotability and the ability to use the tools of firefighting.

Selltiz, Wrightsman and Cook (1976) describe a sensitive rating scale as one that is capable of making distinctions between ratees that are fine enough to meet the intended purposes. If the purpose of the rating scale is to rank order individuals on their job performance ability, then the rating scale must be able to distinguish between the performance ability of any two people that actually perform differently. Kurtosis provides one measure of sensitivity or dispersion of the score distribution. Kurtosis is a measure of the shape of the distribution of scale scores. If the kurtosis is positive, then the distribution of scores is narrow and peaks on one value. Such a distribution would not accurately distinguish between individuals. A high negative kurtosis indicates a flat distribution of scores and more easily distinguishes between individuals.

Development of a Behaviorally Anchored Rating Scale

From the standpoint of the psychological criteria of high scale quality, involving the raters and ratees in scale development seems important. One of the first performance rating scales to utilize job incumbents in the scale development process is the behaviorally anchored rating scale (BARS. cf. Dunnette and Borman, 1979). The BARS, described in greater detail below, is a rating scale in which each of the potential scale values are "anchored" by a behavioral statement describing the level of performance associated with that particular number. Including job incumbents in the scale development procedure is believed to be a way of reducing rater error, especially when the critical incidents method is used (cf. Smith and Kendall, 1963). Critical incidents are easily understandable, highly observable behaviors that are included as scale items, rather than using more general personality traits. Although the overall purpose for the scale is to assess a person's general performance score, critical incidents are used in an attempt to make each item very specific and easily understood. With the decision to include specific rather than global tendencies as items comes the duty of ensuring that all relevant dimensions are included in the test. This is usually accomplished through an extensive job analyses. Also, by extensively tapping the knowledge of those individuals who will eventually be using the scale, the scale developed will be clearly job oriented, rather than personality-trait oriented (Schwab, Heneman and DeCotiis, 1975).

Schwab et al, (1975) describe the five major steps in BARS development. The first step is the collection of critical incidents. Utilizing a technique developed by Flanagan (1954), persons with extensive knowledge of the to-be-evaluated position are asked to write specific examples of effective and ineffective job performance. The second step clusters the behavioral incidents into approximately ten performance dimensions. Each cluster, once collected, is defined according to the items within it. Schwab et al. (1975) indicate that having incumbents define the dimensions after writing the items serves to keep participants focused on the "specific behaviors (critical incidents) rather than on traits (global performance dimensions)" (p. 551). However, from a practical point of view, it would be more efficient (and, therefore cheaper) to have incumbents write only those items that are geared to specific performance dimensions. The dimensions or categories would be written first and the critical incidents later. This reverse order is the development sequence used in the present research.

The third and fourth steps of BARS development are performed at the same time. Incumbents are asked to fill out a retranslation and scaling questionnaire. Retranslation involves having job-knowledgeable people read each of the behavioral incidents. The incident is then matched with the job performance dimension that best describes the category to which the example belongs. Those critical incidents with the highest consistency in retranslation to the dimension categorizing them are retained for scale development.

Scaling, the fourth developmental step, consists of a Thurstone-type ranking of each incident on its performance quality level. Subject matter experts, firefighters and officers in this case, are used as judges who place a performance quality scale value on each of the retranslated items. The low numbers on the (usually seven-point) scales indicate.the poorer levels of job performance, the high numbers indicate superior performance. Scale items are selected on the basis of a small standard deviation. The smaller the standard deviation, the more compact the range of perceived scale values for a specific item. Items with small standard deviations indicate a high degree of inter-rater agreement on the scale value of the item.

The fifth step is the development of the final instrument. A set of critical incidents is selected for each of the approximately ten dimensions on the complete scale. Each dimension's scale will contain seven whole number scale values. Critical incidents are chosen to anchor, as best as possible, each of these whole number scale values for each dimension. The critical incidents are printed alongside the appropriate scale value on the seven point vertical scale. Items at the bottom of the scale are examples of the poorer levels of performance. As one reads up the scale the incidents become examples of better levels of performance. At the top of each scale, anchoring statements that represent the best possible levels of performance are printed.

The BARS is used to evaluate the performance of a job-holder by having the rater make a simple "X" somewhere along the vertical rating scale line of each dimension (i.e., subscale). The "X" should be located at the point on the scale that best represents the level of performance quality of that individual. Each of the subscale values are added up to yield a total performance evaluation score.

The benefits of using a BARS performance evaluation scale lie mainly in the scale development procedure. Job incumbents who have input into scale development should know the duties and requirements of the position better than anyone else. The job dimensions and behavioral incidents generated will be appropriate to the actual position (Campbell et al., 1970). The wording of these dimension definitions and behavioral anchors should be meaningful and unambiguous to the raters since they were involved in writing the items. This increased meaningfulness over other types of scales should, in turn, improve the reliability of the scale and make the scale content valid (Schwab et al., 1975). Smith and Kendall (1963) indicate that retranslation only for items that are highly agreed upon should also reduce leniency and central tendency errors.

Meaningful scale items developed by the scale users may also serve to increase the rater's motivation to make accurate ratings (Dunnette, 1966). Ratees who perceive that they have been rated by individuals that have been involved in the scale development procedure, may feel more accepting of the scores given to them. Acceptable evaluation scores based on meaningful behavioral incidents may also be useful in providing helpful feedback to raters. Specific behavioral items can be used as suggestions for improvement or even as the basis for training programs (Blood, 1974; Cummings and Schwab, 1973).

Development of a Mixed Standard Rating Scale.

The second performance evaluation instrument developed in this study is a Mixed Standard Rating Scale (MSS, Blanz and Ghiselli, 1972). The MSS can be developed from the same set of retranslated and scaled items used for the BARS. The MSS format requires three behavioral examples per dimension of proper job performance. The three incidents should each describe a different level of performance of that dimension. One example should describe a superior level of performance. The second example should describe the average level of performance of that same dimension. The third example will describe a substandard level of performance of that dimension. The three items for each dimension of the MSS roughly correspond to the top (first), middle (fourth) and bottom (seventh) sentences on a BARS subscale.

On the final scale, all of the critical incidents are presented individually in random order to the rater. For each item presented, the rater decides whether the ratee performs better than, the same as, or worse than the example. If the ratee performs his duties worse than the example, the ratee is assigned a minus (-) for the item. A performance perceived to be the same as the example is rated with a zero (0). Performances superior to the written example are given a plus (+). Each ratee ends up with one of the three possible marks (i.e., a plus, zero, or minus) on each of the three levels of incidents (i.e., superior, average, or sub-standard) for each performance dimension.

Performance evaluation scores are given by assigning a numerical scale value to each possible mark on each incident level of the dimension. In this research the score weighting scheme developed by Saal (1978) is followed. Item marked with a plus are assigned the numerical score of eight (8), five (5), or two (2), depending on whether the incident is a superior, average, or sub-standard example respectively. Scores of seven (7), four (4), or one (1) are assigned to incidents marked with a zero for superior, average, or sub-standard items respectively. Finally, critical incidents marked with a minus are assigned a scale value of six (6) if the incident is a superior level of performance example, a three (3) if the incident is an average level of performance, and a zero (0) is the incident is an example of a substandard level of performance. All numerical values for each dimension are summed and eight points are then subtracted. Thus, the maximum score on each dimension will be seven points. The minimum score on any dimension will be one point. The sub-scale scores are then summed to yield a total MSS score.

The resulting MSS scale is a three step Guttman-type scale for each dimension. It is expected, for example, that ratees with a plus on a superior item will also receive a plus on the average and substandard items for a specific dimension. Likewise, a ratee with a minus on a substandard item should not get any mark other than a minus on the average and above average examples of that performance dimension.

The benefits of the MSS should be the same as those of the BARS. Both formats get their scale items in exactly the same way. However, an MSS should have benefits not found in a BARS, due to the differences in format. Blanz and Ghiselli (1972) indicate that the MSS should have a reduced halo and leniency error. Since the items are presented in random order, the rater should not be able to form a clear perception of the order of merit of dimension items. The use of a hidden metric in this scale format should reduce halo error from those scales with an obvious metric (e.g., BARS). On a BARS scale, the rater makes a mark of evaluation directly on top of a numerical scale value printed on a vertical line. A rater can quickly figure out how to be easy (lenient) on a ratee and may simply give the same high mark on each vertical scale irrespective of the incident anchor points (halo). Such a rating scale will have higher leniency and halo than a scale with a hidden metric (Blanz and Ghiselli, 1972).

Saal and Landy (1977), however, disagree with the logic that says the halo error will be reduced in an MSS. They indicate that a random mixing of the item may encourage "nondiscriminative rating practices" (p.22). Th assumption they make is that halo error will increase because the mixing of the traits will make it more difficult for a rater to discriminate between them while making the ratings. However, Saal and Landy (1977) agree that the leniency error should be smaller with the MSS method.

An additional potential benefit of the MSS on which both Saal and Landy (1977) and Blanz and Ghiselli (1972) agree is that "less competent" raters, ratees and inadequate dimensions can all be easily identified. The MSS provides a measure of rater ability. Lenient raters have a tendency to make unduly high ratings. The distribution of their rating scores has a small range and is highly negatively skewed. An MSS scale can pick out raters who are inconsistent in their ratings as well as lenient raters. Inconsistent raters give people minus ratings on substandard or average items while giving zeros or pluses on the superior items, for example. Such findings indicate which raters should receive additional training for their rating duties. Inadequate dimensions are those dimensions whose three component ratings are consistently marked outside of their rank order (e.g., superior ="+"; average ="-"; substandard ="O") by various raters. Thus, the format of the MSS provides data to measure the reliability of the ratee, the rater and the scale itself (Blanz, 1965).

Literature Review.

Research on the two rating scale formats has yielded mixed results. Dunnette and Borman (1979), in their Annual Review of Psychology, chapter entitled, "Personnel Selection and Classification Systems," list four studies that show psychometric superiority of BARS (Borman and Dunnette, 1975; Campbell, Dunnette, Arvey, and Hellervek, 1973; Groner, 1974; Keaveny and McGann, 1975). They also list six studies each indicating no consistent advantage of BARS over other scale types (Bernardin, 1977; Bernardin, Alvares and Cranny, 1976; Borman and Vallon, 1974; Burnaska and Hollmann, 1974; DeCotiis, 1977; Zedeck, Kafry and Jacobs, 1976). In addition, three studies are listed which indicate MSS-type Guttman scales having shown no consistent psychometric superiority (Arvey and Hoyle, 1974; Finley, Osburn, Dubin, and Jearneret, 1977; Saal and Landy, 1977). However, Schwab et al. (1975) note that many of the nonsignificant results of both types of formats may have been due to departures from the scale development process. Due to this consideration and the fact that job incumbents are involved in scale development, Dunnette and Borman (1979) support the use of behaviorally based rating scales.

However, it must be pointed out that the use of job-knowledgable incumbents in the scale development process is not limited to behaviorally based rating scales In fact, most scale development begins, in part, with input from the people whom the scale hopes to eventually test. The use of job incumbents in scale development is one method of helping to improve the validity of the scale.

Saal (1979) developed two sets of rating scales for the performance evaluation of police patrol officers. Both a BARS and an MSS format were developed for supervisory and peer raters of the patrol officers. Thus, a total of four performance appraisal scales were developed: two BARS scales, one for supervisors and one for peers, plus two MSS scales also, one each for supervisors and peers. However, since supervisors, and not peers (fellow officers) would be making the actual departmental performance evaluations, the reliability analyses were conducted only on the ratings made by supervisors. Incorporated into Saal's reliability analyses were calculations of leniency error, halo error and central tendency error.

Leniency refers to the tendency of a rater to make ratings that are "higher" or "better" than ratee deserves (Saal, 1979). In practice, this means that the average rating by a rater should not be significantly different from the midpoint of the scale. Analyzed in this way, the supervisory MSS scale developed by Saal (1979) had less leniency error than the supervisory BARS when the effects of the raters and ratees were held constant. However, all subscales on all four scales did have leniency errors to some extent. When supervisors rated officers on the peer-developed scale, the BARS contained relatively less leniency error than the MSS. Saal (1979) interprets these results as indicating that an MSS will yield less leniency effects when utilized by the people involved in the scale development. Analysis of the skewness of the scales, as well as a rater by ratee by dimension analysis of variance, reinforced this interpretation. Leniency was similar between the two scales when the subordinate (peer) developed scales were analyzed according to skewness.

Halo error is a function of the rating scale rather than of the rater specifically. It is the tendency for the different scale dimensions to have a similar high ranking without providing any degree of discrimination between dimensions. In Saal's study, high correlations between dimensions on both supervisor scales indicated a high amount of halo error. Relatively speaking, however, the supervisory MSS ratings had less halo than the supervisory BARS ratings. The subordinate rating scales also contained a significant amount of halo as measured by the intercorrelations between dimensions. But again, the subordinate MSS contained "marginally less" halo than the BARS (Saal, 1979, p.27). Direct comparisons of dimension variances and a rater by ratee by dimension analysis of variance both support the intercorrelation results. In general, both scales had a significant amount of halo error with the BARS having relatively more halo than MSS. However, this interscale difference failed to achieve statistical significance.

Central tendency is the tendency of ratings to pile up around some central value on each dimension. This tendency leads to the problem of the failure to discriminate between individuals on the scale dimensions. According to a kurtosis analysis of Saal's data, the supervisory MSS ratings had less central tendency and greater discriminability between ratees than the supervisory BARS. No differences were found in the subordinate results. Comparisons of the standard deviations of the scales suggest that the greater discriminability of the supervisory MSS rating is quite small.

The subjective rater preference questions asked by Saal failed to provide any clear preference for either of the two rating scales. The BARS scale was slightly preferable in making administrative decisions for promotions and transfers. There was no preference between scales concerning their uses as feedback and as performance improvement guidelines. Since Saal discovered no great preference for either format, the scale recommended for adoption by the Police Department involved was a combination of different subscales from both scale formats.

As cited in Dunnette and Borman (1979), DeCotiis presents an experiment that shows a BARS format police patrolman rating scale not as resistant to error as a numerically anchored scale and a trait rating scale. The three scales were not significantly different from each other on resistance to leniency and central tendency errors. None of the three formats in DeCotiis (1977) were resistant to halo errors. In addition, the BARS format was no better or worse than the other two formats on a number of psychometric and psychological criteria including the extent of interrater reliability and rater preferences of the format for applied usage. Overall, DeCotiis (1977) ranked the numerically anchored scale the most favorable and the BARS scale the least favorable.

The Saal and Landy (1977) study mentioned in Dunnette and Borman (1979), compared supervisory and peer ratings on MSS and BARS formats of police patrol officers. In general, the MSS format had fewer leniency errors in many of the dimension ratings. The BARS scale had higher interrater reliabilities than the MSS. However, Saal and Landy (1977) point out that an inferior MSS scoring routine was used in the study which may have contributed to the reliability irregularities of the MSS.

Landy and Barnes (1978) indicate some of the potential problems in using a BARS scale that would not be found when using a traditional psychometric, Thurstonian paired-comparison judgment scale. The retranslation and scaling procedures of scale development require respondents to make absolute judgments on the quality of a performance. Nunnally (1967) points out that comparative judgments are much more common and more likely to be psychometrically sound. Accuracy may be lost in absolute judgment techniques because there are fewer data points per anchor than in a comparative judgment technique. The same holds true for the BARS scale as a whole. Raters make one judgment per dimension. In the MSS, raters make three judgments per dimension. The more data points (judgments) per dimension the greater the accuracy, all other considerations held constant.

Schwind (1978) points out two additional problems with BARS formats. First, there is a large amount of information waste in the use of critical incidents. Often-times eight to ten critical incidents are required for a single judgment. Schwind (1978) proposes a behavioral description index (BDI) as as improvement on the BARS. The BDI is very similar in format to the MSS except that the rater decides if the ratee exhibits the printed behavior at one of a set number of frequencies of performance. The multiple data points per dimension provide a higher informational content. The BDI developed by Schwind (1978) showed less halo and central tendency errors than a comparable BARS.

The second problem with the BARS format pointed out by Schwind (1978) is that the individual dimension rating scales (i.e., the subscales) are often multidimensional. Multidimensionality makes it extremely difficult for a rater to choose the most typical level of performance of a ratee. The ratee may be quite competent on one dimension and perform poorly on a second dimension within the same subscale. The interrater agreement can also be hampered by having different raters focus on different aspects within the same performance dimension. Technically speaking, this last problem should be controlled by retranslation.

Katcher and Bartlett (1977) conducted an experiment on an MSS developed for police supervisory personnel. The study directly tested the relationship between multidimensional rating scales and inconsistent ratings. The degree of dimensionality within a specific task dimension definition was assessed by principal components analysis. Results indicated that unidimensional subscales will contain fewer inconsistencies in ratings than multidimensional subscales. The problem with a multidimensional scale is that a rater's task of comparing recalled behaviors to scale anchors becomes very difficult. Since it will be impossible for a rater to focus on the level of performance on a single dimensional performance concept, the rater may be forced to base the rating decision on a "general overall impression or to make invalid compromises between different anchors" (Katcher and

Bartlett, 1977, p. 9). In other words, when the rating dimensions become muddled and unclear to the rater, ratings will be based on perceptions of global personality traits, (e.g., halo, leniency), rather than objective, observable behaviors.

The literature directly related to the performance appraisal of firefighters is quite sparse. Matticks (1977), in an article pointing out the difference between in-title performance appraisal and promotability ratings, lists sets of "activity skills" and "knowledge" required by firefighters. He suggests that firefighters must be able to "function under pressures up to crisis level," "make important personal decisions," "perform physical tasks requiring strength and dexterity" and "work as a team member." In addition, firefighters must be knowledgeble of basic hydraulics, physics, chemistry and equipment and apparatus maintenance (p. 51). Although these items are not specific behavioral examples, and thus are not directly useful for an appraisal scale, they do suggest the content of such a scale.

Wenger (1978) suggests that the most important of these skills, abilities and knowledge is the physical fitness of the firefigher. Wenger cites evidence indicating that the overall efficiency and operational costs of a department is directly related to the physical condition of the firefighters. The remainder of the professional firefighter literature surveyed generally presents suggestions for maintenance of these characteristics in fire departments. Davis and Wright (1979) suggest means by which qualified recruits can be selected who will have these skills and knowledge required. Vincent and Lyle (1979) suggest performance standards for various practical tests to help in skill maintenance of nine engine company evalutions (e.g. hydrant and hose use, etc.) Sparr and McKee (1979) suggest that assessment centers are an appropriate institution to set up as a check on whether the various performance skill standards are being maintained. For the purposes of the present research, an assessment center would be very costly.

However, several high quality technical reports on firefighter job analysis are available from the Personnel Research and Development Center of the United States Civil Service Commission Bureau of Policies and

Standards. Van Rijn (1977 a,b) conducted an extensive computer based job analysis of the District of Columbia Fire Department Entry-Level Firefighters position. The information provided in this series of studies is similar to the information gathered in the present study. The aim of the Van Rijn (1977 a,b) study was the eventual development of a written test for the firefighter position. This written test was designed to reflect the information about the knowledge, skills and abilities of the firefighter position assessed in the job analysis. Thus, the job analysis presented provides information on what the District of Columbia firefighters do on the job, the skills required to perform these duties and the knowledge required for a basic understanding of the job. These statements of the knowledge, skills, and abilities of the firefighter position are organized into dimensions of the job. Similiar dimension categories were also used in the creation of the BARS and MSS performance appraisal scales developed in the current study. For the District of Columbia firefighters twelve dimensions were obtained. These dimensions are presented in Table 1. Of particular interest in this list is the noticeable

Table 1

Job Analysis Dimensions of Entry Level Firefighting in the District of Columbia Fire Department

1. Responding to Alarms
2. Performing General Firefighting Operations
3. Performing Ladder Truck and Related Operations
4. Performing Fire Extinguishing and Related Operations
5. Performing Salvage and Overhaul Operations
6. Performing Special Emergency Operations
7. Maintaining Apparatus and Equipment
8. Providing First Aid and Assistance
9. Inspecting, Investigating, and Code Enforcing Activities
10. Training
11. General Management, Administration, Housewatch, and Related Activities
12. Performing Public Relations and Community Activities

lack of dimensions concerning the personalities and characteristics of individual firefighters. Instead, emphasis is placed on the specific classifications of duties of firefighters.

Bownas, Heckman and Anderson (1977) presented a technical report on a nationwide job analysis of the entry level firefighting position. The study was conducted in 100 different fire departments with over 500 people involved from across the country. The eventual aim of this project was the development of a generally applicable task analysis for the entry level firefighter position (Personnel Decisions Inc., 1977). In addition, the appendicies of the Bownas et al. (1977) technical report contain the details of the development of a BARS performance appraisal scale based on the above firefighter task checklist.

The Bownas et al. (1977) task checklist for firefighters is developed around a set of sixteen dimensions of proper firefighter job performance. This list of sixteen dimensions is presented in Table 2. The list provided here is more extensive than the list of dimensions presented in the Van Rijn (1977) District of

Table 2

Firefighter Task Checklist Dimensions Developed in the Bownas, Heckman and Anderson (1977) Study

1. Performing Rescue Operations
2. Performing Salvage and Overhaul
3. Performing Ladder Operations
4. Forcibly Opening Structures and Enclosures
5. Applying Ventilation Procedures
6. Applying Knowledge of Fire Characteristics
7. Following Standard Safety Procedures
8. Performing Hose Evolutions and Applying Extinguishing Agents
9. Operating Apparatus
10. Administering Emergency Care
11. Dealing with the Public
12. Performing Preplanning and Fire Prevention Inspections
13. Participating in Training and Education
14. Reconditioning and Maintaining Equipment
15. Performing Routine Station Duties
16. Getting Along with Peers

Columbia Study despite considerable overlap. In addition to outlining the basic fire duty operations, Bownas et al. (1977) also present dimensions concerning station duties and personal interactions. This points out that firefighters are expected to be more than simply fire-fighting automatons. Firefighters also are acknowledged to spend a good amount of time dealing with each other and with civilians in non-emergency settings.

Behavioral examples for each of the sixteen job performance dimensions are presented in the Bownas et al. (1977) study. Of particular interest in reading through these examples is the specificity and detail involved in the sentences. For example, in the "performing rescue operations" dimension, the following statement is used to describe a superior level of performance, "The firefighter volunteered to go into a tank to rescue a person overcome by highly volatile gas fumes" (Bownas et. al., 1977, Appendix P, p.1). This example points out both the benefits and the problems in using highly specific behavioral statements in performance appraisal. Specific statement are unlikely to be misunderstood by the raters and refer to an exact situation which the rater will clearly have either experienced or not experienced. Very little chance for ambiguity and vagueness occurs. However, such statements are not universally applicable across the fire department. If a particular company has no dealings with volatile gas fumes, then the example is of no use to the rating officer. Such a situation could not have occurred within the rating period under assessment. A more general statement, perhaps involving unconscious victims, would be more likely to be applicable across the department. Thus, there appears to be a point of diminishing returns in the specificity of behavioral examples. This is one of the main reasons that the raters and ratees involved in the eventual performance evaluation are used in scale development.

Summary of the Study

In the study described below, a BARS and an MSS performance appraisal scale were developed for firefighters. Both the officers that make the performance ratings and the firefighters rated were utilized in the scale development techniques. The five step scale development procedure described above was followed (i.e., dimension generation, critical incident generation, retranslation, scaling, final format writing). Once both formats were generated, reliability and validity studies were conducted. Officers were asked to rate the performance of the group of firefighters under their command on each of the scales. Firefighers also rated their fellow company members (usually about five firefighters on each company) on each of the two formats. Reliability, leniency, halo, discriminability and subjective rater preference analyses were based on the total group. Validity analyses was based solely on the performance ratings given by the officers. Validity criteria include the firefighter's performance on a brief practical test taken at the time of the performance rating, the firefigher's rank order on the lieutenant promotional examination list posted shortly after the collection of the performance ratings, and several other self-report indices. A final performance rating scale format is then recommended to the department based upon the above analyses.

Chapter 2

METHOD

Subjects

Firefighters, lieutenants and captains from the fire department in a large midwestern city took part in this study. All personnel had been in their current job title for at least six months, some of them for as long as twenty years or more. The fire department is organized into groups of up to six individuals assigned to a specific truck or apparatus. These companies are composed of one officer (a lieutenant or captain) and firefighters. On any single shift, up to three companies may be located in one station house. Each station house has three shifts, each with one officer in charge. A captain is the usual ranking officer in any individual station. Stations are organized into different battalions. Each battalion has one chief for each shift. This is the major field organization of the fire department. Eighty-three companies were used from various locations around the city. All respondents were male.

Validation Instruments

Both firefighters and officers filled out two semantic differential attitude scales. Each scale assessed the rater's opinions about one of the two performance evaluation instruments. These attitude scales measure the face validity of the scales. In this instance, face validity concerns whether or not the performance appraisal scales are perceived to measure what they are intended to measure according to the firefighters and officers involved. The same six items are used on both semantic differential scales. Following the two attitude scales, three general questions on format preference were asked. First, raters were asked which of the two performance evaluation scales they preferred (or neither). The second question covered preferences about a more specific distinction between the two scales. The two formats differed in that on the BARS scale the rater knew the exact numerical score he was assigning to a firefighter on each subscale. On the MSS, the rater did not know the numerical value he assigned to a firefighter; rather, he assigned either a plus, zero, or a minus to each item. Firefighters and officers stated whether they preferred to know the numerical score they were assigning to a man immediately or not. Third, the behavioral examples in the two scales were a bit different. The BARS had more examples than the MSS. The MSS contained some examples that were aggregates of several BARS scale anchors. Raters were asked to indicate which of the two sentence types they preferred (or state that both or neither format's examples were acceptable). The Semantic Differential Attitude Scale is presented in Appendix A. The six items on the semantic differential were chosen to represent six evaluative viewpoints with respect to the BARS and MSS.

The firefighter Self-Report Questionnaire contains twelve items designed to gather criterion validation criteria. Items assessed the firefighter's scores on various firefighter certification exams, the number of college and fire science courses taught or taken and how often the ratee has performed special duties on the job. The Self-Report Questionnaire is in Appendix B.

The brief firefighter practical performance tests are a collection of eight short tests designed to measure a firefighter's ability to use some of the accoutrements of fighting fires. Firefighters are asked to name and utilize some tools they would be required to use in various emergency situations. Scores were assigned on a 100 point scale to indicate the firefighter's facility with the equipment. In this study, each firefighter performed only one of the eight possible practical performance tests. Tests were randomly assigned over the sample of companies tested. The eight practical performance tests are presented in Appendix C.

A Battalion Chief, who acted as liaison to the research team, did all the ratings on each of the eight practical performance tests. This minimized rater errors. However, wide differences in ability to perform the eight practical tests probably exist, even within the same firefighter on different versions of the practical test. Thus, a firefighter may be able to score quite highly on one test (e.g., ropes or ladders) and do quite poorly on a different test (e.g., spanners). The difficulty level of the eight practical tests may not be identical. High scores on one test may not indicate high scores on another practical test. This puts some doubt on the utility of these practical performance tests as validity criteria.

Since the officers are the men who will eventually be rating the performance of firefighters, once a scale is implemented department wide, their attitudes toward the behavior of making accurate evaluations are very important pieces of information. In order to tap the intention to make accurate evaluations each officer filled out an extensive (46 item) Fishbein-type expectancy-value attitude scale.

Briefly, the theory behind this scale states that the Behavior (B) of accurately rating the performance of the firefighters under the officer's command is a function of the officers intention (I) to do this behavior (Fishbein and Ajzen, 1975). This intention is, in turn, a function of two things: the officer attitude toward the behavior (AB) involved and the subjective norm (SN) of the officer's "significant others" toward accurate ratings. The attitude toward the behavior is defined as the weighted sum of the officer's individual beliefs about the behavior each multiplied by his evaluation of the beliefs. These are the types of questions asked in section three of the questionnaire. The subjective norm is defined as the weighted sum of the officers individual normative beliefs about each of his "significant others" each multiplied by the officer's motivation to comply with that group. These items are in section five of the questionnaire. These theoretical relationships are presented algebraically in Figure 1. The Officer's Attitude Survey is presented in Appendix D.

The scale consists of five major subsets of items. These items were written from information gained in an interview with a Fire Department liaison to the research team. The first question asks the officer to tell the likelihood of it being true that he intends to accurately rate the performance of each of the firefighters under his command. This intention is the major criterion variable of the survey.

Questions two through seven compose a semantic differential attitude scale. This scale is designed to measure the officer's attitude toward accurately filling out his (forced) preferred choice of a performance evaluation instrument. The third section consists of a set of ten pairs of questions. The first item in each set asks the likelihood that some statement of concern

Figure 1

Algebraic Representation of Fishbein's Expectancy Value Attitude Theory

1. $B \sim I = (AB)W_1 + (SN)W_2$

   B = behavior

   I = intention

   AB = attitude toward the behavior

   SN = subjective norm

   W_1 and W_2 = regression weights

2. $AB = \sum_{i=1}^{n} (B_i) \quad (EB_i)$

   B = behavioral belief

   EB = evaluation of the belief

   n = number of salient beliefs

3. $SN = \sum_{i=1}^{n} (NB_i) \quad (MC_i)$

   NB = normative belief

   MC = motivation to comply

   n = number of salient normative beliefs

to accurate performance evaluation is true. The second question asks whether the first statement of the pair is a good or a bad thing to have happen. For example, "Accurately filling out the efficiency mark rating scales will cause an increase of hostility between Lieutenants and Firefighters." This is followed by, "Increasing hostility between Lieutenants and Firefighters is (somewhere on a seven point scale from good to bad)." These ten pairs of items are designed to measure the officer's attitude toward the behavior (AB) of making accurate ratings.

The fourth section is a single item that assesses the likelihood of it being true that most people in the department who are important to the officer, and whose opinions the officer respects, think that the officer should (or should not, on a seven point scale) accurately complete the efficiency mark scale.

The fifth section consists of nine pairs of items that tap in more detail the subjective norm (or normative belief, NB) of the department toward accurately filling out the efficiency mark scale. The importance of the opinions of several groups of people, from firefighters on the officers shift at the fire station, up to the Commissioner of the Fire Department, were assessed. The first questions in each set asked the likelihood of the perception that each group of "significant others" thought that the firefighters should be rated accurately by their officer. The second question in the set assessed if the officer wanted to do what this group of "significant others" wished.

Procedure

The data collection procedure for this study was performed in two major steps. The first step involved development of the two formats of the performance evaluation questionnaire. The second step was to select the better scale via validity, reliability and attitudinal analyses.

The process of developing two different formats of a performance evaluation instrument for firefighters began by holding four job analysis sessions. Each session developed one version of the characteristics of proper firefighter performance. The first two sessions gained information from firefighters about the various dimensions of their proper job performance. The remaining two sessions gained performance evaluation dimensions from lieutenants and captains. Officers had separate meetings from firefighters in order to minimize the degree of influence of the officers over the responses given by firefighters. A total of thirty-four Fire Department personnel attended the sessions. About eight individuals attended each meeting.

The job analysis sessions began by having the session members brainstorm and list all of the traits and characteristics deemed necessary for proper firefighter job performance. Once the list was developed, the session members read the list of items developed during the previous session. Duplicate and redundant traits and characteristics were eliminated. Each of the remaining items were then defined by the group. Items that were previously defined but not eliminated as redundant were subject to the addition or deletion of phrases deemed necessary by the group. Following the final job analysis session, the four versions of the traits and characteristics of proper firefighter job performance were synthesized with the help of a fire department Battalion Chief. The chief corrected any wording that seemed incorrect or inappropriate (see Appendix E). It is noteworthy that there is considerable overlap between this list of dimensions of the firefighter job and the lists of dimensions presented in District of Columbia Fire Department analysis (see Table 1) and the Bownas, Heckman and Anderson (1977) study list of dimensions (see Table 2). This overlap suggests that the dimensions used in the present study show considerable construct validity as far as the dimensions of the job are concerned.

After development of the dimensions and definitions of proper firefighter job performance, a second set of about six meetings was held. Fifty-two firefighters and officers participated, each in their various fire stations around the city. Since participants were on duty, these companies were periodically forced to leave the data collection procedure to go on a run to a potential fire.

These firefighters and officers wrote behavioral examples of each of the definitions (dimensions) of proper firefighter job performance. For each definition, fire department personnel wrote three behavioral examples. The first example described a critical incident in which a firefighter performed a specific dimension of his job in a superior manner. The second example described an example of an average level of performance of the particular dimension. This second example described the standard level of performance. The third example described a sub-standard or poor level of performance of the particular dimension of proper firefighter job performance (see Appendix F).

Once all the behavioral examples were collected, an additional group of one hundred ten firefighters and officers each filled out a portion of the retranslation and scaling questionnaire. This questionnaire contained all of the above written items and required two responses for each behavioral example. First, the firefighter or officer attempted to match the behavioral example with the definition from which it was written. Then, on a seven-point Likert scale the participant judged the level of performance quality of the item. The retranslation question assessed whether a given item was perceived as an example of the dimension it was intended to represent. The scaling question measured the perceived level of performance quality of the item. The scaling question was a check on whether the item was indeed perceived as a superior, average or sub-standard level of performance (see Appendix G).

From this pool of retranslated and scaled items the two formats of the performance evaluation instrument were constructed. Analysis of the data indicated that seven of the original twelve dimensions of proper firefighter job performance were successfully retranslated and scaled. Each of these seven dimensions categorized a large number of statements that were not confused with any other dimension's statements. However, the teamwork and compatibility dimensions were readily confused. Items from these two dimensions were frequently retranslated to both of these two dimensions, but not to any other dimensions. These two dimensions were, therefore, combined into a single dimension. Furthermore, the responsibility and the consistency in performance dimensions were in a like manner combined into the willingness to work dimension. The interrelationship between these three dimensions was not, however, quite as high as the teamwork and compatibility dimensions. The BARS is presented in Appendix H and the MSS in Appendix I. Both formats were subject to reliability, validity and scale preference research. This is the second major portion of the study.

The actual retranslation and scaling questionnaire was divided up into three parts. Each part contained approximately 260 items. Any one respondent retranslated and scaled one booklet of about 260 items. Furthermore, the entire list of 775 items, contained in three different parts, was organized into three different forms. Each form presented the 775 statements in a different order. Form A present the items in the reverse order from Form B. Thus, items one through 260 (part 1) of form A were items 260 through item one (reverse order) of form B, part 1. Form C placed the items in the middle of forms A and B at the beginning and the first item of form A and the first item of form B at the end of form C. Thus, each respondent randomly received one of nine possible combinations of the items.

For the construction of the BARS scale, the most successfully retranslated items for each dimension were typed alongside a seven-point vertical rating scale line at the location of their appropriate scale valve. Although sentences were chosen so as to anchor as many of the scale values as possible, gaps did occur in the rating scales. These gaps tended to occur in the middle and extreme top (high quality) levels of the scales. Apparently it was very difficult to write items that were examples of high quality performance. The problem with selecting average performance level items had to do with the scaling criteria used. Items were considered to be successfully scaled at one level when the standard deviation of the rated scale values for the item was approximately 1.0 or lower. This restriction caused a large number of average level items to be thrown out. Average level items often had large standard deviations. These larger standard deviations may have been due to these ranges ability to vary more greatly than score located at the extremes. Dispite this problem, the BARS scale development procedure was followed very closely. This study's purpose was to evaluate the comparative worth of the BARS with respect to the MSS, not redesign the BARS.

Validation data was collected from eighty-three (83) companies. Each company consists of one officer (either a lieutenant or a captain) and between one and five firefighters. The fire stations in which these data were collected have between one and three companies. Companies were also "detailed" to other fire stations in order to speed up the validation data collection procedure.

At the beginning of each session all personnel, except for the engineers, assembled in the kitchen and were given a brief introduction and explanation of the approximately three-hour session. The station house members were told that they were participating in a research project to attempt to validate and collect information about the Fire Department's preferences toward the two rating scale formats. They were told that one of the two formats would be recommended as the new performance evaluation instrument for the Fire Department.

The directions for filling out each of the two formats (BARS and MSS) were read aloud to the group and any questions that arose in this regard were answered. When all members in the house understood how to fill out the scales, they rated the other members in their company. The officer rated each of the firefighters under personal command that were present at the meeting on each of the two performance evaluation scales. Firefighters rated the performance of the other firefighters in their company that were present at the session on both scales. No one rated themself nor did anyone rate the officer. Firefighter candidates (who had been on the job less than six months) were not rated. The general feeling among fire department personnel was that these persons had not been company members long enough to accurately rate or be rated. Individual firefighters that were temporarily "detailed" to a company present at the session also were not rated. Finally, each firefighter and officer was free to choose whether or not to participate in the scale validation procedure.

As the firefighters and officers proceeded to rate their fellow employees, each individual firefighter was called out of the kitchen into the garage where they were tested on one of the eight different short performance tests. These performance tests served as behavioral validation criteria for the performance appraisl instruments.

Upon completing all of the rating scale forms the firefighters completed two additional surveys. The first was a set of semantic differential attitude scales. Six bipolar adjective items assessed the firefighter's attitude toward the BARS format. The same six items were used to assess the firefighter's opinion toward the MSS format. This attitude survey concluded by asking three general scale preference questions (see Appendix A).

The final questionnaire completed by the firefighters was a self-report criterion survey (see Appendix B). Firefighters answered questions about various other potential performance appraisal scale criteria plus descriptive information.

Officers also completed two additional questionnaires after rating the firefighters under their command. The first questionnaire is the exact same semantic differential attitude survey completed by the firefighters. Officers gave their preference for one of the two scales. The second questionnaire completed by the officers was a Fishbein expectency-value attitude scale. Officers indicated their intention to accurately rate the efficiency of the firefighters under their command. This survey also measured the officer's attitude toward the behavior of accurately rating the firefighters and the department wide subjective norm toward accurate performance ratings. Upon completion of the data collection procedure all participants gathered once more in the kitchen of the station to be debriefed. Any questions about the day's proceedings or the uses of the data collected were answered.

The scoring of the BARS and the MSS are done in different ways. For the BARS format, ratings are made by placing an "X" over a whole number scale value from one to seven. This number reflects the ratee's level of quality on that particular dimension. Nine such dimension scales are filled out. The subscale score is the number (scale value) over which the rater places the "X". The BARS total score is calculated by adding up the nine subscale scores. The scoring of the MSS has been described earlier. Since the nine MSS dimension scores (i.e., subscale scores) are computed by adding three numbers together, while the BARS subscale scores are assigned directly, items on the two formats cannot be directly compared. Instead, the lowest common similarity, the subscale scores, will be directly compared.

Chapter 3

RESULTS

Reliability Analyses

Cronbach's (1970) coefficient alpha reliability is a number that represents the internal consistency of the scale. Internal consistency is the extent to which all of the scale items relate to each other. The alpha reliability of the BARS, based on the ratings given by all officers, is 0.88293. The alpha reliability of the MSS for all officers is 0.83169. Both alpha reliabilities are respectable, with the BARS appearing to have a higher reliability than the MSS. The BARS alpha reliability and item-total statistics are presented in Table 3. These same analyses for the MSS are presented in Table 4. The BARS "item-total correlation" and "alpha if item deleted" columns indicate two bad items: physical fitness and appearance. These same statistics for the MSS indicate one bad item: initiative. Were these items deleted from their respective scales, the alpha reliabilities would actually go up.

Table 3

BARS Alpha Reliability and Item-Total Statistics

	MEAN	VARIANCE	STANDARD DEVIATION	ALPHA
SCALE	48.02747	46.21471	6.79814	.88293
ITEMS	5.33639	.01767	0.1329	.88384 (std. item alpha)

Subscale	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-total *r*	R^2	Alpha if Item Deleted
KOJ	42.692	35.231	0.717	0.642	0.862
WTW	42.676	36.508	0.719	0.562	0.863
PF	42.484	39.599	0.412	0.330	0.888
TAC	42.588	36.807	0.703	0.545	0.864
SC	42.599	36.286	0.704	0.642	0.864
L	42.879	35.322	0.679	0.665	0.866
I	42.852	37.409	0.677	0.497	0.867
A	42.637	39.514	0.430	0.388	0.886
PR	42.813	36.711	0.661	0.520	0.868

SOURCE	*df*	*F*	*p*
Between People	181		
Within People	1456		
Between Measure	8	5.35081	0.0000
Residual	1448		
Non-additivity	1	9.27632	0.00236
Balance	1447		
Total	1637		

Table 4

MSS Alpha Reliability and Item-Total Statistics

	MEAN	VARIABLE	STANDARD DEVIATION	ALPHA
Scale	45.70000	52.84804	7.26967	.83169
Items	5.07778	0.14434	.37992	.83103 (std. item alpha)

Subscale	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total r	R^2	Alpha if Item Deleted
KOJ	40.794	40.723	0.653	0.607	.8009
WTW	40.000	43.597	0.583	0.398	.8109
PF	40.544	41.043	0.702	0.546	.7968
TAC	40.283	45.232	0.408	0.261	.8280
SC	40.827	39.964	0.620	0.616	.8044
L	40.544	42.339	0.584	0.537	.8095
I	40.672	46.724	0.309	0.180	.8377
A	41.372	39.073	0.605	0.445	.8071
PR	40.561	45.019	0.399	0.277	.8293

Source	df	F	P
Between People	179		
Within People	1440		
Between Measures	8	26.289	0.00000
Residual	1432		
New additivity	1	25.772	0.00000
Balance	1431		
Total	1619		

Comparing these two alpha reliabilities, however, does not present the entire picture of the quality of the internal structure of the two formats. Three additional factors must be considered: restriction of range, leniency and halo. These three factors all have an influence on the alpha reliability. They may each serve to artificially inflate the reliability. Even though the BARS has a higher reliability, this effect may be diminished or even changed when these three other error factors are considered.

Restriction of range refers to the fact that the entire range of possible scores on the efficiency mark scales was not used. Theoretically, both formats could have subscale scores ranging from a minimum of 1.0 to a maximum of 7.0. In fact, the BARS scale had a restricted range on seven of the nine subscales. The physical fitness and initiative subscales had a minimum actual score of 3.00. Knowledge of the job, willingness to work, performance under stressful conditions, appearance and public relations all had minimum scores of 2.00. The MSS had a restricted range on five of the nine subscales. Willingness to work, physical fitness, teamwork and compatibility, leadership and initiative all have a minimum score of 2.0, rather than the theoretically possible minimum of 1.0. The maximum of 7.0 is achieved on all subscales of both formats. The theoretical range of total scores is from a minimum of nine to a maximum of sixty-three. The actual minimum of the BARS is twenty-eight, while the minimum MSS total score is twenty-six. Both formats had a maximum total score of sixty-three.

Cronbach (1970) indicates that the alpha reliability is a function of two main factors. The coefficient depends on the spread or distribution of scores and the number of items making up a person's score. The BARS had a more restricted range of scores than the MSS. The smaller the range of possible scores, the more likely the scores will fall on or near the same digit across the subscales. The more likely the scores to be consistent across subscales, the higher the apparent reliability. The second consideration, the number of items making up the score, does not play a part since the reliabilities were calculated on the basis of nine subscale scores for each format. The higher reliability of the BARS may in part be due to the restriction of range of that scale.

Restriction of range also has an influence on the leniency and halo error analysis. Leniency error analysis can be performed with respect to the theoretical definition of leniency or on the basis of the actual distribution of scores. The theoretical definition of leniency, as operationally defined here, is the tendency of a rater to give all ratees a score that is significantly above the midpoint of the scale. A *t-test* was performed between the mean subscale score for each ratee (i.e., the average of the nine subscales for each ratee) and the theoretical midpoint of the average of the subscales (4.00) for each of the two formats. With the knowledge about the restricted range, a large amount of leniency was expected in both scale formats. This was, in fact, the case. The BARS was found to have an almost unbelievably large amount of leniency, t (182) = 1014.68, $p < .001$. The MSS, while still having a large amount of leniency, was not quite as large as the BARS, t (180) = 100.18, $p < .01$. The mean of the mean subscale ratings for the BARS scale was 5.34. When the nine MSS subscales were averaged for each subject and then averaged over these averages, the value was 5.08. A <u>t-test</u> between correlated means was conducted and showed that the BARS did contain significantly more leniency than did the MSS, t (180) = 41.33, $p < .01$ according to the theoretical definition.

Leniency can also be assessed by examining the skewness of the distribution of scores from the two formats. The skewness of a distribution indicates whether or not scores tend to be pulled toward one end of the distribution. Large negative values of skew reflect a general leniency of ratings across all raters within the distribution of scores. Table 5 lists the values of skewness of the nine subscales for both formats. The BARS contained six negatively skewed subscales with significant differences from a non-skewed distribution: knowledge of the job, willingness to work, physical fitness, teamwork and compatibility, leadership, and appearance. The MSS also contained six subscales with a significant negative skew value: willingness to work, physical fitness, teamwork and compatibility, performance under stressful conditions, leadership, and public relations. The two sets of

Table 5
Descriptive Statistics of the BARS and MSS Scales (N = 220)

BARS SUBSCALE	MEAN	STD. DEV.	SKEWNESS	KURTOSIS
KOJ	5.3352	1.1384	-0.5972***	-0.0507 NS
WTW	5.3516	1.0014	-0.3192*	+0.3080 NS
PF	5.5440	1.0954	-0.5535**	-0.3631 NS
TAC	5.4396	0.9884	-0.6979***	+1.5228 ***
PSC	5.4286	1.0421	-0.2816 NS	-0.4499 NS
L	5.1484	1.1776	-0.5997***	+0.5901 NS
I	5.1758	0.9530	+0.0665 NS	-0.2244 NS
A	5.3901	1.0388	-0.4205*	+0.2661 NS
PR	5.2142	1.0423	+0.0335 NS	-0.3735 NS
SUM TOTAL	48.0275	6.7981	-0.122	0.063
ITEM MEAN	5.0778	1.0530		

MSS SUBSCALE	MEAN	STD. DEV.	SKEWNESS	KURTOSIS
KOJ	4.9056	1.2627	-0.2413 NS	-0.1347 NS
WTW	5.7000	1.0565	-1.0986***	1.2886 ***
PF	5.1555	1.1619	-0.3942 *	0.2677 NS
TAC	5.4167	1.1476	-0.3091 *	0.0592 NS
PSC	4.8722	1.3944	-0.4442 **	-0.2397 NS
L	5.1556	1.1951	-0.3442 *	-0.3334 NS
I	5.0278	1.1406	-0.1234 NS	-0.6494 *
A	4.3278	1.5165	-0.2594 NS	-0.4147 NS
PR	5.1389	1.1948	-0.5695 **	1.5781 ***
SUM TOTAL	45.7000	7.2697	-0.380	0.190
ITEM MEAN	5.3364	1.1126		

* $p < .05$
** $p < .01$
*** $p < .001$

subscales seem to be about equally matched on leniency based on the skewness of the distributions.

The skewness of the format total scores was also assessed. Neither scale's final score were significantly different from a non-skewed distribution (BARS = -.122, MSS = -.380). Table 6 presents the mean skew per subject and the mean absolute value of the skew per subject. These numbers indicate the average amount of skewness in a subject's ratings. The mean item skewness of the MSS is -0.0132 and is not significantly different from zero, t (177) = -0.1739, NS. Since this analysis averages skewness across both positive and negative values, a near zero value may occur if large positive and large negative skew values cancel each other out. For this reason, a mean absolute value of skew within a subject's ratings was also calculated. The MSS was found to have a mean absolute value skew of 0.7423, F (1,178) = 14.442, p <.05. The mean skew of the BARS items within a subject was -0.0949, t (172) = -1.406, NS. The mean absolute value skew of the BARS was 0.6049, F (1,173) = 12.147, p <.05. Thus it appears that within a subject, the amount of leniency error is, in effect, canceled out by the amount of

Table 6

Mean, Standard Deviation and Standard Error of the Mean for Combined Subscales by Ratee

	N	MEAN	STD. DEV.	STD. ERROR OF MEAN
MSS SKEW	178	-0.0132	1.0124	0.0759
MSS SKEW on Absolute Values	178	0.7423	0.6863	0.0514
MSS KURTOSIS	178	0.1882	2.3136	0.1734
MSS STD. DEV.	180	0.9756	0.4200	0.0313
BARS SKEW	173	-0.0949	0.8876	0.0675
BARS SKEW ON ABSOLUTE VALUES	173	0.6049	6549	0.0498
BARS KURTOSIS	173	0.0359	2.3577	0.1792
BARS STD. DEV.	180	0.7188	0.3082	0.0230

severity in scoring. The skewness of a ratee's scores is essentially zero.

A t-test between correlated mean skews of the two formats was performed. No difference was discovered between the mean skews of the formats t (170) = 0.96, NS. Since the skewness values may have been canceling out across ratees, a t-test was also conducted on the mean absolute value of skewness of the two scales. The MSS was discovered to have a higher mean absolute value of skew than the BARS, t (170) = 2.04, $p < .05$. The MSS contains more leniency than the BARS utilizing the mean absolute value of skew within a ratee measure (See Table 5).

Halo has been defined as the inability to distinguish between the different levels of performance of an individual across the various dimensions of job performance. In other words, halo effect is the inability for an individual to get different scores on different dimensions. Halo effect, under this definition, is discriminability within an individual, rather than between individuals. The amount of variance present in the subscale scores of an individual is a good measure of the amount of variation in the scores assigned to a person. If not much variance is present in an individual's scores, then halo effect would be present under this definition. The two formats were compared for different amounts of variance. The format with a larger amount of variance has less halo effect. For each format, the variance within each subject's nine subscale ratings was calculated. From this set of within subject variances the mean variance over all subjects within the rating scale format was calculated. A t-test between correlated mean variances of the two formats showed that the MSS had a larger mean variance than the BARS, t (179) = 5.60, p <.001. The MSS contained comparatively less halo effect than the BARS. The standard deviations of the subscales and total scores of the formats are presented in Table 5.

A second method for examining halo effect is by calculating the intercorrelations of the within format subscales. Although these correlations are not solely based on within ratee analysis as is stipulated in the above definition of halo, intercorrelations are an accepted practice (Saal, 1979). Intercorrelations, and other correlational techniques, are indicative of the internal structure of each format. Thus, correlational techniques give a great deal more information than simply halo effect. Campbell and Fiske (1959) outlined an intercorrelational technique for the analysis of the internal structure of tests know as the multitrait-multimethod matrix. This matrix contains the intercorrelations of the nine subscales both within and between rating scale formats (see Table 7).

The within format correlations are considered to be indicative of halo effect by Saal (1979). Saal states that "High correlations, reflecting strong linear relationships, are suggestive of halo. Low correlations, which indicate that the ratings on one dimension are not related to the ratings on a second dimension, suggest the absence of halo" (Saal, 1979, p. 21). Campbell and Fiske (1959) consider these intraformat correlations to be method variance. Method variance, which contains the concept of halo error, shows the amount of variance within a set of evaluation test scores that is due to the format of the test, rather than due to useful variance in actual performance (Brown, 1976). All of the intercorrelations of the BARS

Table 7

BARS and MSS Intercorrelation Matrix:

A). MSS Intercorrelation Matrix

	KOJ	WTW	PF	TAC	PSC	L	I	A	PR
KOJ	1.0								
WTW	.3932	1.0							
PF	.5584	.5434	1.0						
TAC	.2972	.2281	.3282	1.0					
PSC	.7197	.4365	.5296	.2115	1.0				
L	.6355	.4531	.4090	.3069	.6590	1.0			
I	.1570	.2712	.2792	.0551	.1919	.1157	1.0		
A	.3693	.4034	.5701	.3865	.3501	.3200	.3726	1.0	
PR	.2198	.3253	.3626	.3650	.1314	.1295	.2267	.4095	1.0
ITEM - TOTAL	.7473	.6749	.7787	.5354	.7314	.6870	.4477	.7291	.5329

Table 7 (continued)

BARS and MSS Intercorrelation Matrix:

B). MSS By BARS Intercorrelation Matrix

BARS	MSS									
	KOJ	WTW	PF	TAC	PSC	L	I	A	PR	ITEM - TOTAL
KOJ	.7352	.3918	.4694	.2401	.7092	.5842	.1829	.3456	.1182	.2924
WTW	.4777	.4449	.5768	.2059	.4731	.4020	.2067	.4566	.3510	.3252
PF	.2085	.1048	.1668	.5470	.2010	.2500	.0981	.4057	.2132	.1415
TAC	.4300	.5472	.4879	.1476	.4525	.4084	.3494	.4235	.2875	.2813
PSC	.6132	.3472	.3733	.1659	.6384	.6232	.1873	.3517	.0636	.3148
L	.6376	.4341	.4347	.1236	.7269	.5893	.3127	.3268	.1428	.3954
I	.4072	.3325	.4652	.2576	.3646	.3005	.3550	.6222	.3030	.2006
A	.1422	.1101	.2180	.2113	.1300	.1267	.2541	.3724	.5760	.1210
PR	.3323	.2850	.3129	.0799	.3862	.2865	.5474	.3364	.2177	.2002
ITEM - TOTAL	.6190	.4604	.5370	.2986	.6362	.5528	.3811	.5538	.3437	.9129

Table 7 (continued)

BARS and MSS Intercorrelation Matrix:

C.) BARS Intercorrelation Matrix

	KOJ	WTW	PF	TAC	PSC	L	I	A	PR
KOJ	1.0								
WTW	.5649	1.0							
PF	.2969	.3343	1.0						
TAC	.5361	.6523	.2875	1.0					
PSC	.7072	.5801	.3782	.5455	1.0				
L	.7375	.5740	.1520	.5607	.6907	1.0			
I	.5158	.5485	.4082	.4864	.5023	.4493	1.0		
A	.2159	.3879	.3985	.3971	.1509	.1873	.4215	1.0	
PR	.5025	.4779	.2491	.5677	.4491	.5411	.5793	.4582	1.0
ITEM - TOTAL	.7934	.7866	.5372	.7727	.7774	.7669	.7497	.5508	.7430

subscales are significant at an alpha of .05. For the MSS, all correlations are similarly significant, except for those between initiative, and teamwork and compatibility; and between initiative, and leadership. The mean interitem correlation for the BARS is 0.4581, the mean interitem correlation for the MSS is 0.3534. A t-test between the mean correlations indicate that the BARS contains a significantly higher mean intercorrelation than the MSS, t (35) = 2.79, $p < .01$. The BARS scale appears to have more halo error than the MSS using the correlational definition. The BARS higher mean intercorrelation is also reflected in that format's internal consistency (alpha) reliability.

The two intraformat correlation matrixes are also described as the method variance triangles in the multitrait-multimethod matrix (see Table 7). By adding the interformat correlation matrix of the nine BARS subscales with the nine MSS subscales, the entire multitrait-multimethod matrix is formed. The two off-diagonal triangles within the interformat correlation matrix form the discriminant validity correlations. These are the correlations between separate traits measured by different methods. The main diagonal between the discriminant validity triangles contain the convergent validity correlations. These correlations measure the relationship between the same trait as measured by the two different formats.

The major use of the multitrait-multimethod matrix is to assess convergent and discriminant validity. When the correlations on the validity diagonal are larger than the correlations in the method variance triangles and larger than the correlations involving different traits and different methods, then discriminant validity is achieved. The validity correlation for knowledge of the job is larger than all other correlations for MSS and is surpassed by only one correlation for the BARS. Performance under stressful conditions and leadership also display a high amount of discriminant validity for both formats. Willingness to work, initiative and appearance seem to surpass the off diagonal correlations for about half of the cases or more with MSS doing slightly better. The other subscales (i.e., physical fitness, teamwork and compatibility and public relations) appear to have more method variance and other artifactual variance rather than discriminant validity. One caveat should be mentioned in interpreting the discriminant validity of the scales. Even though the methods were different, the BARS and MSS scales had a substantial overlap in the content and meaning of the items. The methods were not all that different. Thus, discriminant validity may have been artificially inflated.

Since all of the validity diagonal correlations were statistically significant, both scales appear to achieve convergent validity. Discriminant validity appears to be achieved on both formats for three of the nine subscales (knowledge of the job, performance under stressful conditions and leadership).

A factor analysis was performed separately on the subscales of the two formats. Results of the factor analysis lend support to the conclusions based on the multitrait-multimethod matrix. Two factors were found to best explain the data for both the BARS and the MSS. Factors were considered to be significant if and only if their eigenvalve was 1.0 or greater. The two factors in the MSS accounted for a total of 58.6% of the variance. The two factors in the BARS accounted for a total of 66.3%.

The factor analysis on the MSS appears in Table 8. Three subscales load heavily on factor I, knowledge of the job, performance under stressful conditions and leadership. The remaining six subscales all load higher on the second factor. Factor I explains 80.5% of the variance in the two factors. For the BARS factor analysis (see Table 9) the same three subscales load heavily on factor I (knowledge of the job, performance under stressful conditions, and leadership). In addition, willingness to work and teamwork and compatibility also load on factor I, but not nearly as heavily. Factor I accounts for 84.0% of the explained variance. In both formats, four subscales load on factor II: appearance, physical fitness, initiative and public relations. It appears that the three subscales with the highest convergent and discriminant validity for both formats also compose the first factor of the factor analyses. From these factor analyses one may conclude that despite substantial differences in the factor structure of the two formats, the BARS and MSS did have quite similar loadings on the first and largest factor.

Table 8

Factor Analysis Summary of the MSS Subscales

FACTOR	EIGENVALUE	PERCENT OF VARIANCE	CUMULATIVE PERCENT OF VARIANCE	EIGENVALUE	PERCENT OF VARIANCE
1	3.9386	43.8	43.8	3.49967	80.5
2	1.3355	14.8	58.6	0.84663	19.5
3	0.9905	11.0	69.6		
4	0.6684	7.4	77.0		
5	0.5615	6.2	83.3		
6	0.5579	6.2	89.5		
7	0.4048	4.5	94.0		
8	0.2800	3.1	97.1		
9	0.2627	2.9	100.0		

VARIMAX ROTATED FACTOR MATRIX

SUBSCALE	FACTOR 1	FACTOR 2	COMMUNALITY
KOJ	.78275**	0.25727	.67888
WTW	.41686	.46960**	.39429
PF	.47941	.61299**	.60558
TAC	.20752	.41433**	.21474
PSC	.84578**	.17990	.74771
L	.75672**	.17953	.60485
I	.09582	.39003**	.16130
A	.24488	.71955**	.57772
PR	.04188	.59957**	.36124

** - highest loading of subscale

Table 9

Factor Analysis Summary of the BARS Subscales

FACTOR	EIGENVALUE	PERCENT OF VARIANCE	CUMULATIVE PERCENT	EIGENVALUE	PERCENT OF VARIANCE
1	4.73552	52.6	52.6	4.34046	84.0
2	1.23540	13.7	66.3	0.82803	16.0
3	0.81673	9.1	75.4		
4	0.56316	6.3	81.7		
5	0.46969	5.2	86.9		
6	0.40882	4.5	91.4		
7	0.28915	3.2	94.6		
8	0.26072	2.9	97.5		
9	0.22079	2.5	100.0		

VARIMAX ROTATED FACTOR MATRIX

SUBSCALE	FACTOR 1	FACTOR 2	COMMUNALITY
KOJ	.82703**	.22384	.73408
WTW	.59212**	.48317	.58406
PF	.19251	.45818**	.24699
TAC	.56344**	.48340	.55113
PSC	.79518**	.21225	.67736
L	.84297**	.15669	.73516
I	.45509	.56898**	.53084
A	.02276	.77900**	.60736
PR	.47154	.52837**	.50152

** - highest loading of subscale

Sensitivity Analyses

Sensitivity or central tendency was defined earlier as the ability to distinguish between individuals on the basis of their subscale scores. Two approaches to establishing the differential sensitivity of the formats were used. First, the kurtosis of the subscale scores and the total score distributions were calculated. Kurtosis indicates a narrow and peaked distribution of scores that will not be very powerful in distinguishing between ratees. A negative kurtosis indicates a flat distribution of scores that can more easily distinguish between ratees.

The kurtosis of the nine subscales and the total score for each format are listed in Table 5. The kurtosis of the BARS total score and eight of the nine BARS subscales are all not significantly different from zero. The teamwork and compatibility subscale does have a significant kurtosis (kurtosis = 1.5228; $p < .001$). Teamwork and compatibility on the BARS is not sensitive to the differences between individuals.

For the MSS, willingness to work and public relations both have significant positive kurtosis (willingness to work = 1.289; $p < .001$; physical fitness = 1.578; $p < .001$). Neither of these subscales is sensitive to differences between ratees. However, initiative on the MSS has a significant negative kurtosis (Initiative = -0.649, $p < .05$). Initiative appears to be sensitive to fine differences between ratees.

A kurtosis was calculated for the distribution of each subject's scores along with the mean kurtosis across all subjects and within format subscales. The mean kurtosis for the MSS was 0.188 and was not significantly different from zero, $t(177) = 1.085$, NS. The mean kurtosis for the BARS was 0.0359 and was not significant, $t(172) = 0.200$, NS. On the average, neither scale provided platykurtic distributions that could easily discriminate between ratees. However, examining the skewness and kurtosis of both scales points out that the distributions of scores appear to fit a unit normal distribution and provide an adequate amount of sensitivity. A second interpretation of these non-significant kurtoses is that low discriminability reflects the fact that in reality there are few true differences between the ratees. The ratees may not be all that different in the skills measured by these performance appraisal instruments.

The second method for assessing sensitivity is by the analyses of the standard deviations (or variances) of the subscales across all ratees. T-test for the difference between two correlated variances were computed for each of the nine subscales between the two formats (cf. Glass and Stanley, 1970, p. 306). Four of the nine sets of variances were not significantly different from each other (i.e., willingness to work, physical fitness, leadership, and public relations, see Table 10). These subscale were equally sensitive to differences between ratees on both formats. However, the MSS was significantly more sensitive (i.e., has higher variance) on each of the remaining five subscales (i.e., knowledge of the job, t (178) = 2.255, $p < .05$; Teamwork and compatibility, t (178) = 5.101, $p < .01$; initiative, t (178) = 2.592, $p < .05$; appearance, t (178) = 5.566, $p < .01$). Thus, on the second measure of sensitivity, the MSS appears more sensitive than the BARS on five out of nine subscales and equally sensitive

Table 10

Summary of the Sensitivity t-tests for differences between mean variances of the BARS and MSS Subscales

Subscale	BARS Variance	MSS Variance	t-test	Probability	Scale of Greater Sensitivity
KOJ	1.268	1.595	2.255	.05	MSS
WTW	1.002	1.115	0.796	NS	-
PF	1.199	1.350	0.803	NS	-
TAC	0.976	1.318	2.036	.05	MSS
PSC	1.086	1.943	5.101	.01	MSS
L	1.388	1.428	0.234	NS	-
I	0.908	1.302	2.592	.05	MSS
A	1.079	2.298	5.566	.01	MSS
PR	1.086	1.428	1.879	NS	-

on the other four. However, a t-test on the difference between the variances of the total scores shows no difference between the BARS and the MSS on sensitivity, t (179) = 0.6789, NS. A summary of the reliability and sensitivity analyses is presented in Table 11. The kurtosis and variance sensitivity analyses are consistent with the restriction of range analyses. These three measures of the distribution of scores on the two formats show both the BARS and MSS containing adequate amounts of sensitivity with the MSS format displaying a slightly better amount.

Validity Analyses

Criterion validity data were collected on five dependent variables. These five dependent variables were (1) the rank order of the firefighter on a pre-established, but not yet published Fire Lieutenant promotional examination, (2) the amount of seniority of a firefighter as measured by his file number, (3) a composite variable indicating the educational background of the firefighter, (4) a composite variable measuring the job related experience of the firefighter, and (5) the standardized scores on the short, practical performance test. Each dependent variable was regressed

Table 11

Summary of Reliability and Sensitivity Analyses on the BARS and the MSS.

		MSS	BARS	SCALE JUDGED TO BE SUPERIOR
1.	Alpha reliability	.83169	.88293	BARS
2.	Restriction of Range-items	5 of 9	7 of 9	MSS
	Total score range restriction	26-63 points	28-63 points	
3.	Leniency			
	T-test of scale mean and midpoint	5.07778	5.33639	Neither
	T-test of MSS vs. BARS means	45.7000	48.0275	MSS
	Skewness of subscales	6 neg.,1 pos	6 neg.	MSS
	Total score skewness	-0.380 (NS)	-0.122 (NS)	Neither
	Mean item skew per subject	-0.0132(NS)	-0.0949(NS)	Neither
	Mean absolute value skew per subject	0.7423(.05)	0.6049(.05)	BARS
4.	Halo			
	Mean Variances	.9756	.7188	MSS
	Mean interitem correlations	.35337	.45812	MSS
	Factor Analysis (% of Variance)	58.6%	66.3%	
5.	Sensitivity			
	Kurtosis of Subscales	2 bad, 1 good	1 of 9 bad	Neither
	Mean Kurtosis	.188(NS)	.0359(NS)	Neither
	Standard Deviations	5 superior	none superior	MSS
	Total Score Variances	7.2692	6.79814	Neither
6.	Totals on Superiority Judgments	6 MSS	2 BARS	6 Neither

on each of the two format's total score and on each format's nine subscales. Each dependent variable was also analyzed separately for lieutenant raters, captain raters and combined lieutenant and captain raters.

Promotional Rank One of the major purposes of the performance evaluation scale is for promotions. A large group of firefighers in the sample (N=83) had taken the Fire Lieutenant promotional examination several months previous to the performance evaluation project. None of these firefighters, nor their commanding officers, knew the rank order of the list nor any of the promotional examination scores. A summary of the promotional rank order criterion regressions is presented in Appendix J.

The total scores on the BARS and the MSS were calculated and regressed on the firefighter rank order on the Lieutenant Promotional examination. The BARS total score could account for 14.09% of the variance in the promotional rank, F (1, 78) = 13.451, p <.01. The MSS total score accounted for 21.83% of the variance in promotional rank, F (1, 78) = 22.62, p <.001. Both rating scale formats significantly predicted the promotional rank of the firefighters. The MSS accounted for more variance than the BARS in promotional rank.

The effect of the BARS total score and the MSS total score were also analyzed separately for lieutenant and captain raters. Lieutenant total score ratings of firefighters on the BARS could explain 22.10% of the variance in promotional rank, F (1, 59) = 16.737, $p <$ .001. Lieutenant's ratings using MSS total scores explained 22.55% of the variance in promotional rank, F (1, 59) = 17.183, $p <$ .01. Captain's ratings on the BARS could explain a nonsignificant 1.62% of the variance in promotional rank, F (1, 20) = 1.329, NS. Captain's ratings on the MSS explained 19.36% of the variance in promotional rank, but was not statistically significant, F (1, 20) = 4.80, $p <$ .1. The lack of significance of the captain's MSS ratings may be due, in part, to the small sample size for captain.

Using the nine BARS subscales from all raters (i.e., lieutenants and captains) on promotional rank, a total of 26.223% of the variance was accounted for, F (9, 73) = 2.883, $p <$ .001. Utilizing a hierarchical regression with a stepwise inclusion level based on the greatest accounted for variance (hereafter referred to as "stepwise"), the only subscale to account for a significant portion of the promotional rank based on the

BARS is initiative (R^2= .16552, F (1, 73) = 16.378, p < .001). Entering the initiative variable last in a hierarchical regression (hereater referred to as "hierarchical") shows that the unique variance accounted for is 4.572%, F (1, 73) = 5.1435, p < .01. Of the nine BARS subscales based on all raters, only initiative accounts for a significant portion of unique variance in promotional rank.

The nine MSS subscales based on all raters could predict 29.695% of the variance in the promotional rank order of the firefighters, F (9, 73) = 3.426, p < .001. Analyzing the nine subscales individually in a stepwise regression, appearance accounts for the only statistically significant portion of the variance, F (1, 73) = 23.49, p < .001, while knowledge of the job approaches significance, F (1, 73) = 4.669, p <.1. Analyzed by loading appearance hierarchically last in the regression equation, a total of 8.172% of the unique variance in promotional rank can be accounted for F (1, 73) = 8.485, p <.01. Knowledge of the job, F (1, 73) = 1.38, NS, and initiative, F (1, 73) = 1.56, NS each accounted for a little over 1% of the variance. Thus, appearance accounts for the most unique variance in promotional rank based on the nine MSS ratings of all ranks of raters.

Analyzed separately for lieutenant raters (N = 63), the nine BARS subscales could account for a total of 36.212% of the variance in promotional rank, F (9, 51) = 3.216, p <.01. The initiative subscale is the only subscale that accounts for a significant portion of the variance in a stepwise regression, F (1, 51) = 16.675, p < .001. However, initiative can only account for a nonsignificant 1.747% of the variance uniquely, F (1, 51) = 1.397, NS. Physical fitness uniquely accounts for 3.35% of the variance, F (1, 51) = 2.697, NS; teamwork and compatibility uniquely accounts for 2.22% of the variance, F (1, 51) = 1.791, NS; and leadership uniquely accounts for 2.03% of the variance in promotional rank, F (1, 51) = 1.623, NS. None of these subscales accounts for significant portions of the variance, yet they all account for more unique variance than initiative.

The nine MSS subscales utilized by lieutenant raters in predicting promotional rank accounted for 35.07% of the variance in promotional rank, F (9, 51) = 3.06, p <.01. The appearance subscale accounted for the only significant portion of the variance in a stepwise hierarchical regression. Appearance entered the regression equation first and accounted for 27.71% of the total variance, F (1, 51) = 21.77, p <.01. The MSS subscale, knowledge of the job, loaded second in the stepwise herarchical regression for lieutenants, accounting for an additional 4.27% of the variance, F (1, 51) = 3.35, p <.2. Other than appearance, none of the remaining eight subscales could account for a significant stepwise portion of the variance. The appearance MSS subscale, when loaded last in a hierarchical regression uniquely accounted for 12.04% of the variance in promotional rank, F (1, 51) = 9.45, p <.01. The MSS knowledge of the job subscale uniquely accounts for a nonsignificant 2.38% of the variance, F (1, 51) = 1.87, NS. Thus, the MSS appearance subscale accounts for most of the variance in promotional rank for lieutenant raters.

For captain raters, the nine BARS subscales account for a total of 33.05% of the variance, F (9, 12) = 0.659, NS, while the nine MSS subscales account for 31.30% of the variance, F (9, 12) = 0.6076, NS. Since the number of captains involved in this portion of the validity study was small (N = 22) and the number of predictor variables large (nine subscales), the percent of variance accounted for is artificially high. None of the stepwise hierarchical regressions, nor the unique percentages of variance accounted for by individual subscales, approaches acceptable significance levels.

In any case, the physical fitness subscale of the MSS loaded first in the stepwise regression on promotional rank with captain raters and accounted for 17.72% of the variance, F (9, 12) = 3.096, NS. The MSS public relations subscale accounted for the next smaller amount of variance, 4.53%, F (9, 12) = 0.79, NS. The MSS physical fitness subscale uniquely accounted for 0.001% of the variance in promotional rank, F (1, 10) = 0.009, NS. The MSS public relations subscale for captains accounts for 5.06% of the variance in promotional rank, F (1, 10) = 0.737, NS. Willingness to work accounts for 7.39% of the variance in captain's ratings on MSS, the largest single amount for the nine subscales, F (1, 10) = 1.076, NS.

The physical fitness subscale of the BARS ratings by captains uniquely accounted for 13.29% of the variance in promotional rank, F (1, 10) = 1.985, NS. Physical fitness loaded second, behind the performance under stresful conditions subscale on the stepwise hierarchical regression. In the stepwise regression, physical fitness accounted for an additional 6.39% of the variance in promotional rank for captain raters, F (9, 12) = 1.14, NS. The performance under stressful conditions BARS subscale accounted for 8.66% of the variance when entered first, F (9, 12) = 1.55, NS, but uniquely could explain only 0.55% of the variance in promotional rank, F (1, 10) = 0.08, NS.

In summary, a statistically significant portion of the promotional rank order criterion validity measure was predicted by both the BARS and the MSS formats. The MSS predicted a substantially greater portion of the criterion than the BARS. Thus, the MSS appears to have achieved a greater criterion validity than the BARS on the promotional rank criterion. Furthermore, lieutenant raters appeared to have made ratings that accounted for more variance in the criterion than did captain raters.

Lieutenants are judged to have a higher promotional rank criterion validity than captains.

Seniority The amount of seniority that a firefighter has is an important criterion variable for the validation of performance scores. Most firefighters prefer to believe that there performance ability steadily rises throughout their career until their personal maximum is reached. However, a more likely possibility is a quadratic relationship between seniority and performance, rather than the linear one suggested above. A firefighter's performance ability, and therefore that firefighter's performance evaluation scores may be at their lowest point for new recruits. Scores would rise to a maximum sometime in mid-career and then tail-off as the firefighter gets older. All seniority criterion regressions are summarized in Appendix K.

Seniority was measured by the firefighter's file number. File numbers are assigned in consecutive order with lower numbers indicating more time on the job. Thus, file numbers represent an ordinal level measure of length of service.

The linear analyses were conducted first. The BARS total scores based on the ratings of all officers could account for 0.846% of the variance in seniority, *F* (1, 178) = 1.536, NS. The MSS total score based on the ratings of all officers accounted for 0.767% of the variance in seniority, *F* (1, 178) = 1.375, NS. Neither the total score on the BARS nor the MSS can successfully predict a firefighter's seniority when based on the ratings of all officers.

When seniority is analyzed separately for the ratings of the lieutenants, the BARS can account for 0.916% of the variance *F* (1, 119) = 1.0098, NS. The MSS total scores based on the lieutenant ratings could account for 0.964% of the variance in seniority, *F* (1, 119) = 1.158, NS. Neither of the format's total scores assigned by lieutenants could significantly predict seniority. These same findings hold true for captain raters. The captain's BARS total scores could account for 0.791% of the variance in seniority, *F* (1, 56) = 0.446, NS. The MSS total scores based on the captains' ratings accounted for 0.032% of the variance in seniority, *F* (1, 55) = 0.017, NS.

When the BARS ratings for all raters are analyzed using the nine subscales as separate independent variables, a total of 26.998% of the variance in seniority is accounted for, $\underline{F}$ (9, 156) = 6.410, $\underline{p}$ <.01. When analyzed for the individual BARS subscales for all raters, leadership, physical fitness and knowledge of the job each account for significant portions of the seniority variance. Utilizing a stepwise hierarchical regression, leadership entered the prediction equation first and accounted for 8.875% of the variance, $\underline{F}$ (1, 156) = 18.965, $\underline{p}$ < .01. Physical fitness entered the prediction equation second and accounted for 10.43% of the variance in seniority, $\underline{F}$ (1, 156) = 22.30, $\underline{p}$ < .01. Knowledge of the job entered the prediction equation third and accounted for 3.226% of the seniority variance, $\underline{F}$ (1, 156) = 6.894, $\underline{p}$ < .01. Analyzed for unique variance, however, only physical fitness could account for a significant portion of the variance, $\underline{R}^2$= 0.10257, $\underline{F}$ (1, 156) = 21.9185, $\underline{p}$ <.01.

The nine MSS subscales based on all officer ratings accounts for 7.946% of the variance in seniority, $\underline{F}$ (9,156) = 1.496, NS. Performance under stressful conditions, teamwork and compatibility, and public relations entered the prediction equation first, second and third, respectively. None of these subscales accounted for a significant portion of the variance, either stepwise or uniquely. Performance under stressful conditions accounted for 1.733% of the variance in seniority when entered first in the prediction equation, $\underline{F}$ (1, 156) = 2.0168, NS and uniquely accounted for 1.258% of the variance, $\underline{F}$ (1, 168) = 2.29, NS. Teamwork and compatibility accounted for 1.582% of the variance when entered second in the prediction equation, $\underline{F}$ (1, 156) = 2.68, NS, and uniquely accounted for 2.687% of the variance, $\underline{F}$ (1, 168) 4.90, NS. Public relations entered the prediction equation third and accounted for 2.478% of the variance, $\underline{F}$ (1, 156) = 4.199, NS, while uniquely accounting for 2.144% of the variance, $\underline{F}$ (1, 168) = 3.913, NS.

When the seniority data are analyzed separately for the lieutenant ratings, the nine BARS subscales accounted for 33.77% of the variance, $\underline{F}$ (9, 99) = 5.6088, $\underline{p} < .01$. Knowledge of the job entered the prediction equation first and accounted for 10.62% of the variance, $\underline{F}$ (1, 99) = 15.87, $\underline{p} < .01$. Of this variance, knowledge of the job uniquely accounted for 5.88%, F (1, 99) = 8.79, p < .01. Physical fitness entered the prediction equation second and accounted for an additional 11.559% of the variance, F (1, 99) = 17.278, p <.01. Physical fitness uniquely accounted for 8.852% of the variance, F (1, 99) = 13.23, p <.01.

The MSS for lieutenant ratings accounted for 9.335% of the variance when utilizing the nine subscales as predictor variables, F (9, 99) = 1.13, NS. None of the subscales accounted for a significant portion of the variance. Physical fitness entered the prediction equation first and accounted for 3.108% of the variance, F (1, 99) = 3.39, NS. Physical fitness uniquely accounted for 1.35% of the seniority variance, F (1, 111) = 1.65, NS. Teamwork and compatibility entered the equation second, accounting for an additional 2.35% of the variance, F (1, 99) = 2.568, NS, while uniquely accounting for 3.092% of the variance, F (1, 111) = 3.78, NS.

Analyzing the nine BARS subscales for captain raters a total of 28.276% of the variance was accounted for, F (9, 47) = 2.0588, NS. Although the percentage of variance accounted for by captain ratings is comparable to that of the lieutenant BARS ratings, the smaller number of captain raters didn't allow sufficient degrees of freedom for statistical significance. Physical fitness entered the prediction equation first and accounted for 11.764% of the variance, F (1, 47) = 7.7088, p < .05. Performance under stressful conditions entered the equation second and accounted for 11.017% of the variance, F (1, 47) = 7.2193, p < .01. Uniquely, the performance under stressful conditions variable accounted for 1.304% of the variance, F (1, 47) = 0.8545, NS.

The nine MSS subscales utilized by captain raters accounted for 9.942% of the variance in seniority, F (9, 47) = 0.5766, NS. None of the nine subscales accounted for a significant portion of the variance. Performance under stressful conditions uniquely accounted for 5.011% of the variance, F (1, 47) = 2.6152, NS. Appearance uniquely accounted for 1.416% of the variance, F (1, 47) = 0.739, NS. Public relations entered the prediction equation third and accounted for 2.513% of the variance, F (1, 47) = 1.3115, NS. Public relations uniquely accounted for 2.257% of the variance, F (1, 47) = 1.1779, NS.

In order to check for this possibility, polynomial contrasts were run using the seniority data as an independant variable checking for linear, quadratic and cubic trends. Oneway analyses of variance with Newman-Keuls post-hoc analyses were also performed. The seniority variable was recorded into six groups of thirty firefighters each. The first group of firefighters had the least amount of seniority; the sixth group, the most seniority. Although the exact age breakdown of these six groups was not calculated, roughly speaking the groups corresponded to intervals of four years of experience beginning with the first group having one through four years experience and ending with the sixth group having twenty-one through twenty-four years experience.

The BARS total score contained statistically significant linear and quadratic trends. The mean score of each group for these trends and the results from the remaining polynomial trend analyses are presented in Table 12. A total of 3.058% of the recoded seniority variable's variance could be accounted for by the BARS total score, $\underline{F}$ (1, 178) = 5.61, $\underline{p}$ <.01. Significant differences between group one ($\bar{X}$ = 44.44) and groups three ($\bar{X}$ = 49.93), four ($\bar{X}$ = 49.34) and five ($\bar{X}$ = 50.11) were discovered. The remaining groups showed no difference from any of the groups. A closer analysis shows a statistically significant increasing linear trend over seniority, F (1, 162) = 10.84, p < .01. In genéral, the more seniority a firefighter has, the higher the firefighter's BARS total score. However, a significant quadratic trend also fit the BARS total score data, F (1, 162) = 6.583, p < .02. Mean BARS performance appraisal total scores rose from a low in group one ($\bar{X}$ = 44.44) to a high in group five (X = 50.11) and then drops off slightly in group six (X = 48.40).

The MSS total score distribution is also linearly and quadratically related to the recoded seniority variable. Although in the simple regression the MSS could only account for a non-significant 1.90% of the variance, F (1, 178) = 3.44, NS, the polynomial trend analysis of variance did reach significance F (5, 160) = 2.91, p <.02. The linear trend had higher F ratios than the quadratic trends, linear F (1, 165) = 7.548, p <.01; quadratic F (1, 160) = 4.96, p <.03. A general increase in MSS total scores across seniority occured. Groups one ($\bar{X}$ = 42.74) and four ($\bar{X}$ = 48.21) were significantly different from one another. The significant quadratic trend showed that performance appraisal scores tended to rise from a minimum in group one to a maximum in group four, then fall off in group five ($\bar{X}$ = 47.96) and group six ($\bar{X}$ = 45.57). Both the BARS and MSS total score distributions support the linear and quadratic explanations of the score distributions.

Polynomial trend analyses of variance were conducted on each of the nine subscales for the BARS and MSS in order to assess which of the subscales were contributing to these effects. The results are presented in Table 12. For the BARS scale, the following subscales displayed statistically significant linear and quadratic trends with the highest ratings generally occuring in the fourth or fifth groups with lower scores on either side: knowledge of the job, physical fitness, performance under stressful conditions and leadership. The BARS willingness to work subscale

Table 12

Group Means of Significant ANOVA Trend Analyses for All Raters

Seniority Criterion Variable Subscales

1. Total Score

MSS

Group:	1	2	3	4	5	6
Mean:	42.74	43.22	46.07	48.21	47.96	45.57

Trends: Linear, Quadratic

Similar Groups*: I: 1,2,3,5,6
II: 2,3,4,5,6

BARS

Group:	1	2	3	4	5	6
Mean:	44.44	45.32	49.93	49.39	50.11	48.40

Trends: Linear, Quadratic

Similar Groups*: I: 1,2,6
II: 2,3,4,5,6

2. Knowledge of the Job

MSS

Group:	1	2	3	4	5	6
Mean:	4.11	4.15	5.17	5.39	5.64	4.97

Trends: Linear, Quadratic

Similar Groups*: I: 1,2
II: 3,4,5,6

BARS

Group:	1	2	3	4	5	6
Mean:	4.33	4.89	5.55	5.64	5.81	5.70

Trends: Linear, Quadratic

Similar Groups*: I: 1
II: 2
III: 3,4,5,6

3. Willingness to Work

MSS

(NS)

BARS

Group:	1	2	3	4	5	6
Mean:	4.93	4.75	5.62	5.43	5.50	5.67

Trends: Linear

Similar Groups*: I: 1,2
II: 1,4,5
III: 3,4,5,6

4. Physical Fitness

MSS

(NS)

BARS

Group:	1	2	3	4	5	6
Mean:	5.67	5.86	5.86	5.50	5.50	4.90

Trends: Linear, Quadratic

Similar Groups*: I: 1,2,3,4,5
II: 4.5,6

Table 12 (continued)

5. Teamwork and Compatibility

MSS							BARS
Group:	1	2	3	4	5	6	
Mean:	5.81	5.63	5.31	5.54	5.54	4.70	(NS)

Trends: Linear, Cubic

Similar Groups*: I: 1,2,3,4,5

II: 6

6. Performance Under Stressful Conditions

MSS							BARS						
Group:	1	2	3	4	5	6	Group:	1	2	3	4	5	6
Mean:	3.41	4.41	5.28	5.32	5.24	5.50	Mean:	4.81	5.03	5.41	5.89	5.77	5.60

MSS — Trends: Linear, Quadratic

Similar Groups*: I: 1

II: 2

III: 3,4,5,6

BARS — Trends: Linear, Quadratic

Similar Groups*: I: 1,2,3

II: 2,3,6

III: 3,4,5,6

7. Leadership

MSS							BARS						
Group:	1	2	3	4	5	6	Group:	1	2	3	4	5	6
Mean:	4.48	4.85	5.21	5.64	5.52	5.30	Mean:	4.22	4.53	5.45	5.61	5.42	5.63

MSS — Trends: Linear, Quadratic

Similar Groups*: I: 1,2

II: 2,3,4,5,6

BARS — Trends: Linear, Quadratic

Similar Groups*: I: 1,2

II: 3,4,5,6

* Similar Groups - groups listed after a roman numeral have mean scores not significantly different from each other based on a Neuman-Keuls statistic following an ANOVA.

showed only a significant linear trend that increases from group one ($\overline{X}$ = 4.93) through group six ($\overline{X}$ = 5.67).

For the MSS scales, linear and quadratic trends of the same low-high-low scoring pattern occur for the following subscales: knowledge of the job, performance under stressful conditions and leadership. A linear and cubic significant trend occured for the teamwork and compatibility subscale. The linear trend began at its highest point in the compatibility subscale. The linear trend began at its highest point in group one ($\overline{X}$ = 5.81) and reached its lowest point at group six ($\overline{X}$ = 4.70). The cubic trend is due to a rise in the performance appraisal scores in groups four and five over the third and sixth group scores. None of the remaining subscales showed significant polynomial trends.

When the categorical seniority data was analyzed for polynomial trends for lieutenant and captain raters separately, a very interesting result occured--lieutenant raters appear to be responsible for the polynomial trends discovered in the data. The captain rater data across both formats contained only one significant polynomial trend.

For the lieutenant raters, the BARS total score categories predicted a significant portion of the seniority variance, F (5, 103) = 3.773, $p < .01$. The lieutenant BARS total score ratings contained both a significant linear trend, F (1, 103) = 7.244, $p < .01$ and a significant quadratic trend, F (1, 103) = 7.854, $p < .01$. The linear trend shows a general increase in performance appraisal scores from the group with the lowest amount of seniority to the group with the highest amount. The quadratic trend shows generally low performance appraisal scores in the extreme groups and the highest score for group three in the middle (see Table 13).

The lieutenant MSS total scores did not reach acceptable levels of statistical significance, F (5, 103) = 2.063, $p = .0761$, and therefore, contained no stable polynomial trends.

Neither the captain rater BARS total scores nor the captain rater MSS total scores, when analyzed as categorical data, reached statistical significance, BARS F (5, 51) = 1.287, NS; MSS F (5, 50) = 1.002; NS. Neither format's seniority criterion total score

Table 13

Group Means of Significant ANOVA Trend Analyses for Lieutenant Rater Seniority Criterion Variable Subscales

1. Total Score

MSS: (NS)

BARS

Group:	1	2	3	4	5	6
Mean:	44.07	45.37	51.17	50.33	49.28	49.00

Significant Trends: Linear, Quadratic

Similar Groups*: I: 1,2,5,6
II: 2,4,5,6
III: 3,4,5,6

2. Knowledge of the Job

MSS

Group:	1	2	3	4	5	6
Mean:	4.00	4.12	5.33	5.33	5.61	5.00

Trends: Linear, Quadratic

Similar Groups*: I: 1,2
II: 3,4,5,6

BARS

Group:	1	2	3	4	5	6
Mean:	4.21	4.94	5.78	5.78	5.71	5.77

Trends: Linear, Quadratic

Similar Groups*: I: 1
II: 2,3,4,5,6

3. Willingness to Work

MSS: (NS)

BARS

Group:	1	2	3	4	5	6
Mean:	5.07	4.75	5.67	5.44	5.48	5.68

Trends: Linear

Similar Groups*: I: 1,2,4,5
II: 1,3,4,5,6

4. Physical Fitness

MSS: (NS)

BARS

Group:	1	2	3	4	5	6
Mean:	5.64	5.87	5.94	5.72	5.47	5.00

Trends: Linear, Quadratic

Similar Groups*: I: 1,2,3,4,5
II: 1,2,4,5,6

Table 13 (continued)

5. Teamwork and Compatibility

MSS

Group:	1	2	3	4	5	6
Mean:	6.07	5.44	5.39	5.44	5.48	4.82

Trends: Linear

Similar Groups*: I: 1,2,3,4,5
II: 2,3,4,5,6

BARS

6. Performance Under Stressful Conditions

MSS

Group:	1	2	3	4	5	6
Mean:	3.36	4.50	5.56	5.44	5.14	5.64

Trends: Linear, Quadratic, Cubic

Similar Groups*: I: 1
II: 2,5
III: 3,4,5,6

BARS

Group:	1	2	3	4	5	6
Mean:	4.71	5.19	5.56	6.00	5.67	5.50

Trends: Linear, Quadratic

Similar Groups: I: 1,2
II: 2,3,4,5,6

7. Leadership

MSS

Group:	1	2	3	4	5	6
Mean:	4.57	4.81	5.39	5.67	5.43	5.45

Trends: Linear

Similar Groups*: I: 1,2,3,4,5,6

BARS

Group:	1	2	3	4	5	6
Mean:	4.29	4.50	5.56	5.72	5.38	5.68

Trends: Linear, Quadratic

Similar Groups: I: 1,2
II: 3,4,5,6

* Similar Groups - groups listed after a roman numeral have mean scores not significantly different from each other based on a Neuman-Keuls statistic following an ANOVA.

regressions contained significant linear or quadratic trends. The data are summarized in Table 14.

This difference in trend significance between lieutenant and captain raters shows up in the polynomial trend analyses for the individual subscales. For the lieutenant BARS format, the following subscales contained significant linear and quadratic trends (see Table 13 for the category means): knowledge of the job, physical fitness, performance under stressful conditions and leadership. The willingness to work subscale contained only a linear trend.

For the lieutenant MSS subscales the knowledge of the job dimension contained significant linear and quadratic trends. The teamwork and compatibility dimension and the leadership dimension contained significant linear trends. The performance under stressful conditions subscale contained significant linear, quadratic and cubic trends. For lieutenant raters, the BARS format contained more significant linear and quadratic trends than the MSS both on the basis of total score and subscale score analyses.

Only one subscale for the captains' BARS format reached statistical significance in the trend analysis.

Table 14

Group Means of Significant ANOVA Trend Analyses for Captain Rater Seniority Criterion Variable Subscales

1. Leadership

MSS	BARS						
(NS)	Group	1	2	3	4	5	6
	Mean	4.15	4.58	5.28	5.22	6.00	5.50
	Significant trends:			Linear			

The leadership dimensions contained a significant linear relationship in which performance appraisal scores generally increased from the low seniority levels through the higher seniority levels (see Table 14). None of the other subscales for the BARS ratings by captains reached statistical significance. In addition, none of the captain rater MSS subscales reached statistical significance with a linear trend.

Education A firefighter's education level was also considered an important validity criterion variable. A firefighter with some college courses or night school fire science courses should be able to perform his fire duties with a good deal of intelligence and understanding. This greater knowledge of firefighting should lead to higher efficiency marks. A firefighter's education level should be predicted by his efficiency mark score.

The education criterion variable was constructed from four self-report variables. These variables assessed two major aspects of firefighting relevant education: the amount of college courses and the number of fire science courses that the firefighter had taken. If a firefighter had attended some college but had not received a degree, he was given one point. If a firefighter had attended some college and received an associates (two year) degree, he received a point for attending college and receiving a two year degree. If the firefighter had received a bachelor's degree, he received points for some college, a two-year degree and a four year degree. These were the first three variables comprising the education criterion. The fourth and final variable composing the education variable was whether or not the firefighter had taken any fire science courses. These four variables defined the education criterion. The minimum possible score was zero and the maximum was four. The mean of the education variable was 0.9588 and the standard deviation was 1.04 (N = 194). The smallest intercorrelation among the four variables was 0.15999 (N = 167, NS) between the college attendance and college graduate variables. The largest inter-correlation was between the associates degree and college graduate variables ($\underline{r}$ = .57343, $\underline{p}$ < .01). All the intercorrelations were positive. The education criterion variable had an alpha reliability of 0.62586. This relatively low reliability, low mean and small standard deviation makes the education criterion variable somewhat suspect in its usefulness as an accurate indication of the firefighter's education. Very little variance was present in this criterion.

The BARS total score based upon the ratings of lieutenants and captains combined accounted for a total of 7.667% of the education variable, F (1, 177) = 14.697, $p < .01$. However, the Pearson Product Moment Correlation between the BARS total score and the education variable was negative ($r = -0.2769$, $p < .001$). These same results, although slightly smaller, hold true for the MSS total scores based on all raters. The MSS total score accounted for 1.994% of the variance, F (1, 176) = 3.581, NS. The correlation between MSS total score and education was -0.1412, $p < .03$. All education criterion regressions are summarized in Appendix L.

When analyzed separately for total scores by lieutenant raters, the BARS total score accounted for 5.076% of the variance in education, F (1, 119) = 6.36, $p < .05$. The correlation between lieutenant BARS and education was -0.2253, $p < .01$. The lieutenant MSS total scores accounted for 0.496% of the education variance, F (1, 119) = 0.5932, NS. The correlation between MSS total score and education was -0.0704, NS.

For captain raters, the BARS total scores accounted for 11.129% of the variance in education, F (1, 56) = 7.379, $p < .05$. The correlation between captain MSS total scores and education was -0.3336, $p < .01$. The captain assigned MSS total scores accounted for a total of 5.700% of the education variance, F (1, 55) = 3.385, NS. Once again the education -MSS total score correlation was negative, $r = -0.2388$, $p < .05$.

The nine BARS subscales based on the ratings of all officers accounted for a total of 16.394% of the variance in the education criterion variable, F (8, 170) = 4.1668, $p < .01$. Willingness to work entered the prediction equation first and accounted for a total of 10.997% of the variance, F (1, 170) = 22.36, $p < .01$. Willingness to work uniquely accounted for 3.496% of the variance in education, F (1, 169) = 7.067, $p < .01$. Willingness to work was significantly negatively correlated with the education variable, $r = -0.332$, $p < .01$. All of the other subscales except physical fitness, $r = .0036$, NS, correlated negatively with the experience variable. However, willingness to work was the only BARS subscale to predict a significant portion of the education variance.

The nine MSS subscales based on the ratings of all officers accounted for 10.914% of the variance in education scores, $\underline{F}$ (9, 168) = 2.59, $\underline{p}$ <.01. In a stepwise hierarchical regression, knowledge of the job entered the prediction equation first and accounted for 6.271% of the variance, $\underline{F}$ (1, 169) = 11.898, $\underline{p}$ <.01. Of this variance, 2.445% is uniquely accounted for variance in education, $\underline{F}$ (1, 168) = 4.611, $\underline{p}$ <.1, not a statistically significant amount. Teamwork and compatibility entered the prediction equation second and accounted for an additional 2.44% of the variance, $\underline{F}$ (1, 169) = 4.632, $\underline{p}$ <.1. Teamwork and compatibility uniquely accounted for 2.539% of the variance, $\underline{F}$ (1, 168) = 4.788, $\underline{p}$ <.1. None of the MSS subscales uniquely accounted for a significant portion of the education variance. Teamwork and compatibility, and initiative were the only two MSS subscales to correlate positively with education (teamwork and compatibility $\underline{r}$ = .077, NS, initiative $\underline{r}$ = .0386, NS).

Analyzing the nine BARS subscale data separately for lieutenant ratings, a total of 14.511% of the education variance was accounted for, $\underline{F}$ (1, 11) = 2.09, $\underline{p}$ <.06. Of the nine BARS subscales, willingness to work entered the stepwise hierarchical regression prediction equation first accounting for 8.298% of the variance, F (1, 111) = 10.774, p <.01. Willingness to work also uniquely accounted for a significant 4.415% of the education variance, F (1, 111) = 5.732, p <.05. As is true of the other eight subscales, willingness to work correlated negatively with the education criterion variable, r = -0.288, p <.01. None of the other eight BARS subscales accounted for significant portions of the variance in the education variable.

With the nine MSS subscales utilized by lieutenant raters, a total of 9.464% of the education variance was accounted for, F (8, 112) = 1.463, NS. Knowledge of the job entered the stepwise hierarchical regression equation first and accounted for 2.809% of the variance, F (1, 112) = 3.48, NS. Knowledge of the job also uniquely accounted for a significant 3.48% of the variance, F (1, 111) = 4.296, p < .05. The correlation between knowledge of the job and education was negative, r = -.168. Teamwork and compatibility entered the prediction equation second and accounted for an additional 4.447% of the variance, F (1, 112) = 5.501, p <.05. Of this variance, teamwork and compatibility uniquely accounted for 4.805%, F (1, 111) = 5.89, p <.05. Teamwork and compatibility and knowledge of the job were the only MSS subscales to account for significant portions of the education variance. Teamwork and compatibility had a significant positive correlation, r = .15778, p <.05, while knowledge of the job was negatively correlated with the education criterion, r = -.1676, p < .05.

The nine BARS subscales analyzed separately for captain ratings accounted for a total of 22.539% of the education variance, F (8, 49) = 1.782, NS. Leadership entered the stepwise hierarchical regression equation first and accounted for 17.813% of the variance, F (1, 49) = 11.268, p <.01. Leadership could uniquely account for only 1.872% of the variance in education, F (1, 49) = 1.184, NS. The correlation between education and leadership was, oddly enough, significant and negative, r = -.422, p <.01. Only physical fitness correlated positively with education r = .05, NS. Physical fitness entered the stepwise regression prediction equation second and accounted for an additional 1.894% of the variance, F (1, 58) = 1.198, NS. Physical fitness uniquely accounted for 1.652% of the education variance, F (1, 49) = 1.045, NS.

The nine MSS subscales used by captain raters accounted for a total of 19.634% of the variance in education, F (6, 50) = 2.036, NS. Knowledge of the job entered the stepwise hierarchical regression equation first and accounted for the only significant portion of the education variance, 13.067%, F (1, 50) = 8.13, p <.01. Knowledge of the job uniquely explained, however, only 2.563% of the variance in education, F (1, 50) = 1.578, NS. Performance under stressful conditions entered the stepwise hierarchical regression prediction equation second and accounted for an additional 1.72% of the variance, F (1, 50) = 1.07, NS. Performance under stressful conditions could uniquely explain 3.02% of the variance in education after the variance due to the other eight subscales was removed, F (1, 50) = 1.88, NS. Both performance under stressful conditions and knowledge of the job correlated moderately and negatively with education (knowledge of the job, r = -.36148, p <.01; performance under stressful conditions r = -.35472, p <.01). It appears if these two variables explained approximately the same set of variance in education. Only the initiative subscale correlated positively with education (*r* = .09351, NS).

Promotion Related Experience A firefighter's amount of experience with duties commonly assigned to officers, or firefighters ready for promotion was considered an important indication of the quality of work. An experience criterion variable was calculated on the basis of four items on the self-report questionnaire. The four variables composing the experience criterion were: (1) the number of times in the previous month that the firefighter drove a fire truck (2) the number of times in the previous month that the firefighter drove a battalion chief's truck (3) the number of times in the previous month that the firefighter performed as an acting lieutenant and (4) the number of times in the previous month that the firefighter performed as an acting fire engineer. Each of these variables was scored the same way. A "zero" was assigned for no performance, a "one" was given for performing a duty once or twice and a "two" was assigned for performing a duty three times or more often.

The resulting experience criterion variable had a minimum possible score of zero and a maximum of eight.

The mean of the experience variable was 3.204 with a standard deviation of 2.4111 (N = 155). The item intercorrelations are all small, but positive. The smallest correlation is between the acting fire engineer variable and the battalion chief driver variable, $\underline{r}$ = .092, NS, and the largest is between the acting fire engineer and driving a rig variables, $\underline{r}$ = .449, $\underline{p}$ <.01. The alpha reliability of the experience variable is 0.633.

The BARS total score based on the ratings of all officers could account for 5.22% of the variance in the experience variable, $\underline{F}$ (1, 153) = 8.441, $\underline{p}$ <.01. The Pearson product moment correlation between the BARS total score and experience was a statistically significant 0.2286, $\underline{p}$ <.01. The MSS total score based on the ratings of all officers accounted for 4.334% of the variance in experience, $\underline{F}$ (1, 151) = 6.84, $\underline{p}$ <.01. The simple correlation between the MSS total score and the experience variable was 0.2082, $\underline{p}$ <.01. Both correlations were positive and modest in weight. The experience criterion variable regressions are summarized in Appendix M.

When analyzed separately for lieutenant and captain raters, the results were not consistent. The 101 lieutenant BARS total scores correlated 0.199 with the experience variable and accounted for 3.997% of the variance, F (1, 99) = 4.12, p <.05. The 101 lieutenant MSS total scores correlated 0.08681 with the experience variable and accounted for 0.754% of the variance, F (1, 99) = 0.752, NS. For the 51 captain raters however, the MSS correlations with experience were larger than the BARS correlations with experience. The captain rater BARS total scores correlated 0.283 with experience ratings and accounted for 8.031% of the variance, F (1, 49) = 4.279, p <.05. The captain MSS total scores correlated .401 with experience and accounted for 16.107% of the variance F (1, 49) = 9.408, p <.01.

The nine BARS subscales based on the ratings of all officers together accounted for a total of 15.836% of the experience variable variance, F (7, 147) = 3.951, p <.05. The leadership subscale entered the prediction equation first and accounted for the only significant portions of the variance in experience, r^2 = 0.12669, F (1, 147) = 4.422; p <.01. The leadership subscale of the BARS ratings of all officers was also able to uniquely account for a significant portion of the variance in the experience variable, unique r^2 = 0.02535, F = (1, 145) = 4.367, p <.05. None of the remaining BARS subscales accounted for significant portions of the variance in the experience criterion variable.

The nine MSS subscales based on all officers ratings accounted for a total of 19.865% of the variance in the experience variable, F (9, 143) = 3.939, p <.01. experience variable, F (9, 143) = 3.939, p <.01. The performance under stressful conditions subscale of the MSS ratings based on all raters entered the prediction equation first and accounted for the only statistically significant protion of the variance, r^2= 0.147, F (1, 143) = 9.573, p <.01. Performance under stressful conditions was also able to uniquely account for a significant portion of the experience variance, r^2 = 0.05364; F (1, 143) = 9.573, p <.01. This was the only MSS subscale based on all officer's ratings to account for a unique portion of the variance.

The nine subscales of the BARS and MSS were analyzed separately for lieutenant and captain raters. The nine BARS subscales for lieutenant raters were able to account for 15.831% of the variance in the experience criterion, F (8, 92) = 2.16, p <.05. The knowledge of the job subscale entered the prediction equation first and was able to account for 10.343% of the variance, F (1, 92) = 4.525, p <.05. The leadership subscale entered the prediction equation second and was able to account for a statistically significant 2.008% of the variance, F (1, 92) = 4.212, p <.01. Both of these subscales utilized by lieutenant raters were able to uniquely account for significant portions of the experience variance. Knowledge of the job was able to uniquely account for 4.140% of the variance, F (1, 91) = 4.476, p <.05. The leadership subscale used by lieutenant raters was able to uniquely account for 3.77% of the experience variance, F (1, 91) = 4.076, p <.05.

The nine MSS subscales used by lieutenant raters were able to account for a total of 19.549% of the variance in the experience criterion variable, F (9, 91) = 2.457, p <.05. Performance under stressful conditions entered the prediction equation first and accounted for 11.269% of the variance, F (1, 91) = 10.620, p <.01. None of the other subscales accounted for statistically significant portions of the experience variance in the standard regression equation. The performance under stressful conditions subscale was also able to uniquely account for 9.389% of the experience variable variance, F (1, 91) = 10.62, $p < .01$.

When analyzed separately for captain raters, the nine BARS subscales together accounted for 32.971% of the variance in the experience total score, F (9, 42) = 2.296, $p < .05$. Performance under stressful conditions entered the prediction equation first and accounted for 17.835% of the variance in the experience variable. This amount was not statistically significant, F (1, 42) = 2.553, NS. The physical fitness subscale of the captain's BARS entered the standard regression equation second and accounted for 6.533% of the variance, F (1, 42) = 2.918, NS. Again, this is not a statistically significant amount. Performance under stressful conditions was able to uniquely account for a nonsignificant 4.075% of the variance, F = (1, 42) = 2.553, NS. Physical fitness uniquely accounted for 4.658% of the experience variance in captain's BARS ratings. This was not significant, F (1, 42) = 2.918, NS.

The captain rater's nine MSS subscales were able to account for 43.589% of the variance in the experience criterion variable, F (9, 41) = 3.52, $p < .01$. The leadership subscale entered the prediction equation first and accounted for 24.673% of the variance, F (1, 41) = 6.184, $p < .05$. The public relations subscale entered the prediction equation second and accounted for 4.371% of the variance, F (1, 41) = 5.464, $p < .05$. The teamwork and compatibility subscale entered the prediction equation third and accounted for 8.655% of the variance in experience, F (1, 41) = 6.464, $p <.05$. Each of these three subscales were also able to uniquely account for significant portions of the experience variance. The leadership subscale uniquely accounted for 8.509% of the variance, F (1, 41) = 6.184, $p <.05$. The public relations subscale uniquely accounted for 7.517% of the experience variance, F (1, 41) = 5.464, $p <.05$. Teamwork and compatibility uniquely accounted for 8.894% of the variance in the experience criterion, F (1, 41) = 6.464, $p < .01$.

To summarize, both the BARS and the MSS were able to account for statistically significant portion of the promotions related experience criterion variance. The amount of variance explained was minimal, with the BARS total score explaining slightly more variance than the MSS. However, when analyzed according to the nine subscales, the MSS outperformed the BARS. Finally, captain raters were able to better predict the experience criterion than were lieutenant raters.

The Practical Performance Tests During the performance appraisal data gathering sessions, each firefighter being evaluated was given one of eight short practical performance tests. Every member of a single session was given the same practical performance test. The tests were randomly assigned to the different data collection sessions. Eight different tests were used, mainly for practical reasons. If only a single test was used, news of the test and its items would have quickly spread around the City's Firehouses. Firefighters tested near the end of the data collection period would have scored higher than those near the start.

The eight randomly assigned tests were developed by the Fire Department for in-station practice drills. Although readily available for use, none of the fire stations tested had previously conducted any of these eight practice drills. The number of firefighters tested on each practical performance test, along with the mean, standard deviation and correlations with the total scores on the BARS and MSS for all raters are presented in Table 15. Since the means and standard deviations varied widely among the eight tests, the test scores were standardized, i.e., each of the eight practical performance tests was transformed to a mean of zero and a standard deviation of 1.00. However, it appears from the correlations with BARS and MSS total scores that the practical performance tests were measuring widely different things. Three of the practical performance tests correlated negatively with the BARS total scores (spanners, ropes and ladders) with the spanner test being statistically significant, $\underline{r}$ = -.3328, $\underline{p}$ <.05. Two of the practical performance tests correlated negatively with the MSS total scores based on all raters (spanners and ropes), neither one is statistically significant. The pattern of correlations of the practical performance tests with the BARS and MSS is similar. All of the remaining practical performance tests were positively correlated with the BARS and MSS total scores. Only the practical performance test measuring the use of "two 1-1/2 inch lines for overhaul"

Table 15

Eight Practical Performance Tests Descriptive and Correlational Statistics

	Test Number	Number Tested	Mean	Standard Deviation	BARS Item-total *r*	MSS Item-total *r*
Utilizing two 1½ inch lines for overhaul	1	6	94.0	7.899	.8110**	.7064*
Small tools and fittings	2	34	76.9	18.217	.1207 NS	.0232 NS
Spanner	3	26	79.7	25.945	-.3328**	-.2626*
Ropes	4	20	68.2	26.661	-.1129 NS	-.0391 NS
Point of vantage leadout	5	31	80.9	13.543	.2582*	.2649*
Ladders	6	25	63.0	16.894	-.1551 NS	.0160 NS
Distributor nozzel layout	7	8	84.4	12.374	.3267 NS	.6098*
Securing a charged hoseline in a circle	8	12	71.7	26.227	.2919 NS	.3271 NS
Standardized Practical Performance Test		185	.044	.898	.0334 NS	.0731 NS

* $p < .05$
** $p < .01$

reached significance in correlation with the BARS total score, $r = .811$, $p < .025$. However, the number of ratees utilizing this test is extremely small. Thus, it is not clear that these eight concatenated practical performance tests can be said to be measuring the same performance abilities in the different firefighter groups. However, there was no a priori or a posteriori justification for eliminating one or several of these practical performance tests. The validity analysis on practical performance was done utilizing each firefigher's score on which ever one of the eight tests he took.

The BARS total score based on all rater ranks were regressed against the standardized practical performance test. The BARS accounted for only 0.11% of the practical test's variance, F (1, 177) = 0.1988, NS. The overall correlation between the BARS total scores and the practical test was a non-significant 0.0334. The results for the MSS based on all raters was similar. The MSS total scores accounted for 0.53% of the variance in the practical test, F (1, 176) = 0.9509, NS. The correlation between MSS total scores based on all raters and practical performance scores was a non-significant

0.0731. The practical performance test criterion regressions are summarized in Appendix N.

When analyzing the BARS and MSS total scores separately for lieutenant raters and for captain raters, the simple regressions on practical performance scores remain non-significant; but, the correlations are negative for lieutenants and positive for captains. For lieutenant raters, the BARS total scores accounted for 0.032% of the variance in the practical performance test, $\underline{F}$ (1, 119) = .032, NS. The correlation between the lieutenant BARS and the practical performance test was a non-significant -0.018. The lieutenant rater MSS total scores accounted for 0.0025% of the variance, $\underline{F}$ (1, 119) = 0.003, NS. The correlation was a non-significant -0.005.

For the captain raters, however, the relationship between total performance evaluation scores and practical performance test scores was positive. The captain raters BARS accounted for 1.46% of the variance in practical test scores, $\underline{F}$ (1, 56) = 0.8455, NS. The correlation was 0.1209. The MSS total scores given by captain raters accounted for 4.83% of the variance in the practical performance test scores, $\underline{F}$ (1, 55) = 2.84, NS. The correlation between the captain raters total MSS scores and the practical performance test scores was a significant 0.2198, $p < .05$.

Standardization of the eight component scales of the practical performance tests was done to minimize the effect of the differences between tests. The effect due to the individual test on the practical performance test scores was limited as a possible source of error variance. However, one additional source of error variance was still possible. Since the same test was used for each firefighter within a performance rating sesssion, it was possible that word of the test items could have secretly spread to the later test takers. If this was the case, then firefighters taking the practical performance test later in the experimental session would get higher test scores than those at the start. In order to control for this possibility, each of the following criterion regressions using the nine BARS or MSS subscales was done after the effect for testing order was entered into the prediction equations. In none of the six sets of regressions did test taking order account for a significant portion of the variance.

The nine BARS subscales, together with the test taking order variable accounted for a total of 5.16% of the variance in the practical performance test criterion variable, $\underline{F}$ (10, 123) = .664, NS. The test taking order was entered into the prediction equation first and allowed to account for as much variance as it was able. Test taking order, when entered with the nine BARS subscales for all raters, accounted for 0.134% of the variance, $\underline{F}$ (1, 123) = 0.1738, NS. Since the test order variable did not account for a significant portion of the variance, it may be assumed that the variance attributed to the test order variable is error variance. Thus, no variance due to the test order was removed from the prediction equation prior to the calculations of significance tests for the nine BARS subscales.

Of the nine BARS subscales for all raters, none accounted for a significant portion of the variance in the practical performance test score. Willingness to work entered the stepwise hierarchical regression equation first (after the test order variable) and accounted for 0.844% of the variance, $\underline{F}$ (1, 123) = 1.095, NS. Willingness to work uniquely accounted for 0.736% of the variance, $\underline{F}$ (1, 123) = 0.9545, NS. The correlation between willingness to work and the practical performance test was positive, $\underline{r}$ = .09486, NS. Performance under stressful conditions ($\underline{r}$ = .04590), leadership ($\underline{r}$ = .08719), initiative ($\underline{r}$ = .04820) and appearance ($\underline{r}$ = .03672) all correlated positively with the practical performance test, but were not statistically significant. The remaining BARS subscales all correlated negatively. Knowledge of the job entered the stepwise hierarchical regression second, after willingness to work, and accounted for 1.716% of the variance, $\underline{F}$ (1, 123) = 2.2255, NS. Knowledge of the job uniquely accounted for 0.808% of the variance, $\underline{F}$ (1, 123) = 1.048, NS.

The nine MSS subscales plus the test order variable based on the ratings of all officers accounted for a total of 9.740% of the variance in the practical performance test, $\underline{F}$ (9, 123) = 1.47, NS. Test order, forced to enter the prediction equation before the nine MSS subscales, accounted for a non-significant 0.086% of the variance, $\underline{F}$ (1, 123) = 0.1172, NS. Again, since the test order variable could not explain a significant portion of the variance, the amount attributed to test order was allowed to remain in the prediction equation.

None of the MSS subscales could explain significant portions of the variance in the practical performance test. Appearance was the first MSS subscale to enter the stepwise hierarchical regression equation. Appearance accounted for 3.57% of the variance, F (1, 123) = 4.8649, p <.1, while uniquely accounting for 4.315% of the variance, F (1, 123) = 5.8324, p <.05. Appearance correlated positively with the practical performance test, r = .18976, p <.01. Performance under stressful conditions was the second MSS subscale to enter the practical performance test predictions equation. Performance under stressful conditions accounted for an additional 3.261% of the variance, F (1, 123) = 4.44, p < .05, while uniquely accounting for 1.346% of the variance F (1, 123) = 1.82, NS. Performance under stressful conditions correlated negatively with the criterion r = -.07919, NS, and initiative, r = -.00155, NS. The other subscales correlated positively.

Analyzing the standardized practical performance test data separately for lieutenant raters, the nine BARS subscales and the test taking order variable accounted for a total of 10.23% the variance, F (9, 78)

= 0.9877, NS. The test order variable was entered first into the regression equation and could account only for a non-significant 0.043% of the variance, $\underline{F}$ (1, 78) = 0.0374, NS. This variance was kept in the equation, since it could not be statisticlly significantly distinguished from error variance. The first BARS subscale to enter the stepwise hierarchical regression equation was knowledge of the job. It accounted for 3.009% of the practical test variance, $\underline{F}$ (1, 78) = 2.6145, NS, while uniquely accounting for 2.711% of the variance $\underline{F}$ (1, 77) = 2.3254, NS. Knowledge of the job correlated negatively with the practical test variable ($\underline{r}$ = -.1715). Willingness to work entered the stepwise regression equation second, after knowledge of the job. Willingness to work accounted for 2.509% of the variance, $\underline{F}$ (1, 78) = 2.2165, NS, while uniquely accounting for 1.191% of the variance, $\underline{F}$ (1, 77) = 1.0216, NS. Willingness to work correlated positively with the dependent variable ($\underline{r}$ = .1414), performance under stressful conditions ($\underline{r}$ = .1795), teamwork and compatibility ($\underline{r}$ = .0485) and initiative ($\underline{r}$ = .00021).

The nine MSS subscales and the test taking order variable analyzed for lieutenant ratings accounted for

16.406% of the variance in the practical performance test criterion, F (10, 77) = 1.51, NS. The test order variable accounted for 0.043% of the variance, F (1, 77) = 0.0396, NS. This non-significant variance remained in the prediction equation, and was not removed. The first MSS subscale to enter the prediction equation was knowledge of the job, and was the only subscale to account for a significant portion of the variance in the criterion. Knowledge of the job accounted for 7.915% of the variance, F (1, 77) = 7.29, p <.01. Knowledge of the job uniquely accounted for a significant 6.526% of the variance, F (1, 77) ≐ 6.01, p <.05. However, the Pearson Product Moment Correlation between knowledge of the job and the criterion variable was negative, r = -.27767, p < .01. In other words, knowledge of the job--the only MSS subscale for lieutenant raters to account for a significant portion of the variance--was negatively related to the criterion. Firefighters that scored well on the practical performance test tended to get low scores on knowledge of the job. This certainly does not speak highly of the relationship between the MSS scale and the criterion variable.

The nine BARS subscales plus the test taking order variable, when analyzed separately for captain ratings, accounted for 15.878% of the variance in the practical performance test, $\underline{F}$ (10, 35) = .66, NS. The test order variable accounted for 0.348% of the variance, $\underline{F}$ (1, 35) = 0.1448, NS. This random error was allowed to stay in the prediction equation. The first BARS subscale to enter the stepwise hierarchical regression equation was initiative. Initiative accounted for 5.832% of the variance, $\underline{F}$ (1, 35) = 2.423, NS. Initiative uniquely accounted for 3.687% of the variance, $\underline{F}$ (1, 35) = 1.334, NS. The correlation between initiative and the criterion variable was a positive 0.244. Appearance, the only BARS subscale to correlate negatively with the practical performance test, $\underline{r}$ = -.04141, NS, loaded second in the stepwise hierarchical regression equation. Appearance accounted for an additional 2.147% of the variance, $\underline{F}$ (1, 35) = 0.8933, NS. Appearance uniquely accounted for 2.46% of the variance, $\underline{F}$ (1, 35) = 1.027, NS.

The nine MSS subscales for captain raters plus the test order variable accounted for a total of 28.858% of the variance in the practical performance test, $\underline{F}$ (10,

34) = 1.379, NS. The test order variable accounted for a non-significant 0.14% of the variance, F (1, 34) = 0.0669, NS. This variance remained in the equation. The first MSS subscale to enter into the stepwise hierarchical regression equation was appearance. Appearance was the only MSS subscale for captain raters to account for a significant portion of the variance in the practical performance tests, 13.017%, F (1, 34) = 6.22, NS. Appearance uniquely accounted for 14.711% of the variance, F (1, 34) = 7.03, p <.01. The correlation between appearance and the practical performance test was a significantly positive 0.36131, p <.01. The only MSS subscale to correlate negatively with the criterion variable was initiative, r = -.10788, NS. Initiative loaded second in the stepwise hierarchical regression equation behind appearance. It accounted for 6.985% of the variance, F (1, 34) = 3.338, NS. Uniquely, initiative accounted for 4.793% of the variance, F (1, 34) = 2.29, NS.

A summary of the validity analyses on the BARS and MSS using percentages of variance accounted for and subscales responsible for the significant effects is presented in Table 16.

Table 16

Summary of Validity Analyses on the BARS and MSS Using Percentages of Variance Accounted for and the Subscales Responsible for Significant Effects

	MSS	BARS
1. Promotional Rank		
Total score		
All Raters (N=83)	21.83%**	14.09%**
Lieutenants (N=63)	22.55%**	22.10%**
Captains (N=20)	19.36% (.1)	1.62% NS
Subscales		
All Raters	29.695%*	26.22%**
	Appearance	Initiative
	Knowledge of the job	
Lieutenants	35.07%**	36.212%
	Appearance	Initiative
Captains	31.30% NS	33.05% NS
2. Seniority (linear regressions only)		
Total Score		
All Raters (N= 179)	.767% NS	.846% NS
Lieutenants (N=121)	.964% NS	.916% NS
Captains (N= 57)	.032% NS	.791% NS
Subscales		
All Raters	7.964% NS	26.998%**
		Leadership
		Physical Fitness
		Knowledge of the Job
Lieutenants	9.335% NS	33.77%**
		Knowledge of the Job
		Physical Fitness
Captains	9.942% NS	28.276% NS
3. Education***(Alpha Reliability = .62586)		
Total Score		
All Raters (N=178)	(-) 1.994%*	(-) 7.667%**
Lieutenants (N=121)	(-) .496% NS	(-) 5.076%*
Captains (N= 57)	(-) 5.700% NS	(-)11.129%*
Subscales		
All Raters	10.914%*	16.394**
	(-) Knowledge of the Job	(-) Willingness to Work
	(+) Teamwork and Compatibility	
Lieutenants	9.464% NS	14.511% (.06)
		(-) Willingness to Work
Captains	19.634% NS	22.539% NS

Table 16 (cont.)

	MSS	BARS
4. Promotion Related Experience (Alpha Reliability = .633)		
Total Score		
All Raters (N= 155)	4.334%**	5.22%**
Lieutenants (N= 101)	.754% NS	3.997%
Captains)N= 54)	16.107%**	8.031%
Subscales		
All Raters	19.865%** Performance Under Stressful Conditions	15.836% Knowledge of the Job Leadership
Lieutenants	16.406%** Knowledge of the job	10.23%
Captains	43.582% Leadership Public Relations Teamwork and Compatibility	32.971%*
5. Practical Performance Tests		
Total Score		
All Raters (N= 178)	.53% NS	.11% NS
Lieutenants (N= 121)	.2225% NS	.032% NS
Captains (N= 57)	4.83% NS	1.46% NS
Subscales		
All Raters	9.74% NS	9.74% NS
Lieutenants	(-) 16.406% NS	10.23% NS
Captains	28.858% NS	15.878% NS

* $p < .05$

** $p < .01$

*** sign in parentheses is the direction of the correlation

Table 16 (cont.)

6. Correlations of BARS and MSS total scores with the five criteria

	BARS	MSS
Promotional Rank	-.376	-.467
Seniority	-.110	-.090
Education	-.277	-.141
Experience	-.084	-.073
Practical Performance Test	.033	.073

7. Intercorrelations of the five criteria

	PR	S	Ed	Ex
Promotional Rank (PR)	———			
Seniority (S)	.11244	———		
Education (Ed)	-.11515	.37707	———	
Experience (Ex)	-.06575	.68004	.26674	———
Practical Performance	-.07651	-.01958	-.00052	.06843

Table 16 also presents the intercorrelations between the five criterion variables with each other and with the BARS and MSS total scores. The intercorrelations between the five criteria show the reason why a single composite criterion composed of the five separate criteria is inadvisable. The pattern of positive, near zero and negative intercorrelations between the criteria suggest that each criteria was measuring an independent component of the rated performance. The consistency of the correlations of the five criteria with the BARS total score and the MSS total score, in terms of the amount and direction of the correlations across the two formats, suggest that the criteria were assessing similar things in the two formats. For example, the promotional rank order criteria correlated moderately and negatively with both the BARS and the MSS. Thus, the five criteria each measured approximately the same five concepts in the BARS and MSS. Furthermore, these five concepts were generally independent of each other (except for seniority and experience). A single composite criterion would have glossed over these differences between the five criteria and important information would have been lost. Finally, it would be inadvisable to sum together criterion scores that measure different concepts. One would be hard pressed to state exactly what this composite criterion was measuring. Rather, it was deemed wiser to assess all the information in the five criteria separately.

Attitude and Scale Preference Surveys

Every firefighter, lieutenant and captain participating in the criterion validity study was given a chance to rate the performance of at least one other person in his company on both the BARS and the MSS. Following the ratings on the two performance evaluation scales, all participants filled out a performance evaluation scale preference survey. The scale preference survey contained two identical, six-item semantic differential attitude scales. One attitude scale assessed opinions toward the MSS and the other assessed opinions toward the BARS. Each item in the semantic differential scales was a seven point checklist. A "one" indicated an "extremely" positive evaluation. A "seven" indicated an "extremely" negative evaluation on the bipolar adjective pair. One item was reverse scored in order to be consistent with this numbering scheme. The six item scores were summed up to yield a BARS and MSS semantic differential attitude total score. Lower numbers indicate more positive attitudes. The MSS semantic differential attitude scale had an alpha reliability of 0.86175, while the BARS semantic differential attitude scale had an alpha reliability of 0.85713. Both alpha reliabilities are quite respectable.

The means and standard deviations for each of the six semantic differential items are presented in Table 17. In each case, the mean attitude score on the MSS is lower than the mean of the BARS. This indicates that the MSS had a better evaluation than the BARS. A t-test between the total MSS semantic differential attitude score and the BARS attitude score, based on the ratings by both officers and firefighters, indicated that the MSS was preferred over the BARS, t (220) = 3.14, p = .002, two-tailed. The semantic differential attitude scale total scores were analyzed separately for firefighters only. The MSS preformance evaluation scale was again preferred over the BARS, t (178) = -3.14, p = .002, two-tailed. However, when analyzed only for officers, there was no difference between semantic

Table 17

Semantic Differential Attitude Scale, Scale Preferences Score Preference and Sentence Preference Analyses

	MSS		BARS	
Adjectives	MEAN	STD	MEAN	STD
good/bad	3.000	1.543	3.387	1.603
hard/easy (reverse scored)	3.115	1.810	3.482	1.889
useful/useless	2.987	1.462	3.273	1.619
relevant/irrelevant	3.009	1.494	3.221	1.603
personally liked/disliked	3.106	1.665	3.634	1.794
liked by Firefighters/dis-liked	3.304	1.520	3.729	1.670
TOTAL	18.571	7.324	20.631	7.784

I. BARS vs. MSS Preference t-Tests based on the Sementic Differential

1. Combined Firefighters and Officers
   t (220) = 3.14 p = .002 (two-tailed)

2. Firefighters Only
   t (178) = -3.14 p = .002 (two-tailed)
   $\bar{x}$ MSS = 18.7486 STD = 7.217
   $\bar{x}$ BARS = 21.151 STD = 7.731

3. Officers Only
   t (41) = -0.79 p = .435 (two tailed)
   $\bar{x}$ MSS = 17.2381
   $\bar{x}$ BARS = 18.6190

II. Scale Preference*

	Combined Freq.	Firefighters Freq.	Officers Freq.
(1) BARS	75	57	18
(2) MSS	114	93	21
(3) Neither	31		
CHI-SQUARE	8.048	8.640	0.231
p	.005	.003	.631

Table 17 (cont.)

III. Score Preference*

	Combined Freq.	Firefighters Freq.	Officers Freq.
(1) Know Numerical Score immediately	170	142	28
(2) Not Know Numerical Score immediately	51	38	13
(3) No Answer/Missing	34		
CHI-SQUARE (1)	64.077	60.089	5.488
p	.000	.000	.019

IV. Sentence Preference*

	Combined Freq.	Firefighters Freq.	Officers Freq.
(1) MSS sentence prefered	90	72	18
(2) BARS sentence prefered	57	43	14
(3) Both Acceptable	48		
(4) Neither prefered	23		
(5) Missing	37		
CHI-SQUARE	7.408	7.313	0.500
p	.006	.007	.480

* Chi-Square Statistics are based on alternatives one and two only.

differential total scores for the BARS and the MSS, t (41) = -.79, NS, two-tailed. Thus, it appears that overall, fire department personnel prefer the MSS. Firefighters, the men being rated, prefer the MSS. However, the officers who will actually be doing the performance evaluations do not have a preference for either the BARS or the MSS. Both performance evaluation scales are equally acceptable to the officers.

Three additional questions were asked in the scale preference survey. First, participants were asked directly which of the two performance evaluation scales they preferred. A total of 114 officers and firefighters preferred the MSS, 75 preferred the BARS and 31 either did not respond or did not like either format. A chi-square statistic was calculated on the BARS and MSS preference frequencies and indicated that there was a statistically significant difference between the two frequencies, Chi-square (1) = 8.048; p = .005. The general opinion of the fire department personnel was that the MSS was preferred over the BARS. Scale preference was also broken down by firefighter and officer. Once again, firefighters preferred the MSS while officers supported both formats equally. A total of 93 firefighters preferred the MSS, while 57 preferred the BARS, chi-square (1) = 8.64; $\underline{p}$ <.003. Of the officers, 21 preferred the MSS, while 18 preferred the BARS, chi-square (1) = 0.231, NS.

The second question asked on the scale preference survey attempted to assess a preference for one of the two formats according to a specific difference between the scales. In using the BARS scale, the rater knows the exact numerical score that he is assigning to a ratee. In the MSS scale, the rater does not know the numerical score he is assigning to a ratee. Rather, the rater simply marks down a plus (+), a zero (0), or a minus (-) depending on whether the ratee could be expected to perform his duties better, the same as, or worse than the statement, respectively. Using the combined data for both officers and firefighters, 170 preferred to know the numerical score immediately, 51 preferred not to know the numerical score immediately and 34 either did not answer or had no preference. The difference between the frequency of respondents preferring to know the numerical score immediately and those not preferring to know the numerical score immediately was statistically significant, chi-square (1) = 64.0, $p < .001$.

When analyzed separately for firefighter responses, the same results held. Of the 180 firefighters responding, 142 preferred to know the numerical score immediately and 38 preferred not to know the numerical score immediately, chi-square (1) = 60.089, $p < .000$. This same result held true for the officer responses; 28 officers preferred to know the numerical score immediately while 13 preferred not to know immediately, chi-quare (1) = 5.488, $p < .02$. Thus, the respondents, in general, prefer to know the numerical score being assigned to a ratee immediately.

The final major distinction between the rating scale formats was the type of sentences found within them. The BARS contained sentences taken word-for-word from the pool of sentences written by firefighters. The MSS contained sentences that were composites of several firefighter sentences as written by a battalion chief in cooperation with the research team. Using the combined data for both officers and firefighters, 48 men indicated that both sentence types were acceptable, 23 men said neither type was acceptable and 37 men did not respond. A total of 90 men indicated that they preferred the sentences in the MSS, while 57 preferred the sentences in the BARS, chi-square (1) = 7.408, *p* <.01. Of the 115 firefighters with a preference, 72 preferred the MSS sentence types while 43 preferred the BARS sentences chi-square (1) = 7.313; *p* < .01. Of the 32 officers responding, 18 preferred the MSS sentences while 14 preferred the BARS sentences, chi-square (1) = 0.50, NS. The officers did not clearly prefer either format's type of sentences, while the firefighters preferred those of the MSS.

Officer Rater Attitude-Behavior Consistency Each of the officers that rated firefighters under their command on BARS and MSS, also completed an attitude-behavior consistency survey. This survey was based on Fishbein and Ajzen's (1975) theory of behavioral expectations. The theory indicates that a person's behaviors are controlled by their intentions to behave in some specific fashion. The intention of concern in this survey is to "fill out the performance appraisal (efficiency mark) scale accurately for each firefighter under (the officer's) command, knowing that they will find out the marks given to them and that (the officers) should not give everybody the same high grade."

The intention is, in turn, a function of two subcomponents. An intention is a weighted sum of the attitude toward performing the behavior, plus the subjective norm toward performing the behavior. The attitude toward the behavior was measured in two ways. First, a semantic differential attitude scale toward the behavior in question was used. This six item adjective pair scale assessed the officers opinion on "accurately filling out the efficiency mark rating scale." The six items were summed to yield a total score. The lower the score the more favorable opinion. The mean attitude was 5.4746 with a standard deviation of 6.0325. The alpha reliability of the semantic differential attitude scale was 0.65564.

The theoretical definition's measure of the attitude toward the behavior is the sum of each individual salient belief's importance ratings and the evaluation of the belief. These ten beliefs were multiplied by their individual evaluation ratings and then summed to yield the measure of the attitude toward the behavior. The mean of this attitude toward the behavior was 16.9661 with a standard deviation of 28.1161 (N = 59). The alpha reliability of this ten item scale was 0.83486.

The subjective norm was also measured in two ways. The subjective norm was directly assessed by asking the officers how likely it is that "most people who are important to (them) and whose opinions (they) respect think that (they) should accurately complete the efficiency mark scale." On a seven point scale the mean rating was 1.7119 with a standard deviation of 1.1898. This indicates that it is at least "moderately likely" that officers feel that their co-workers think they should accurately complete the efficiency mark scale.

The theoretical definition of the subjective norm is the sum of each individual normative belief's importance rating multiplied by the motivation to comply rating. A total of nine normative beliefs were used to comprise the subjective norm. The mean subjective norm rating was 18.30159 with a standard deviation of 34.423. The alpha reliability of the subjective norm scale was 0.89. The alpha reliabilities as well as the means and standard deviations of expectancy-value theory components are presented in Table 18.

Table 18

Alpha Reliabilities, Means and Standard Deviations of the Expectancy Value Theory Components*

Component	Alpha Reliability	Mean	Standard Deviation
(B_i) (EB_i)	.83486	16.9661	28.1161
$(NB_i)(MC_i)$	.89769	18.9661	34.6868
Attitude toward the Behavior (AB)	.75132	1.9492	1.2652
Subjective Norm (SN)		1.7119	1.1898
Intent		2.2712	.9619

* N = 59, for all components

Simple regressions were run predicting the semantic differential score for the theoretical attitude toward the behavior measure and predicting the subjective norm direct measure from the theoretical subjective norm measure. The weighted sum of the salient beliefs times the evaluations of those beliefs, the theoretical measure of the attitude toward the behaviors, was able to account for 11.409% of the variance in the sematic differential, F (1, 57) = 7.60, $p < .01$, $r = .34$. The theoretical subjective norm, the weighted sum of the nine normative beliefs times the individual motivations to comply, was able to account only for 0.116% of the variance in the direct measure of the subjective norm, F (1, 57) = 0.098, NS, $r = .03$. The theoretical attitude toward the behavior is a better predictor of the direct attitude measure than the theoretical subjective norm is a predictor of the direct subjective norm measure. The expectancy value theory regressions are summarized in Appendix O.

However, in the multiple regression predicting the officers intention from the theoretical measure of attitude toward the behavior and subjective norm, the subjective norm becomes the more powerful predictor.

Overall the attitude toward the behavior and the subjective norm can account for a total of 8.428% of the variance in the officers intention, $\underline{F}$ (2, 56) = 2.715, $\underline{p}$ <.1. Entering each of the two predictor variables last into the regression equation points out that the subjective norm accounts for all of the significant variance in the intention. The attitude toward the behavior uniquely accounts for only 0.187% of the variance in intention, $\underline{F}$ (1, 56) = 1.12, NS. The subjective norm entered the stepwise prediction equation first and could uniquely account for 8.395% of the intentions variance, $\underline{F}$ (1, 56) = 5.41, $\underline{p}$ <.05. It appears that the officers intentions to accurately rate the performance of firefighters under their command is due solely to the influence of the officer's colleagues, co-workers, and supervisors on his normative behalf. In other words the officers will rate their firefighter's performance in the same way as they perceive is expected of them.

Some caution should be taken in interpreting the above results however. The attitude toward the behavior theoretical measure was a better predictor of the direct AB measure. But when used to predict the intention, the

AB measure essentially did not correlate with the intention. For the subjective norm, the direct and theoretical measures were not significantly related, yet the theoretical measure was able to account for a significant portion of the intention. It is not clear if this difference is due to a poor AB or a poor SN measure, or if the intention measure itself was at fault. This limits the utility of the expectancy-value theory analyses.

Chapter 4

DISCUSSION

The purpose of this study was to evaluate the differences in quality of two performance appraisal rating scale formats for the position of firefighter. Rating scale format quality was assessed by measures of reliability, sensitivity, criterion related validity on five different measures, rater and ratee scale preference attitude assessment and rater attitude-behavior consistency measures.

Reliability Reliability measures used in this study are summarized in Table 11. Cronbach's alpha reliability indicates that the BARS scale is slightly more reliable than the MSS. Both alpha reliabilities are in the eighties, however. As noted above, reliability is influenced by the distribution of scores and the number of items making up the score. The BARS had a more restricted format than the MSS. This artificially increased its reliability. Furthermore, the BARS rating format of placing an "X" over the scale value allowed for very few fluctuations in score. The MSS has three scores comprising a single dimension scale value, rather than only one. The MSS allows for the possibility of a lower reliability by permitting a greater fluctuation in scores. In any case, the reliability measures were calculated on the nine subscale scores. This allows comparability between the BARS and MSS on reliability. Although the BARS did, in fact, show a higher alpha reliability, this may have been due to artificial restrictions on the variation of scores.

The restriction of range on the nine subscales and on the total performance evaluation scores on the BARS and MSS support the above conclusion. The BARS showed a restricted range on seven of the nine subscales, whereas the MSS showed a restricted range on only five of the nine subscales. The MSS was judged to be superior on the basis of range restriction.

The leniency analyses provide the first indicator that neither format should be invoked by the fire department without some refinement on the scale items. Leniency was examined from two theoretical points of view. First, leniency was analyzed as a function of the pool of the performance appraisal total scores. Both the BARS and the MSS were discovered to contain a great deal of leniency. However, the MSS contained a statistically significant lesser amount of this between rater leniency than did the BARS. The skewness of the subscale scores was another measure of the between rater leniency definition. The single significantly positive skew of one subscale of the MSS indicates a slight superiority of the MSS. Neither format could be judged superior on the total performance evaluation score skewness distribution.

The second definition of leniency is a function of the individual rater. The mean subscale skew per subject and the mean absolute value skew per subject were analyzed as indicies of the within rater leniency definition. Both the BARS and the MSS had nonsignificant, negative mean within subject subscale skews. Neither scale format could be judged superior by this measure. However, both scales had statistically significant amounts of mean within subject skew when the absolute values of the skews were taken. This indicates that some subscales had positive skews and others showed negative skews so that in total, the two cancelled each other out. However, the BARS contained less leniency than the MSS on this mean absolute value of the subscale skewness measure and is judged to be superior.

Halo effect was assessed by rating the mean variances of the two formats within a ratee. The statistically significant larger mean variance of the MSS indicates that it contains less halo effect than the BARS. Halo effect can also be examined by calculating intercorrelations. The statistically significant lower mean interitem correlations of the MSS indicate that it also contains less halo effect than the BARS on this measure. This greater amount of halo in the BARS is reflected in the higher alpha reliability of the BARS.

The intercorrelation matricies of the two formats' subscales also provided information on convergent and discriminant validity. The formats appear to achieve convergent validity. Discriminant validity was achieved on both formats for the knowledge of the job, performance under stressful conditions and leadership dimensions. These three subscales composed the first of two ascertainable factors in the factor analyses. Thus, neither format is clearly superior based on the factor analyses.

Sensitivity Sensitivity was analyzed by the kurtosis and standard deviations of score distributions of the two formats (see Table 11). Both formats' distributions approximate unit normal distributions, thus providing reasonable amounts of discriminability between ratees. Noting the kurtoses of the individual subscales, the MSS contained two subscales (willingness to work and physical fitness) that showed peaked distributions not useful for discriminability, but the MSS initiative subscale was flat, providing for a good amount of sensitivity. The BARS teamwork and compatibility subscale showed a peaked kurtosis--indicating a poor amount of sensitivity. Thus, eight BARS subscales and seven MSS subscales had acceptable amounts of sensitivity. One MSS subscale was superior in sensitivity to all these eight BARS subscales. Neither format can be judged to have superior sensitivity on the basis of the subscale kurtoses. The mean kurtosis of the subscales across all subjects supports this conclusion. Neither format's means kurtosis was significant.

However, five of the MSS subscales displayed standard deviations found to be significantly larger than their counterparts on the BARS. The MSS was judged superior in sensitivity according to the subscale standard deviations. However, there was no difference in total score variance between the MSS and BARS.

The conclusion of the reliability and sensitivity analyses is that first, neither format was judged to be superior on six of the fourteen measures. In other words, both were equally acceptable. The MSS was judged to be superior on six measures and the BARS was judged superior on two measures. However, the BARS was superior on the alpha reliability. This has already been indicated to be, in part, due to range restriction and high leniency and halo effects.

The other measure indicating superiority of the BARS was the leniency measure based on mean absolute value of skews. This measure indicates that the BARS had a lesser overall amount of within subject severity and leniency combined.

Thus, both measures were equal in quality on seven of the indicies (including alpha reliability) and the MSS judged to be superior on six measures to the BARS single superior judgment. The MSS is recommended for use over the BARS based on the reliability and sensitivity measures.

<u>Promotional Rank Validity</u> The promotional rank order criterion regressions and all of the other five sets of criterion regressions are summarized in Table 16. The rank order listing on the fire lieutenant promotional exam could be successfully predicted by both the MSS and the BARS. Utilizing all raters and total scores, the MSS accounted for more variance in promotional rank than the BARS. This effect appears to be due mainly to the difference between the BARS and MSS scores to promotional rank given by captain raters. The captain MSS ratings accounted for 19.36% of the variance ($p < .1$) while their BARS ratings accounted for only 1.62% (NS) of the variance in promotional rank. The lieutenant MSS and BARS total scores each accounted for approximately 22% of the variance. Thus, on the basis of total scores, lieutenant ratees do equally well with either the MSS or BARS when judged on the criterion of predicting promotional exam rank order listings. The captains clearly do better with the MSS, although neither regression reached statistical significance. The result is that over all ratees the MSS total scores

are better at predicting promotional rank than the BARS total scores. This effect may be due, in part, to the small sample size of captains involved in the promotional rank criterion (N = 20).

When the promotional rank order criterion was predicted by the nine BARS and MSS subscales, both formats did equally well. The BARS initiative subscale accounted for most of the significant variance by all raters and the lieutenant raters. While the captains' nine subscale ratings accounted for 33% of the variance, this was not significant. Of the nine MSS subscales, the appearance subscale was the only significant unique contributor to the prediction of promotional rank order. Appearance on the MSS explained the majority of the variance for lieutenant raters. Again, captains' subscale scores accounted for a substantial portion (31%) of the variance, but was not statistically significant, due in part to the small sample size.

The total scores, as well as individual subscales, of both the BARS and the MSS were successful in predicting a firefighter's rank order listing on the fire lieutenant promotional exam. The total score results favor the use of the MSS. However, the initiative subscale of the BARS that predicts the criterion suggests a higher construct validity than for the MSS appearance subscale predictor of promotional rank. One would prefer to believe that promotions are based on some internal, stable quality such as initiative rather than some external, stable quality like appearance. Furthermore, it seems that lieutenants are better predictors of promotional rank order than captains, although this effect may be due to the small number of captains in this sample.

Seniority Validity The seniority criterion variable was analyzed in two different ways. First, as a continuous variable and second, as a categorical variable with six groups. As a continuous variable, seniority could not be significantly predicted by either the BARS total scores or the MSS total scores. Neither the lieutenant raters separately nor the captain raters separately could account for even one percent of the variance in the seniority criterion.

The nine MSS subscales used as predictor variables for the continuous seniority criterion variables could not predict significant portions of the variance. Neither all raters, the lieutenants alone, nor the captains alone could predict significant portions of the seniority variance. However, the nine BARS subscales did account for significant portions of the seniority variance. Three BARS subscales based on the ratings of all officers could account for significant protions of the variance: leadership, physical fitness and knowledge of the job. For lieutenants alone the knowledge of the job and physical fitness subscales accounted for significant portions of variance. The ratings based on the captains could not account for a significant portion of the variance although the physical fitness subscale did reach statistical significance by itself.

When analyzed as a continuous variable, neither the MSS nor the BARS total scores were useful in predicting seniority. However, some individual BARS subscales could account for significant portions of the seniority variance. Furthermore, the lieutenant raters again outperformed the captain raters, but only on the BARS subscales.

The seniority data was also analyzed as categorical data. In these analyses, both the BARS and MSS total scores, as well as several subscale scores contained both linear and quadratic trends. The BARS and MSS total scores both showed stronger linear trends than quadratic trends. Significant differences were discovered between group one, the youngest group, and the group(s) near the average length of seniority. The quadratic trends for both scale formats are due to a drop in performance appraisal scores in the highest seniority groups only. However, these high seniority group performance ratings do not drop to the same level as is found in the lowest seniority groups, but rather drop to a level near to the scores obtained by the average seniority groups. The MSS total scores showed a stronger quadratic trend than the BARS total scores.

The knowledge of the job, performance under stressful conditions and leadership subscales contribute to the linear and quadratic trends in both the BARS and the MSS. In addition, the BARS format also contained linear and quadratic trends in the physical fitness subscale and linear trends only in the willingness to work subscale. One may conclude that some of the significant linear findings for both formats may be due to an elimination of error variance in score fluctuations by the various seniority categories.

However, the BARS format contained three significant subscale regressions. These three subscales - knowledge of the job, physical fitness and leadership - also contain quadratic relationships to the criterion. It thus appears that the BARS subscales are more strongly linearly related to seniority than the MSS, but are also somewhat more strongly quadratically related to seniority than the MSS. Thus, both formats contain performance appraisal scores that rise over a firefighter's length of seniority and then have a tendency to diminish slightly in the highest seniority categories.

Since it is highly logical to expect a drop in performance appraisal scores for the long length of service firefighters, the format most sensitive to this effect would be preferred. The MSS total scores show a steeper inverted U-shaped curve than the BARS total scores. Both formats contain basically the same number of subscales demonstrating a U-shaped relationship with the seniority criteria. Therefore, the MSS is preferred to the BARS for the seniority criteria.

<u>Education Validity</u> The most important consideration in the analysis of the education criterion

variable is the negative correlations with both the BARS and the MSS total scores as well as several of the statistically significant subscales. Thus, the greater the amount of education a firefighter reported, the lower were his efficiency scores. However, the BARS total score accounted for more of this negatively correlated variance than did the MSS. Neither the separate lieutenant ratings nor the separate captain ratings on the MSS accounted for significant portions of the education total score variance. Both groups of raters predicted significant portions of the variance in education for the BARS format.

When considering the nine subscales, this same significant negative correlation with the education criterion occurs for the BARS format based on the ratings of all officers and approaches significance for the lieutenant group of raters. The willingness to work subscale accounts for the majority of the significant variance in both of these cases. However, for the nine MSS subscales based on the ratings of all officers, one of the significant subscales, knowledge of the job, correlates negatively with the education criterion, while the other, teamwork and compatibility, correlates positively. Neither the lieutenant nor captain raters could account for significant portions of the education variance with the nine MSS subscales.

Thus, some difficulty arises in attempting to choose which of the two rating scale formats is preferred on the basis of their relationship with the education criterion. Part of the problem may be within the education variable itself. The alpha reliability of the education criterion was only 0.62, not highly reliable. But more important, the variable may have placed too much emphasis on college academic course work rather than courses directly related to fire duty. With the boom in people obtaining college educations during the 1960's and 1970's, an inordinate number of younger firefighters may have obtained college educations. As was noted in the seniority criterion analyses above, these younger firefighters had lower performance appraisal scores than firefighters with more average lengths of seniority. It could easily be that firefighters with more average lengths of seniority, and hence higher performance appraisal scores, also had less formal college education. This would lead to the negative correlation between education and preformance appraisal scores.

Experience Validity Both the BARS and the MSS formats could account for similar significant quantities of the variance in the promotion related experience criterion variable. The BARS total score accounted for a little more than five percent of the variance while the MSS accounted for a little better than four percent. However, substantial differences occur when the total scores are analyzed separately for lieutenant and captain raters. First, for the lieutenant raters, only the BARS total score could account for significant portions of the experience variable. For the captains, the MSS total scores could account for twice as much variance as the BARS total scores. Both formats for captain raters could account for significant portions of the variance, though.

When analyzing the experience variable with respect to the nine subscales, both formats could account for significant portions of the variance with both lieutenant and captain raters. Generally, the MSS subscales accounted for more variance than the BARS subscales. For the MSS format, performance under stressful conditions was the subscale that accounted for the only unique portion of the variance for all raters combined and for lieutenant raters alone. For the captain raters on the MSS, three subscales, leadership, public relations, and teamwork and compatibility, uniquely accounted for significant portions of the variance. For the BARS subscales leadership accounted for the only significant unique portion of the variance for all ratees. For lieutenants on the BARS, leadership and knowledge of the job each accounted for significant, unique portions of the variance. Although the captains BARS subscale ratings could together account for a significant portion of the education variance, none of the individual subscales could uniquely account for significant portions of the variance.

Thus, it generally appears that both formats do acceptably well in predicting the promotion related experience criterion. However, the MSS can be considered to do slightly better on the individual subscales while both do approximately equal based on total scores. Also, captain ratees appear to do considerably better in predicting the experience criterion than do the lieutenant raters.

Performance Test Validity The most obvious result of the practical performance test regressions is that the criterion variable probably was not useful. The eight different brief practical tests measured widely differing constructs. Of the eight tests, three correlated negatively with the BARS and two of these three tests correlated negatively with the MSS. These three tests, spanners, ropes and ladders, are the basic skills required of firefighters. Practically all fire duty involves the use of these skills. With so common a set of skills, it seems extremely unlikely that the firefighters would do poorly on these skill tests. One possibility is that the firefighters who were tested on these practical performance tests were so familiar with them that they had each developed an idiosyncratic style of performance that did not meet the tough standards on the test. For example, one item in the test of the spanner wrench is the proper wearing of the tool for later quick use in a fire scene. Many firefighters did not properly wear the tool, thus lowering their score on the criterion.

In this way, the firefighters most recently graduated from the training academy could be expected to recall the department's correct procedures. They would not have developed idiosyncratic styles of wearing the spanner, using ladders or tying knots. However, the remaining practical performance tests were positively correlated with performance appraisal scores. Thus, for BARS and MSS total scores, no significant amount of variance in the practical performance test could be predicted from the ratings by all ratees, lieutenants only or captains only. Furthermore, neither the nine BARS subscales taken as a whole nor the nine MSS subscales as a whole could account for significant portions of the practical performance test variance. However, for the MSS based on all ratees, the appearance subscale correlated positively with the criterion and accounted for a significant portion of the variance. The appearance subscale also correlated positively with the captains ratings on the MSS and accounted for significant portions of the variance. For the lieutenants, however, the knowledge of the job subscale was able to account for a significant portion of the variance and correlated negatively with the criterion.

Thus, captain raters utilizing the MSS subscales were more accurate, positive predictors of the practical performance tests than were lieutenants on either format. For all raters combined, the MSS outperforms the BARS, but neither reaches acceptable significance levels.

Attitude and Scale Preference Surveys Perhaps the major consideration in choosing between the different formats of a performance appraisal system with comparable psychometric properties is the preferences noted by the raters and ratees involved. Before considering the results of the attitude and scale preference surveys, one caveat should be noted. Although the members of the Fire Department involved in this study may have a tendency to agree with the general psychometric analyses of the performance appraisal scales, their reasons for preferring one scale format over the other have more in common with their reactions against past performance appraisal systems with features similar to the new formats, than with either of the new formats themselves.

For the combined group of raters and ratees, the MSS was clearly preferred. The firefighters, when analyzed alone, also preferred the MSS to the BARS. However, officers displayed no statistically significant preference for either format. The MSS did receive a slightly more favorable rating though. These results held true for both the semantic differential attitude questionnaire and for the item in the self report survey directly calling for a preference judgement. The combined group of raters and ratees, as well as the firefighters alone also preferred the type of sentences with MSS over the BARS.

However, when called to state whether they preferred to know their numerical score immediately, as in the BARS format, or not to know the numerical score immediately, as in the MSS, the BARS type of scoring was preferred by all groups. While 170 raters and ratees preferred to know their numerical score immediately, a total of 51 preferred not to know immediately. Thus a trade-off has presented itself. The Fire Department members clearly preferred the MSS scale, in general, but not its method for assigning numerical scores.

Looking closer at this problem, two questions come to mind. First, is this result due to the fact that in the BARS, the rater places an "X" directly over a number, while on the MSS the rater does not see or use any numbers? If this is the case, then a simple recoding of the "+" "o" and "-" signs to "H", "S" and "L" (for high, same, low) respectively might be utilized. The exact same scoring routine could be used as was used in this paper. A change similar to this involving the numbers "3" "2" and "1" was made in a study by Wood, Cook and Specht (1980) without adverse results. What is important for scale usage is a clear demarcation between superior, average and below average rankings on each item.

Second, the problem with knowing the numerical score immediately may be due to the fact that the firefighters actually want to see their performance rating scores immediately following the rating session. One major consideration may make this desire impossible. The American Psychological Association (APA) Standards for Educational and Psychological Testing (1974) strongly suggest that scores of this type be standardized so that the average and standard deviation of scores assigned by any particular group of raters are the same. This standardization procedure would eliminate the possibility of having firefighters know their final numerical score immediately. However, if a firefighter's only concern is that his marks not be altered in some illegal fashion, he could easily be shown the performance appraisal instrument sheets before they leave his rater's presence to be scored and after scoring and standardization. Thus, the MSS still appears to be preferred over the BARS.

Attitude-Behavior Consistency The rater officer attitude-behavior consistency analysis is an important indication of how closely the performance appraisal scale adopted by the Fire Department will meet the standards established in this reliability and validity study. The major finding is that an officer's intention to accurately rate the performance of the firefighters under his command on the appraisal instrument of his choice is best accounted for by the officer's perception of what his colleagues and fellow workers expect of him (the subjective norm). If the department wide attitude is one that feels the new performance appraisal instrument is worthy of one's best efforts, in general, those best efforts would be given. Poor efforts result from perceived negative attitudes.

Chapter 5

RECOMMENDATIONS

Scale Format Selection In addition to conducting research to assess the various strengths and weaknesses of the BARS and MSS performance appraisal scales for firefighters, a recommendation must be made as to which format should be selected by the Fire Department. The decision is based upon the reliability, sensitivity, criterion regressions and scale preference attitude surveys.

The MSS format was judged to be superior to the BARS on the basis of the reliability and sensitivity analyses.

The results of the five criterion variable regressions were a bit more complex. For promotonal rank order prediction, the MSS out-performed the BARS on total score. Both formats were about equally acceptable based on the subscale regressions.

For the seniority criterion, the MSS total score was found to have a stronger quadratic relationship with seniority than the BARS format. The BARS subscales appear to be more strongly quadratically related to seniority.

Neither format's total score could be recommended on the basis of the education variable. However, the MSS did slightly better than the BARS based on the analysis of the nine subscales.

On the promotion related experience criterion variable, both formats explained approximately the same amount of variance based on total score. Both formats also explained significant portions of the experience variance with the nine subscales, with the MSS explaining slightly more.

Finally, the MSS subscales appear to have done slightly better in predicting the practical performance test scores than the BARS. However, no firm decisions can be based on the practical performance test criterion. Thus, the MSS is judged to be superior to the BARS on two total score criteria, of equal quality on one criterion and neither format is prefered based on the final two total score criteria. The MSS performed better or slightly better than the BARS on three subscale criterion regressions, the same on one criterion and worse than the BARS on only one criterion.

The MSS appears to be preferred over the BARS based on the set of five criterion regressions. Before put into regular use, however, the MSS must be revised and improved to try to maximize the above psychometric considerations.

The selection of the MSS format over the BARS is supported by the scale preference attitude measures. In general, the MSS was preferred over the BARS by all participants. As noted above, the firefighters strongly preferred the MSS, while the officers showed no particular preference.

Rank of Rater With the decision made to select the MSS format, two final considerations deserve some discussion. These two considerations are the relative quality of lieutenant versus captain raters and the recommendations concerning continued use of some of the individual subscales of the MSS based on the item analyses.

Due to restrictions in the collection of biographical data in order to insure anonymity of raters and ratees, the reliability and sensitivity data could not be analyzed separately for raters of different ranks. However, the criterion data did show substantial differences between lieutenant raters and captain raters.

On the promotional rank order criterion the lieutenant raters appear to have outperformed the captain raters on both total score and subscale scores of the MSS. Similar amounts of variance were accounted for by both ranks, however. The major explanation for the superiority of lieutenant raters over captain raters appears to be the small sample size of captain raters involved in the promotion rank order criterion regression. This problem limits the weight that can be placed on these results.

For the seniority criterion, when analyzed as a continuous variable, there were no substantial differences between lieutenant and captain raters. However, when analyzed separately as categorical variables, substantial differences occurred between lieutenant and captain raters. Essentially, the lieutenant raters accounted for nearly all of the statistically significant effects. Linear and quadratic effects occurred for five BARS subscales for lieutenants and two MSS subscales. No captain rater subscales for either format contained significant quadratic effects.

It appears that lieutenants outperformed captains in predicting a firefighter's seniority in a quadratic fashion with performance appraisal scores.

Although none of the education criterion variable regressions with MSS reached statistical significance, the captain raters consistently outperformed lieutenant raters. Captains were the preferred raters based both on total score and subscale scores.

The promotion related experience criterion showed that the captain raters again outperformed the lieutenant raters on the MSS. For total score, captain raters accounted for the only significant portion of the explained variance. For the nine MSS subscales, captain raters accounted for better than double the explained variance due to lieutenant raters.

This tendency for captain raters to explain more variance in the criterion variable again holds true for the practical performance test. Similar to the results of the education criterion, neither rater rank could account for significant portions of the variance. But for both total score and subscale score regressions with the MSS, captain's ratings explained just about twice as much variance as did lieutenant's ratings.

It thus appears that captain raters are somewhat better predictors of three of the five criterion variables in this study. Lieutenants are able to account for more variance in two criterion variables. Thus, captains would appear to be a bit more preferred as raters than lieutenants.

However, several considerations must be taken into account before a decision to recommend captain raters should be made. First, each of the officers involved in this study rated the firefighters under their own command. Due to the structure of the fire department involved in this study, about one-third of the companies were lead by captain raters, while the remainder were lead by lieutenants. The results suggest that captains are better raters of the firefighters under a captain's command than lieutenants are for the companies lead by lieutenants. These results do not show that captains are better raters overall, just for the men under their personal command.

Second, it is unlikely that captains could be expected to triple their current efforts in performance appraisal by assuming the rating duties of the lieutenants under their command. The amount of paper work involved would be prohibitive. Currently each company officer is responsible for rating the performance of the firefighters under his command.

One speculation that appears to be permissible is that what is of importance in the rank of rater analysis is the improvement of rating quality as an increasing function of rank of field officers. If this is the case, then one might expect the next higher rank of officer, battalion chief, to be a better rater of performance than captains. This possibility will be taken into consideration in the next version of the MSS performance appraisal scale. If this speculation is supported, then two raters per firefighter, a battalion chief and the company officer, might be possible. The battalion chief supervises the company officer's work.

Subscale Selection The final consideration for the revision of the MSS presented in this paper concerns which of the subscales should be included. The best source of information for this decision is found in the item-analysis data presented in Table 4. The overall alpha reliability is .83169. In the last column of Table 4 the numbers listed are the scale alpha reliability if the particular subscale dimension were deleted. It appears that if the initiative item were deleted the alpha would rise to .8377. Furthermore, the item-total correlation for the initiative subscale is the lowest of the set, .309. Thus, it is recommended that the initiative subscale be revised or deleted from the next version of this scale.

Noting the "scale variance if item deleted" column, it appears that the knowledge of the job, performance under stressful conditions and appearance subscales account for large portions of the variance in total score. If possible, these are items that should be considered to be expanded on to raise the number of items in the scale. This would result in a higher alpha reliability for the scale. One possibility is to expand the knowledge of the job dimension to state specifically the various components of job knowledge (e.g., ladder operations and hose operations, etc.) as were found in the Bownas, Heckman and Anderson (1977) study.

Reference Notes

1. Groner, D.M. <u>Reliability and susceptibility to bias of behavioral and graphic rating scales</u>. Unpublished doctoral dissertation, University of Minnesota, 1974.
2. Katcher, B.L. & Bartlett, C.J. <u>Rating errors of inconsistency as a function of dimensionality of behavioral anchors</u>. Unpublished Masters Thesis, University of Maryland, 1977.
3. Landy, F.J. & Barnes, J. <u>Scaling behavioral anchors</u>. Paper presented at the meeting of the American Psychological Association, Toronto, August, 1978.
4. Saal, F.E. <u>The mixed standard rating scale: A consistent system for numerically coding inconsistent response combinatinons</u>. Paper presented at the meeting of the Midwest Psychological Association, Chicago, May, 1978.
5. Saal, F.E. <u>Development of Job-Related Rating Scales for Evaluating the Performance of Police Patrol Officers in the City of Topeka, Kansas</u>.

Manhattan, Kansas: Kansas State University, Department of Psychology, January, 1979.

6. Schwind, H.F. Job behavior sampling: A method to increase the information content of performance feedback. Paper presented at the meeting of the American Psychological Association, Toronto, August, 1978.

References

American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. Standards for Educational and Psychological Tests. Washington, D.C. American Psychological Association, 1974.

Anastasi, A. Psychological Testing (4th ed.). New York: MacMillan, 1976.

Arvey, R.A. & Hoyle, J.C. A Guttman approach to the development of behaviorally based rating scales for system analysis and programmer analysts. Journal of Applied Psychology, 1974, 59, 61-68.

Bernardin, H.J. Behavioral expectation scales versus summated scales: A fairer comparison. Journal of Applied Psychology, 1977, 62, 422-428.

Bernardin, H.J., Alveres, K.M., & Cranny, C.J. A recomparison of behavioral expectation scales to summated scales. Journal of Applied Psychology, 1976, 61, 564-570.

Blanz, F., & Ghiselli, E.E. The mixed standard scale: A new rating system. Personnel Psychology, 1972, 25, 185-199.

Blood, M.R. Spin-offs from behavioral expectation scale procedures. Journal of Applied Psychology. 1974, 59, 513-515.

Borman, W.C. & Dunnette, M.D. Behavior based versus trait-oriented performance ratings: An empirical study. Journal of Applied Psychology, 1975, 60, 561-565.

Borman, W.C. & Vallon, W.R. A view of what can happen when behavioral expectation scales are developed in one setting and used in another. Journal of Applied Psychology, 1974, 59, 197-201.

Bownas, D.A. & Heckman, R.W. Job analysis of the entry-level firefighter position. (Technical Study 77-8) Minneapolis, Minnesota: Personnel Decisions Inc., December 1976. (NTIS No. PB271-534/AS).

Burnaska, R.F. & Hollman, T.D. An empirical comparison of the relative effects of rater response biases on three rating scale formats. Journal of Applied Psychology, 1974, 59, 307-312.

Campbell, J.P., Dunnette, M.D., Arvey, R.D., & Hellervic, L.V. The development and evaluation of behaviorally based rating scales. Journal of Applied Psychology, 1973, 57, 15-22.

Campbell, J.P., Dunnette, M.D., Lawler, E.E., III, & Weick, K.E., Jr. Managerial Behavior Performance and Effectiveness. New York: McGraw-Hill, 1970.

Cronbach, L.J. Essentials of Psychological Testing (3rd ed.). New York: Harper and Row, 1970.

Cummings, L.L. & Schwab, D.P. Performance in Organizations: Determinents and Appraisal. Glenview, Illinois: Scott, Foreman, 1973.

Davis, P.O. & Wright, Major H.F. Elements of a recruit selection process. Fire Command, January 1979, 12-13.

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

Dunnette, M.D. Personnel Selection. Belmont, California: Wadsworth, 1966.

Dunnette, M.D., & Borman, W.C. Personnel Selection and Classification Systems. In M.R. Rosenzweig & L.W. Porter (Eds.), Annual Review of Psychology (Vol. 30). Palo Alto, California: Annual Review, 1979.

Finley, D.M., Osborn, H.G., Dubin, J.A., & Jeanneret, P.R. Behaviorally based rating scales: Effects of specific anchors and disguised scale continua. <u>Personnel Psychology</u>, 1977, <u>30</u>, 659-669.

Fishbein, M. & Ajzen, I. <u>Beliefs, Attitudes, Intentions and Behavior: An Introduction to Theory and Research</u>. Reading, Massachusetts: Addison-Wesley, 1975.

Flanagan, J.C. The Critical Incident Technique. <u>Psychological Bulletin</u>, 1954, <u>51</u>, 327-358.

Keaveny, T.J. & McGann, A.F., A comparison of behavioral expectation scales. <u>Journal of Applied Psychology</u>, 1975, <u>60</u>, 695-703.

Matticks, K.R. Are you promoting the best man? <u>Fire Command</u>, September 1977, 51-52.

Nunnally, J.C. <u>Psychometric Theory</u>. New York: McGraw-Hill, 1967.

Personnel Decisions Inc. <u>Firefighter Task Checklist</u> (Technical Study 77-8A) Minneapolis, Minnesota: Personnel Decisions Inc., December 1976. (NTIS No. PB271-534/AS).

Saal, F.E. & Landy, F.J. The mixed standardrating scale: An evaluation. <u>Organizational Behavior and Human Performance</u>, 1977, <u>18</u>, 19-35.

Schwab, D.P., Heneman, H.G., & DeCotiis, T.A. Behaviorally anchored rating scales: A review of the literature. Personnel Psychology, 1975, 28, 549-562.

Selltiz, C., Wrightsmen, L.S., & Cook, S.W. Research Methods in Social Relations (3rd ed.). New York: Holt, Rinehart and Winston, 1976.

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, 147-155.

Sparrs, J.E. & McKee, J.B. Assessment centers and accountability. Fire Command, January 1979, 16-17.

Van Rijn, P. Job analysis of entry-level firefighting in the District of Columbia Fire Department: A task/duty approach (Technical Memorandum 77-5, Test Services Section, Personnel Research and Development Center, United States Civil Service Commission). Washington, D.C.: U.S. Government Printing Office, 1977a.

Van Rijn, P. Job analysis of entry-level firefighting in the District of Columbia Fire Department:

Supervisory perception of the entry level job. (Technical memorandum 77-8, Test Service Section, Personnel Research and Development Center, United States Civil Service Commission). Washington, D.C.: U.S. Government Printing Office, 1977b.

Vincent, J.P. & Knowles, L. Using performance standards to maintain manipulative skills. Fire Command, January 1979, 14-15.

Wenger, D.R. Survival of the Fittest. Fire Command, August 1978, 42-43.

Zedeck, S., Katry, D. & Jacobs, R. Format and scoring variations in behavioral expectation evaluations. Organizational Behavior and Human Performance, 1976, 17, 171-184.

Appendix A

Firefighter and Officer Semantic Differential Attitude Scale

Directions for Completing the Firefighter and Lieutenant Efficiency Mark Attitude Scale.

The following scales will assess your opinions toward the two new efficiency mark rating scales. After the attitude scales, you will answer several items specifically asking which of the two formats of the efficiency mark rating scale you would prefer using, plus other information required for statistical analysis of the scales. Your responses will be taken into consideration when the final format of the efficiency mark rating scale is decided.

The first seven items will require you to indicate how strongly you feel that the mixed Standard Rating Scale (the plus, zero, minus check mark scale) is, for example, good and useful. Thus, if you think the mixed Standard Rating Scale is "slightly good", you would mark the attitude scale as follows:

good	______ :	______ :	X :	______ :	______ :	______ :	______	bad
	extremely	moderately	slightly	neutral	slightly	moderately	extremely	

Continue down the list of seven items making "X's" on each scale in a position that reflects your attitude. Please be careful to note whether the right side or the left side of each scale is positive. Sometimes placing an "X" near the right endpoint will indicate a positive attitude, at other times an "X" near the left endpoint will indicate a positive attitude.

The scale on the second page asks the very same questions regarding the other rating scale (the Behaviorally Anchored Rating Scale). Please make your responses in the same way as on the preceding scale.

Finally, several separate questions will be found directly asking for your preferences to the scales. Make all of your answers accurate. If you have any questions regarding what you should do, please feel free to ask.

FIREFIGHTER AND LIEUTENANT ATTITUDES TOWARD TWO
PROPOSED FIREFIGHTER EFFICIENCY MARK SCALES

The Proposed New Mixed Standard Rating Scale
(The plus, zero or minus checkmark scale)

	extremely	moderately	slightly	neutral	slightly	moderately	extremely	
good	_______ :	_______ :	_______ :	_______ :	_______ :	_______ :	_______ :	bad
hard to use	_______ :	_______ :	_______ :	_______ :	_______ :	_______ :	_______ :	easy to use
useful	_______ :	_______ :	_______ :	_______ :	_______ :	_______ :	_______ :	useless
relevant to efficiency grading	_______ :	_______ :	_______ :	_______ :	_______ :	_______ :	_______ :	irrelevant to efficiency grading
personally liked the scale	_______ :	_______ :	_______ :	_______ :	_______ :	_______ :	_______ :	personally disliked the scale
liked by the Firefighters	_______ :	_______ :	_______ :	_______ :	_______ :	_______ :	_______ :	disliked by the Firefighters

THE PROPOSED NEW BEHAVIORALLY ANCHORED RATING SCALE

(The "X" in one of Seven Boxes scale)

	extremely	moderately	slightly	neutral	slightly	moderately	extremely	
good	________ :	________ :	________ :	________ :	________ :	________ :	________ :	bad
hard to use	________ :	________ :	________ :	________ :	________ :	________ :	________ :	easy to use
useful	________ :	________ :	________ :	________ :	________ :	________ :	________ :	useless
relevant to efficiency grading	________ :	________ :	________ :	________ :	________ :	________ :	________ :	irrelevant to efficiency grading
personally liked the scale	________ :	________ :	________ :	________ :	________ :	________ :	________ :	personally disliked the scale
liked by the Firefighters	________ :	________ :	________ :	________ :	________ :	________ :	________ :	disliked by the Firefighters

1. Which of the two proposed new rating scales do you prefer to see adopted by the Chicago Fire Department? (check off one)

   _________ Behaviorally Anchored Rating Scale ("X" in box).

   _________ Mixed Standard Rating Scale (Plus, zero, minus).

   _________ No Efficiency Scale Should Be Used.

2. Would you prefer to immediately know the numeral score you are assigning each Firefighter? (check off one)

   _________ Yes, would prefer to know the numerical score immediately.

   _________ No, would not want to know the numerical score.

3. The sentences in the Behaviorally Anchored Rating Scale ("X" in box) are exact copies of items written by Firefighters in the Chicago Fire Department. The sentences in the Mixed Standard Rating Scale (plus, zero, minus) are composites of many of the above sentences written by a senior member of the Fire Department with the help of the research team. Not considering the formats of the two scales - the way they look - which type sentences do you prefer?

   _________ I prefer the sentences in the Mixed Standard Scale.
   _________ I prefer the sentences in the Behaviorally Anchored Rating Scale.
   _________ Both types are good.
   _________ Neither type is good.

Appendix B

Firefighter Self-Report Questionnaire

SELF-REPORT QUESTIONNAIRE

CHICAGO FIREFIGHTER PERFORMANCE
APPRAISAL PROJECT

Name: ______________________________ File Number: __________

Social Security Number: ________________

1) Have you taken the firefighter I certification examination?

_____ Yes _____ No

If yes, what was your score? _____

2) Have you taken the Firefighter II certification examination?

_____ Yes _____ No

If yes, what was your score? _____

3) Do you have a Basic Instructor certificate?

_____ Yes _____ No

5) Do you have a Standard Instructor certificate?

_____ Yes _____ No

6) Did you graduate from high school?

_____ Yes _____ No

7) Have you taken any college courses?

_________ None

_________ Some courses, but no degree

_________ Received an associates degree

_________ Received a bachelor's degree

8) How many college credit fire science courses have you taken? _____

9) How many not - for-credit fire science courses have you taken? _____

10) How often during the last six months have you performed any of the following. (Check off frequency)

	Never	once or twice	more than twice
Drive on apparatus	____	____________	______________
Buggy Driver	____	____________	______________
Acting Lieutenant	____	____________	______________
Acting Fire Engineer	____	____________	______________

11) Have you been a Fire academy Instructor during the last 6 month?

______ Yes _______ No

12) Have you been included in the Chief of Fire Services Merit Roll in the last six months?

______ Yes _______ No

Appendix C

Firefighter Nine Practical Performance Tests

Name________________________________ File #__________
(Last) (First)

Company________________ Platoon________ Date______________

Division________ Battalion_______ Daley Day______ Date of Entry____________

Social Security #__________________ Company Officer____________________

Test (Outside) Point of Vantage Leadout

OBJECTIVE: To test the candidate's ability to select the equipment necessary for the point of vantage operation.

To identify equipment by name and their connection diameter and nozzle size.

To set up layout for operation

Select necessary equipment from following list of displayed items
- Siamese - Divider - 2-1/2"-1-1/2" Reducer - 3-1/2"-2-1/2" Reducer
- 2-1/2"-3-1/2" increasor - Wall Hook - Spanner - 3" hose - 2-1/2" hose
- 1-1/4" S.O.F. - 1-1/4" street pipe - 2-1/2" bell fog

Selection		Identify	Size
3-1/2-2-1/2" Reducer	☐ - 4 pts.	☐ - 4 pts.	☐ - 4 pts.
Siamese	☐ - 3 pts.	☐ - 3 pts.	☐ - 3 pts.
1-1/4" Street Pipe	☐ - 6 pts.	☐ - 3 pts.	☐ - 3 pts.
3" hose	☐ - 3 pts.	☐ - 3 pts.	☐ - 3 pts.
2-1/2" hose	☐ - 3 pts.	☐ - 3 pts.	☐ - 3 pts.
1-1/4" S.O.P.	* ☐ 0 pt.	☐ 3 pts.	☐ 3 pts.
Wall Hook	☐ 6 pts.	☐ 3 pts.	___ Total Points__________ (60 maximum)

Application (Laid out in following order)

2-1/2" - 3" hose (to base)	☐ - 8 pts.
Siamese w/reducer (at base)	☐ - 8 pts.
2 lap-3" hose	☐ - 8 pts.
1-1/4" street pipe	☐ - 8 pts.
1-1/4" S.O.P.	☐ - 8 pts.
Wall Hook (3 wraps)	☐ - 8 pts.

Total Points__________
(40 maximum)

☐ PASS

☐ FAIL

Evaluator's Name ______________________________

Final Score__________

NAME ______________________________ FILE # ____________
(Last) (First)

COMPANY ________ PLATTOON ________ DATE ____________

DIVISION ______ BATTALION ______ DALEY DAY ______ DATE OF ENTRY ____________

SOCIAL SECURITY # ____________ CO. OFFICER ____________

TEST (SPANNER)

OBJECTIVE: TO TEST THE CANDIDATE'S ABILITY TO DEMONSTRATE THE PROPER MANNER IN WHICH A SPANNER IS WORN IN ROUTE TO A FIRE. TO DEMONSTRATE THE PROPER METHOD IN PLACING SPANNER ON A CHARGED LINE AND TO TAKE THE PROPER HEELING POSITION.

SPANNER WORN ON CORRECT SHOULDER (WRENCH HOOK FACING DOWNWARD)	☐	- 15 Pts.
SPANNER RING BROUGHT UNDER HOSE FROM SIDE CANDIDATE IS STANDING	☐	- 15 Pts.
DOUBLE WRAP USED ON LINE	☐	- 15 Pts.
"SPANNERS IN" APPX. 18" FROM BUTT	☐	- 15 Pts.
TAKES POSITION OPPOSITE OF PIPEMAN	☐	- 15 Pts.
TAKES PROPER HEEL STANCE WITH SPANNER ON OUTSIDE SHOULDER	☐	- 25 Pts.

TOTAL SCORE ____________

____
EVALUATOR'S NAME

NAME ______________________________ FILE # ______________________
(Last) (First)

COMPANY ____________ PLATOON ____________ DATE ______________________

DIVISION______ BATTALION______ DALEY DAY______ DATE OF ENTRY ______________

SOCIAL SECURITY # ______________ CO. OFFICER ______________________

TEST (ONE MAN LADDER OPERATION)

OBJECTIVE: TO TEST THE CANDIDATE'S ABILITY TO PROPERLY CARRY A 24 FOOT EXTENSION LADDER (ONE MAN CARRY) TO DESIGNATED LOCATION - PLACE LADDER IN CORRECT POSITION - RAISE LADDER AND EXTEND FULLY AND PLACE HEEL PROPER DISTANCE FROM BUILDING. CONTROL OF LADDER MUST BE MAINTAINED THROUGHOUT RAISE TO WARRANT POINTS ON SPECIFIC ITEM.

CARRYING LADDER PROPERLY (ON EITHER SHOULDER - ARM PASSING THROUGH LADDER AT MIDDLE OF OF LADDER'S LENGTH)	☐	- 10 Pts.
PROPER RAISING POSITION (FLY FACING UP)	☐	- 10 Pts.
(PLACE HEEL AGAINST BUILDING)	☐	- 10 Pts.
(FACE TOP OF LADDER OR GRASP TOP OF LADDER)	☐	- 10 Pts.
GRASP LADDER FIRMLY	☐	- 10 Pts.
MOVE HEEL OUT APPX. 1 FOOT FROM BUILDING	☐	- 10 Pts.
PLACE FOOT OUTSIDE BEAM SECURE KNEE AGAINST BEAM	☐	- 10 Pts.
RAISE FLY TO FULL EXTENSION	☐	- 10 Pts.
SET LADDER RIGHT SIDE OF OPENING	☐	- 10 Pts.
SET HEEL FOR PROPER CLIMBING ANGLE (APPX. 6 FEET)	☐	- 10 Pts.

TOTAL SCORE ______________

____
EVALUATOR'S NAME

NAME ______________________________ FILE # ______________________
(Last) (First)

COMPANY ________ PLATOON ________ DATE ______________________

DIVISION ____ BATTALION ____ DALEY DAY ____ DATE OF ENTRY ______________

SOCIAL SECURITY # ______________ CO. OFFICER ______________________

TEST (ONE MAN LADDER OPERATION)

OBJECTIVE: TO TEST THE CANDIDATE'S ABILITY TO PROPERLY CARRY A 24 FOOT EXTENSION LADDER (ONE MAN CARRY) TO DESIGNATED LOCATION - PLACE LADDER IN CORRECT POSITION -- RAISE LADDER AND EXTEND FULLY AND PLACE HEEL PROPER DISTANCE FROM BUILDING. CONTROL OF LADDER MUST BE MAINTAINED THROUGHOUT RAISE TO WARRANT POINTS ON SPECIFIC ITEM.

CARRYING LADDER PROPERLY (ON EITHER SHOULDER - ARM PASSING THROUGH LADDER AT MIDDLE OF OF LADDER'S LENGTH)	☐	- 10 Pts.
PROPER RAISING POSITION (FLY FACING UP)	☐	- 10 Pts.
(PLACE HEEL AGAINST BUILDING)	☐	- 10 Pts.
(FACE TOP OF LADDER OR GRASP TOP OF LADDER)	☐	- 10 Pts.
GRASP LADDER FIRMLY	☐	- 10 Pts.
MOVE HEEL OUT APPX. 1 FOOT FROM BUILDING	☐	- 10 Pts.
PLACE FOOT OUTSIDE BEAM SECURE KNEE AGAINST BEAM	☐	- 10 Pts.
RAISE FLY TO FULL EXTENSION	☐	- 10 Pts.
SET LADDER RIGHT SIDE OF OPENING	☐	- 10 Pts.
SET HEEL FOR PROPER CLIMBING ANGLE (APPX. 6 FEET)	☐	- 10 Pts.

____

TOTAL SCORE ______________

____
EVALUATOR'S NAME

NAME ______________________________ FILE # ______________________
(Last) (First)

COMPANY __________ PLATOON __________ DATE ______________________

DIVISION_____ BATTALION_____ DALEY DAY_____ DATE OF ENTRY ______________

SOCIAL SECURITY #______________ CO. OFFICER ______________________

TEST (ONE MAN LADDER OPERATION)

OBJECTIVE: TO TEST THE CANDIDATE'S ABILITY TO PROPERLY CARRY A 24 FOOT EXTENSION LADDER (ONE MAN CARRY) TO DESIGNATED LOCATION - PLACE LADDER IN CORRECT POSITION -- RAISE LADDER AND EXTEND FULLY AND PLACE HEEL PROPER DISTANCE FROM BUILDING. CONTROL OF LADDER MUST BE MAINTAINED THROUGHOUT RAISE TO WARRANT POINTS ON SPECIFIC ITEM.

CARRYING LADDER PROPERLY (ON EITHER SHOULDER - ARM PASSING THROUGH LADDER AT MIDDLE OF OF LADDER'S LENGTH)	☐	- 10 Pts.
PROPER RAISING POSITION (FLY FACING UP)	☐	- 10 Pts.
(PLACE HEEL AGAINST BUILDING)	☐	- 10 Pts.
(FACE TOP OF LADDER OR GRASP TOP OF LADDER)	☐	- 10 Pts.
GRASP LADDER FIRMLY	☐	- 10 Pts.
MOVE HEEL OUT APPX. 1 FOOT FROM BUILDING	☐	- 10 Pts.
PLACE FOOT OUTSIDE BEAM SECURE KNEE AGAINST BEAM	☐	- 10 Pts.
RAISE FLY TO FULL EXTENSION	☐	- 10 Pts.
SET LADDER RIGHT SIDE OF OPENING	☐	- 10 Pts.
SET HEEL FOR PROPER CLIMBING ANGLE (APPX. 6 FEET)	☐	- 10 Pts.

____

TOTAL SCORE ______________

____
EVALUATOR'S NAME

NAME ______________________________ FILE # ____________________
(Last) (First)

COMPANY ____________ PLATOON ____________ DATE ____________________

DIVISION ______ BATTALION ______ DALEY DAY ______ DATE OF ENTRY ____________________

SOCIAL SECURITY # ____________ COMPANY OFFICER ____________________

TEST UTILIZING TWO 1-1/2" LINES FOR OVERHAUL

OBJECTIVE: TO TEST THE FIREFIGHTER'S ABILITY TO SELECT THE EQUIPMENT REQUIRED TO UTILIZE TWO 1-1/2" LINES FROM THE 1-1/4" S.O.P. FOR THE PURPOSE OF OVERHAUL. THE FIREFIGHTER IS EXPECTED TO SELECT, NAME AND IDENTIFY, AND SET UP THE LAYOUT IN ITS PROPER ORDER FOR OPERATION.

SELECT NECESSARY EQUIPMENT FROM FOLLOWING LIST OF DISPLAYED ITEMS.

- SIAMESE - DIVIDER - 2-1/2" x 1-1/2" REDUCER -
- 1-1/2" x 2-1/2" INCREASER - 3-1/2" x 2-1/2" REDUCER -
- 3" HOSE - 2-1/2" HOSE - TWO 1-1/2" HOSE - 1-1/4" S.O.P.
- 3/4" S.O.P. - 1-1/2" ADJUSTABLE FOG - DOUBLE 2-1/2" MALE -
- DOUBLE 2-1/2" FEMALE -

SELECTION -	- IDENTIFY -		- SIZE -
1-1/4" S.O.P.	☐ - 6 Pts.	☐ - 4 Pts.	☐ - 2 Pts.
1-1/2" x 2-1/2" INCREASER	☐ - 6 Pts.	☐ - 4 Pts.	☐ - 2 Pts.
2-1/2" x 1-1/2" x 1-1/2" DIVIDER	☐ - 4 Pts.	☐ - 4 Pts.	☐ - 4 Pts.
TWO - 1/2" LINES	☐ - 2 Pts.	☐ - 2 Pts.	☐ - 0 Pts.
3/4" S.O.P.	☐ - 2 Pts.	☐ - 2 Pts.	☐ - 2 Pts.
ADJUSTABLE FOG	☐ - 2 Pts.	☐ - 2 Pts.	☐ - 2 Pts.

TOTAL POINTS ____________ (52 Max.)

APPLICATION. (LAID OUT IN FOLLOWING ORDER.)

1-1/4" S.O.P. USED ON 2-1/2" LINE	☐ - 8 Pts.
1-1/2" x 2-1/2" INCREASER ON 1-1/4" S.O.P.	☐ - 8 Pts.
DIVIDER ON INCREASER	☐ - 8 Pts.
TWO LINES OF 1-1/2" ON DIVIDER	☐ - 8 Pts.
3/4" NOZ. ON ONE 1-1/2" LINE	☐ - 8 Pts.
ADJ. FOG NOZ. ON ONE 1-1/2" LINE	☐ - 8 Pts.

TOTAL POINTS ____________ (48 Max.)

☑ PASS

☒ FAIL EVALUATOR'S NAME ______________________________

Name ______________________________ File# ____________________
(Last) (First)

Company __________ Platoon __________ Date ____________________

Division ________ Battalion ________ Daley Day _____ Date of Entry __________

SOCIAL SECURITY # __________________ Company Officer __________

Test <u>ROPES</u>

OBJECTIVE: To determine the candidate's ability to <u>tie</u> the listed knots using the prescribed methods of <u>the</u> C.F.D. Manual.

<u>Clove Hitch</u> (use rope bars in drill hall)

Ties knot properly in reasonable time	☐	25 pts.
Ties knot properly - demonstrates uncertainty	☐	15 pts.
Unable to properly tie knot	☐	0 pts.

<u>Bowline 'TO'</u>

Ties knot properly in reasonable time	☐	25 pts.
Ties knot properly - demonstrates uncertainty	☐	15 pts.
Unable to properly tie knot	☐	0 pts.

<u>Bowline 'FROM'</u>

Ties knot properly in reasonable time	☐	25 pts.
Ties knot properly - demonstrates uncertainty	☐	15 pts.
Unable to properly tie knot	☐	0 pts.

<u>Draw Knot</u>

Ties knot properly in reasonable time	☐	25 pts.
Ties knot properly - demonstrates uncertainty	☐	15 pts.
Unable to properly tie knot	☐	0 pts.

☐ = Pass Total points __________

☐ = Fail Evaluator ____________________

Name ______________________________ File# ______________
(Last) (First)

Date ______________

Company __________ Platoon __________

Division ________ Battalion ________ Daley Day ________ Date of Entry ______________

SOCIAL SECURITY # ____________________ Company Officer ______________

Test LADDERS

OBJECTIVE: To test candidate's ability to take a proper position (hands and feet) in setting heel. To properly tie the safety hitch on an extension ladder. To take proper beamman position, (hands and feet) stabilizing ladder in avertical position.

Heel Set Position

Correct hand position	☐	10 pts.
Correct feet position	☐	10 pts.
Placement for proper climbing angle.	☐	15 pts.

Safety Hitch Tie

Ties knot properly in reasonable time	☐	35 pts.
Ties knot properly - with uncertainty	☐	15 pts.
Unable to tie knot properly	☐	0 pts.

Beam Man Position

Correct hand position	☐	10 pts.
Correct feet position	☐	10 pts.
Watches top of ladder	☐	10 pts.

☐ = Pass

☐ = Fail

Total Points __________

Evaluator ______________

Name ______________________________ File# _______________
(Last) (First)

Company ________ Platoon ____________ Date _______________

Division ______ Battalion _______ Daley Day _______ Date of Entry _______

SOCIAL SECURITY # ______________ Company Officer ______________

Test <u>SMALL TOOLS & FITTINGS</u>

OBJECTIVE: To test candidate's ability to <u>identify</u> and <u>name</u> connection diameters of each tool or fitting on display and describe the use/s of each.

Identify		Size		Use	
Siamese	☐ 3 pts	2-2½ FM to 3½ Male	☐ 8 pts	To combine 2 hose line into one	☐ 5 pts
				To make up improvised syphone	☐ 5 pts
Divider	☐ 3 pts	2½ FM to 2-1½ Males	☐ 8 pts	To divide one large line (2½ or 3) into 2-1½ line	☐ 5 pts
Reducer	☐ 3 pts	3½ FM to 2½ Male	☐ 8 pts	Used on a siamese to accept a 2½ Female fitting	☐ 5 pts
Increaser	☐ 3 pts	1½ FM to 2½ Male	☐ 8 pts	To accept a 2½ Male fitting off an 1½ S.O.P. or 1½ hose	☐ 5 pts
Chicago Valve	☐ 3 pts	4½ FM to 4½ FM	☐ 8 pts	To allow a 2nd suction to be attached to a flowing hydrant	☐ 5 pts
Displays Certainty	☐ 2 pts		☐ 8 pts		☐ 5 pts
Points ____ (17 max)		Points ____ (48 Max)		Points ____ (35 max)	

☐ = Pass

☐ = Fail Evaluator ____________________ Total Points ________

Name ______________________________________ File# ____________________
(Last) (First)

Company ____________ Platoon ____________ Date ____________________

Division ________ Battalion ________ Daley Day ________ Date of Entry ____________

SOCIAL SECURITY # ____________________ Company Officer ______________

Test <u>SECURING CHARGED HOSE LINE IN A CIRCLE</u>

OBJECTIVE: To test the candidate's ability to <u>place</u> the hose nozzle in proper position on a length of hose and <u>secure</u> hose using the appropriate spanner tie <u>demonstrating</u> a reasonable certainty.

Placement of nozzle on hose line	☐	20 pts.
Provide 15 foot circle for heeling	☐	20 pts.
Correct spanner tie in securing	☐	40 pts.
Demonstrate certainty in above procedures	☐	20 pts.

Total points ____________________

☐ = Pass

☐ = Fail

Evaluator ____________________

Name ____ File# ______________
(Last (First)

Company: ______________ Platoon ______________ Date ______________

Division ______________ Battalion ______________ Daley Day ______________ Date of Entry ______________

Social Security # ______________ Company Officer ______________

Test DISTRIBUTOR NOZZLE LAYOUT

OBJECTIVE: To test the candidate's ability to select the equipment, and other or required items necessary to place a distributor nozzle into operation. Name and identify equipment and their connection diameters of each small tool. To set up the layout in its proper working order.

Candidate will select necessary items from following list of displaced items.

- Siamese - divider - 2½" x 1½" reducer - 3½" x 2½" reducer
- 1½" x 2½" increaser - 3" hose - 1½" hose - distributor nozzle
- chair or other suitable object

2½" length of hose with an 1½' S.O.P. attached will be provided

Selection		Points	Identity			Size		Points
1½" x 2½" increaser	☐	- 10 pts.	Increaser	☐	- 5 pts.	1½" x 2½"	☐	5 pts.
3" Length of Hose	☐	- 10 pts.	3" Hose	☐	- 5 pts.	2½"	☐	5 pts.
Distributor Nozzle	☐	- 10 pts.	Distrib. Nozzle	☐	- 5 pts.	2½"	☐	5 pts.
Chair or other object	☐	- 10 pts	1½" S.O.P.	☐	- 5 pts.	1½"	☐	5 pts.
Total points	☐		Total Points	☐		Total Points	☐	

Application Laid out in (following order)			Points Scored	
1½" x 2½" increaser	☐	- 5 pts.	Selection	☐
3" Hose	☐	- 5 pts.	Identification	☐
Chair or other object	☐	- 5 pts.	Sizes	☐
Distributor Nozzle	☐	- 5 pts.	Application	☐
Total Points	☐		Total Score	☐

☐ PASS

☐ FAIL

Evaluator's Name ______________________________

Appendix D

Fishbein Expectancy Value Attitude Survey for Officers

Directions For The Fire Lieutenant Attitude Survey

The following forty-six (46) question attitude survey will be a detailed investigation into your thoughts, beliefs, opinions and intentions toward using the efficiency mark scale that you prefer the most. You have graded all of the Firefighters under your command on each of the two scales. Answer the following questions with respect to the efficiency mark scale that you would prefer using.

The questions on the following attitude scale are divided into several different sections. However, for each question, your response will be to place an "X" on one of seven (7) blank spaces in between the two adjective endpoints of the scale.

The first question will ask you to specify your intention to accurately use the efficiency mark scale you prefer. Your response will be to indicate how likely it is that you will accurately fill out that scale. Thus, if you believe that it is "moderately likely" that you will accurately complete the efficiency mark scale, you will complete the scale as follows:

I intend to accurately fill out the efficiency mark scale.

likely	______ :	X :	______ :	______ :	______ :	______ :	______	unlikely
	extremely	moderately	slightly	neutral	slightly	moderately	extremely	

Questions two (2) through seven (7) will ask you to give your opinion toward accurately filling out the efficiency mark rating scale. You will be asked to rate how strongly you feel either for or against the scale. The format will be similar to the one above. For example, if you feel that accurately completing the efficiency mark scale is "slightly good," then you would mark the second question as follows:

good	______ :	______ :	X :	______ :	______ :	______ :	______	bad
	extremely	moderately	slightly	neutral	slightly	moderately	extremely	

The third section, beginning with question eight (8), will alternate between asking you to rate the likelihood of some statement being true and questions asking you to rate whether that statement is good or bad.

For example, you will be asked to rate the likelihood that it is true that accurately completing the efficiency mark scale will lead to improved performance of Firefighters. You will also be asked to rate how good or bad it is to improve the performance of Firefighters.

The fourth and final section, beginning with question twenty-eight (28) will ask you to rate how likely it is that several different groups of fellow workers want you to accurately complete the efficiency mark scale. Then for each of these groups of people you will be asked to rate how likely it is that you want to comply with these people by doing what they want you to do.

Please take your time and answer each question honestly and accurately. If you have negative feelings about some item, mark those feelings down. If you feel quite positively about another item, make a clearly positive response. Although you will not be personally identified, your responses will be combined with the responses of the other Lieutenants in this pilot test and reported to the Commissioners of the Fire Department and Personnel Department. Your opinion will also help decide whether the entire project to create a new efficiency mark scale has been useful and accurate. Thank you for your cooperation.

FIRE LIEUTENANT ATTITUDE SURVEY

1) I intend to fill out the performance appraisal (efficiency mark) scale accurately for each Firefighter under my command, knowing that they will find out the marks given to them and that I should not give everybody the same high grade.

likely ________ : ________ : ________ : ________ : ________ : ________ : ________ : unlikely
extremely moderately slightly neutral slightly moderately extremely

* *

My attitude or opinion is that accurately filling out the efficiency mark rating scale is:

2) good ________ : ________ : ________ : ________ : ________ : ________ : ________ : bad
extremely moderately slightly neutral slightly moderately extremely

3) hard to use ________ : ________ : ________ : ________ : ________ : ________ : ________ : easy to use
extremely moderately slightly neutral slightly moderately extremely

4) useful ________ : ________ : ________ : ________ : ________ : ________ : ________ : useless
extremely moderately slightly neutral slightly moderately extremely

5) irrelevant ________ : ________ : ________ : ________ : ________ : ________ : ________ : relevant
extremely moderately slightly neutral slightly moderately extremely

6) I personally like to rate I personally dislike to rate
________ : ________ : ________ : ________ : ________ : ________ : ________ :
extremely moderately slightly neutral slightly moderately extremely

7) The ratings are liked by the Firefighter The ratings are disliked by the Firefighter
________ : ________ : ________ : ________ : ________ : ________ : ________ :
extremely moderately slightly neutral slightly moderately extremely

* *

8) Accurately filling out the efficiency mark rating scales will help improve the performance of Firefighters in the department.
likely ________ : ________ : ________ : ________ : ________ : ________ : ________ : unlikely
extremely moderately slightly neutral slightly moderately extremely

9) Helping improve the performance of Firefighters in the department is:
good ________ : ________ : ________ : ________ : ________ : ________ : ________ : bad
extremely moderately slightly neutral slightly moderately extremely

10) Accurately filling out the efficiency mark rating scales will help improve the performance of Firefighters who have difficulty doing the job.
likely ________ : ________ : ________ : ________ : ________ : ________ : ________ : unlikely
extremely moderately slightly neutral slightly moderately extremely

11) Improving the performance of Firefighters who have difficulty doing the job is
good ________ : ________ : ________ : ________ : ________ : ________ : ________ : bad
extremely moderately slightly neutral slightly moderately extremely

12) Accurately filling out the efficiency mark ratings scales will cause an increase of hostility between Lieutenants and Firefighters.
likely ________ : ________ : ________ : ________ : ________ : ________ : ________ : unlikely
extremely moderately slightly neutral slightly moderately extremely

13) Increasing hostility between Lieutenants and Firefighters is:
good _______: ________: ______: _____: ______: ________: _______: bad
extremely moderately slightly neutral slightly moderately extremely

14) Accurately filling out the efficiency mark rating scale will help improve the relationship between Firefighters in the department.
likely _______: ________: ______: _____: ______: ________: _______: unlikely
extremely moderately slightly neutral slightly moderately extremely

15) Improving the relationship between Firefighters in the department is:
good _______: ________: ______: _____: ______: ________: _______: bad
extremely moderately slightly neutral slightly moderately extremely

16) Accurately filling out the efficiency mark rating scale will help improve the relationship between City Hall "105" and the Firefighters and Lieutenants.
likely _______: ________: ______: _____: ______: ________: _______: unlikely
extremely moderately slightly neutral slightly moderately extremely

17) Improving the relationships between City Hall "105" and the Firefighters and Lieutenants is:
good _______: ________: ______: _____: ______: ________: _______: bad
extremely moderately slightly neutral slightly moderately extremely

18) Accurately filling out the efficiency mark rating scale will truthfully indicate the quality of work or ability of each Firefighter.
likely _______: ________: ______: _____: ______: ________: _______: unlikely
extremely moderately slightly neutral slightly moderately extremely

19) Truthfully indicating the quality of work or ability of each Firefighter is:
good _______: ________: ______: _____: ______: ________: _______: bad
extremely moderately slightly neutral slightly moderately extremely

20) Accurately filling out the efficiency mark rating scale will not matter because the scale and its scoring procedure are not "on the square" and not honest.
likely _______: ________: ______: _____: ______: ________: _______: unlikely
extremely moderately slightly neutral slightly moderately extremely

21) An efficiency mark scale and scoring procedure that are not "on the square" is:
good _______: ________: ______: _____: ______: ________: _______: bad
extremely moderately slightly neutral slightly moderately extremely

22) Accurately filling out the efficiency mark rating scale is important because the Fire Department will use the Lieutenants' ratings as the true job performance ability level of each Firefighter.
likely _______: ________: ______: _____: ______: ________: _______: unlikely
extremely moderately slightly neutral slightly moderately extremely

23) Having the Fire Department use the Lieutenants' ratings as the true job performance ability level of each Firefighter is:
good _______: ________: ______: _____: ______: ________: _______: bad
extremely moderately slightly neutral slightly moderately extremely

24) Accurately filling out the efficiency mark rating scale is important because it can accurately measure each Firefighter's ability to do his job.
likely _______: ________: ______: _____: ______: ________: _______: unlikely
extremely moderately slightly neutral slightly moderately extremely

25) Accurately measuring each Firefighter's ability to do his job is:
good _______ : _______ : _______ : _______ : _______ : _______ : _______ : bad
extremely moderately slightly neutral slightly moderately extremely

26) Accurately filling out the efficiency mark rating scale is useful only if the Lieutenant does not know the exact number that he is assigning to any one Firefighter.
likely _______ : _______ : _______ : _______ : _______ : _______ : _______ : unlikely
extremely moderately slightly neutral slightly moderately extremely

27) The Lieutenant not knowing the exact number that he is assigning to any one Firefighter on the efficiency mark is:
good _______ : _______ : _______ : _______ : _______ : _______ : _______ : bad
extremely moderately slightly neutral slightly moderately extremely

* *

28) Most people who are important to me and whose opinions I respect (e.g. coworkers, other department members) think that I should accurately complete the efficiency mark scale.
likely _______ : _______ : _______ : _______ : _______ : _______ : _______ : unlikely
extremely moderately slightly neutral slightly moderately extremely

29) The Firefighters on my own shift at my station house think that I should grade them accurately.
likely _______ : _______ : _______ : _______ : _______ : _______ : _______ : unlikely
extremely moderately slightly neutral slightly moderately extremely

30) In general, with regard to accurately completing the efficiency mark scales, I want to do what most of the Firefighters on my own shift at my own station think I should.
likely _______ : _______ : _______ : _______ : _______ : _______ : _______ : unlikely
extremely moderately slightly neutral slightly moderately extremely

31) The Firefighters on other shifts at my own station house think that I should accurately grade the efficiency mark scales of the Firefighters under my command.
likely _______ : _______ : _______ : _______ : _______ : _______ : _______ : unlikely
extremely moderately slightly neutral slightly moderately extremely

32) In general, with regard to accurately completing the efficiency mark scales, I want to do what most of the Firefighters on other shifts at my own station house think I should.
likely _______ : _______ : _______ : _______ : _______ : _______ : _______ : unlikely
extremely moderately slightly neutral slightly moderately extremely

33) The Firefighters in my own battalion think that I should accurately grade the efficiency mark scales of the Firefighters under my command.
likely _______ : _______ : _______ : _______ : _______ : _______ : _______ : unlikely
extremely moderately slightly neutral slightly moderately extremely

34) In general, with regard to accurately completing the efficiency mark scales, I want to do what most of the Firefighters in my battalion think that I should.
likely _______ : _______ : _______ : _______ : _______ : _______ : _______ : unlikely
extremely moderately slightly neutral slightly moderately extremely

35) My Captain or immediate supervisor thinks that I should accurately grade the efficiency mark scales of the Firefighters under my command.
likely ______ : ______ : ______ : ______ : ______ : ______ : ______ : unlikely
extremely moderately slightly neutral slightly moderately extremely

36) In general, with regard to accurately completing the efficiency mark scales, I want to do what my Captain or immediate supervisor thinks that I should.
likely ______ : ______ : ______ : ______ : ______ : ______ : ______ : unlikely
extremely moderately slightly neutral slightly moderately extremely

37) The other Lieutenants and Captains at my station house think that I should accurately grade the efficiency mark scales of the Firefighters under my command.
likely ______ : ______ : ______ : ______ : ______ : ______ : ______ : unlikely
extremely moderately slightly neutral slightly moderately extremely

38) In general, with regard to accurately completing the efficiency mark scales, I want to do what other Lieutenants and Captains at my station house think that I should .
likely ______ : ______ : ______ : ______ : ______ : ______ : ______ : unlikely
extremely moderately slightly neutral slightly moderately extremely

39) The Lieutenants and Captains in my battalion think that I should accurately grade the efficiency mark scales of the Firefighters under my command.
likely ______ : ______ : ______ : ______ : ______ : ______ : ______ : unlikely
extremely moderately slightly neutral slightly moderately extremely

40) In general, with regard to accurately completing the efficiency mark scales, I want to do what the Lieutenants and Captains in my battalion think that I should.
likely ______ : ______ : ______ : ______ : ______ : ______ : ______ : unlikely
extremely moderately slightly neutral slightly moderately extremely

41) My Chief and Marshal think that I should accurately grade the efficiency mark scales of the Firefighters under my command.
likely ______ : ______ : ______ : ______ : ______ : ______ : ______ : unlikely
extremely moderately slightly neutral slightly moderately extremely

42) In general, with regard to accurately completing the efficiency mark scales, I want to do what my Chief and Marshal think I should do.
likely ______ : ______ : ______ : ______ : ______ : ______ : ______ : unlikely
extremely moderately slightly neutral slightly moderately extremely

43) The Fire Department Commissioner's Office ("105") thinks that I should accurately grade the efficiency mark scales of the Firefighters under my command.
likely ______ : ______ : ______ : ______ : ______ : ______ : ______ : unlikely
extremely moderately slightly neutral slightly moderately extremely

44) In general, with regard to accurately completing the efficiency mark scales, I want to do what the Fire Department Commissioner's Office thinks I should do.
likely ______ : ______ : ______ : ______ : ______ : ______ : ______ : unlikely
extremely moderately slightly neutral slightly moderately extremely

45) The people at the Department of Personnel think that I should accurately grade the efficiency mark scales of the Firefighter under my command.

likely _______ : _______ : _______ : _______ : _______ : _______ : _______ : unlikely

extremely moderately slightly neutral slightly moderately extremely

46) In general, with regard to accurately grading the efficiency mark scales, I want to do what the people at the Department of Personnel think that I should.

likely _______ : _______ : _______ : _______ : _______ : _______ : _______ : unlikely

extremely moderately slightly neutral slightly moderately extremely

APPENDIX E

JOB ANALYSIS SESSION INSTRUMENTS

FIRE DEPARTMENT FIREFIGHTER
PERFORMANCE APPRAISAL DIMENSIONS

MEETING I

1. Knowledge of the job--knowledge of evolutions, construction types, procedures and of assignment area.

2. Willingness to work--perform work without being told, willingness to help out or to perform disagreeable tasks, performs work with a good attitude.

3. Willingness to take orders--ability to follow instructions.

4. Aggressiveness--ability to perform without a supervisor.

5. Physical fitness--physically able to perform the required tasks on the job.

6. Teamwork--ablilty to work in a group without being told.

7. Compatibility--ability to get along with other people.

8. Ability to perform under stressful conditions--remain calm and do one's own work at emergency situations.

9. Responsibility--willingness to perform any required task or fill any position as required.

10. Fairness--treating fellow workers equally, share equally in intercompany responsibilities and duties.

11. Consistency in performance--regularly performs well.

FIRE LIEUTENTANT PROMOTIONAL
RATING SCALE DIMENSIONS

MEETING 1

1. Leadership--a person whose judgments people respect plus a person who has the ability to make decisions.

2. Initiative--a person who looks for opportunities to learn and practice other jobs.

3. Appearance--a Firefighter who is neat and in proper uniform.

4. Public relations--ability to deal with the public.

5. Personnel Management--the ability to communicate, listen and motivate the men.

FIRE DEPARTMENT FIRE FIGHTER - FIRE LIEUTENANT
PERFORMANCE APPRAISAL DIMENSIONS

MEETING 2

1. Knowledge of the job--Knowledge of evolutions, construction types, procedures and of assignment area.

*2. Willingness to work--ability to perform work without a supervisor, willingness to help out or to perform disagreeable tasks, performs work with a good attitude.

*3. Willingness to take orders--willingness to follow instructions.

*4. Aggressiveness--DROP.

5. Physical fitness--physically able to perform the required tasks on the job.

6. Teamwork--ability to work in a group without being told.

7. Compatibility--ability to get along with other people.

8. Ability to perform under stressful conditions--remain calm and do one's own work at emergency situations.

9. Responsibility--willingness to perform any required task or fill any position as required.

*10. Fairness--treating fellow workers equally in responsibilities and duties.

11. Consistency in performance--regularly performs well.

*12. Leadership--a person whose judgments people respect plus a person who has the ability to give and take orders and make proper decisions.

13. Initiative--a person who looks for opportunities to learn and practice other jobs.

*14. Appearance--a fire fighter who is neat, in proper uniform, and maintains his personal gear.

15. Public relations--ability to deal with the public.

FIRE DEPARTMENT FIRE FIGHTER - FIRE LIEUTENANT
PERFORMANCE APPRAISAL DIMENSIONS

MEETING 2
(Cont.)

*16. Personnel management--the ability to communicate, listen to, motivate men and set an example.

*17. Attendance--being in service, knowing the conditions of the rig, and having consideration for partners.

* Changes of definitions from first meeting.

FIRE DEPARTMENT FIRE FIGHTER
PERFORMANCE APPRAISAL DIMENSIONS

MEETING 3

*1. Knowledge of the job--Knowledge of evolutions, equipment, construction types, procedures and of assignment area.

2. Willingness to work--ability to perform work, without a supervisor, willingness to help out or to perform disagreeable tasks, performs work with a good attitude.

*3. Willingness to take orders--DROP (covered under "Responsibility")

4. Aggressiveness--DROP

5. Physical fitness--physically able to perform the required tasks on the job.

*6. Teamwork--ability to work in a group.

*7. Compatibility--ability to get along with other people and treat fellow workers equally in responsibilities and duties.

8. Ability to perform under stressful conditions--remain calm and do one's own work at emergency situations.

*9. Responsibility--willingness to accept any required task or fill any position as required.

*10. Fairness--DROP (combined with "Compatibility")

11. Consistency in performance--regularly performs well.

*12. Leadership--a person who has the ability to give and take orders and make proper decisions.

13. Initiative--a person who looks for opportunities to learn and practice other jobs.

14. Appearance--a fire fighter who is neat, in proper uniform, and maintains his personal gear.

15. Public relations--ability to deal with the public.

FIRE DEPARTMENT FIRE FIGHTER
PERFORMANCE APPRAISAL DIMENSIONS

MEETING 3
(Cont.)

*16. Personnel mangement--DROP (covered under "Leadership")

*17. Promptness--being on time and in service.

*18. Attitude--avoids unnecessary layups and shows interest in his job.

* Changes of definitions from second meeting

FIRE DEPARTMENT FIRE FIGHTER
PERFORMANCE APPRAISAL DIMENSIONS

MEETING 4

*1. Knowledge of job--knowledge of evolutions equipment, construction types, and practices and procedures.

*2. Willingness to work--willingness to perform all duties with a good attitude.

5. Physical fitness--physically able to perform the required tasks on the job.

*6A. Teamwork in a fire--ability to work in a group at a fire.

*6B. Teamwork in quarters--ability to work in a group in quarters.

7. Compatibility--ability to get along with other people and treat fellow workers equally in responsibilities and duties.

8. Ability to perform under stressful conditions--remain calm and do one's work at emergency situations.

*9. Responsibility--accept any required task or fill any position as required.

11. Consistency in performance--regularly performs well.

*12. Leadership--ability to give and take orders and make proper decisions.

13. Initiative--a person who looks for opportunities to learn and practice other jobs.

14. Appearance--a fire fighter who is neat, in proper uniform and maintains his personal gear.

15. Public relations--ability to deal with the public.

17. DROP

18. DROP

* Changes from the third meeting.

LIST OF THE FINAL NINE
PERFORMANCE APPRAISAL DIMENSIONS AND THEIR ABBREVIATIONS

1. Knowledge of the Job (KOJ)--knowledge of evolutions, equipment, construction types, practices and procedures and area assignment.

2. Willingness to Work (WTW)--responsibility and consistency in performance.

3. Physical Fitness (PF)--physically able to perform the required tasks on the job.

4. Teamwork and Compatibility (TAC)--the ability to work in a group, to get along with other people and treat fellow workers equally in responsibilities and duties.

5. Ability to Perform Under Stressful Conditions (PSC)--remain calm and do one's own work at emergency situations.

6. Leadership (L)--a Firefighter who has the ability to give and take orders and make proper decisions.

7. Initiative (I)--a Firefighter who looks for opportunities to learn and practice other jobs.

8. Appearance (A)--a Firefighter who is neat, in proper uniform, and maintains his personal gear.

9. Public Relations (PR)--ability to deal with the public.

Appendix F

Questionnaire for Developing Behavioral Examples of the Twelve Dimensions

Instructions For Generating Behavioral Statements

The following sheets contain a list of qualities on dimensions that are important for proper firefighter job performance. Along with these dimensions are their definitions. These firefighter job dimensions and definitions were generated by about thirty firefighters and lieutenants. These firefighters and lieutenants were interviewed in one of four meetings held at Engine Company No. 42's station house between June 1st and June 8th. The dimensions and definitions were developed as the first step in completely rewriting the performance evaluation system for firefighters. <u>The old, five dimension system is being completely eliminated.</u>

Today, you will be participating in the second phase of creating a new performance evaluation system for firefighters. The Chicago Fire Department, in cooperation with a Department of Personnel research team, is using a new approach to performance evaluation. The Fire Department's new system will be more objective and less subject to biases from personality conflicts than in the past. Some of the sentences written today will be on the new job performance evaluation check list.

During today's session you will be asked to write brief, precise, specific examples of firefighter job performance. Each of your sentences should be an example of what <u>you believe</u> represents each of the dimensions listed. For each of the dimensions listed you will be asked to write three sentences. The first sentence for each dimension on quality of a firefighter should be an example of <u>superior or above average performance</u> of that dimension. The second sentence for each dimension should be an example of <u>average or standard performance</u> of that dimension. The third sentence will be a specific instance of <u>poor or substandard</u> performance of that dimension. Please make your statements as specific as possible.

The following sentences are an example of what you are expected to write. Suppose that one of the qualities of proper firefighting behavior is the following (Note that this is not one of the dimensions, but merely an example):

<u>Dimensions</u>

1. <u>EXAMPLE</u> = Care of Equipment: Maintenance and care of all equipment used.

Tell exactly what a firefighter did that is a <u>superior</u> example of Care of Equipment. <u>This firefighter checked all of his equipment each work day and took care of all necessary maintenance.</u>

DIMENSIONS AND DEFINITIONS OF FIREFIGHTER

Listed below are some dimensions and definitions of proper firefighter performance. For each dimension, think back over your years of service in the department and recall examples of each dimension. Then, write specific behavioral examples of each dimension using actions you have seen or heard. For each dimension:

First, try to think of a specific instance (critical incident) when a firefighter has performed his job in a superior manner (an excellent level of performance).

Second, write an example of an action that is an average or standard level of performance of the dimension.

Third, write a behaviorial example of a poor or substandard level of performance of the dimension.

Please, DO NOT WRITE ANY NAMES in your sentences describing actions for each dimension. Also, do not copy someone else's examples. The research staff needs many different examples of each dimension.

Dimensions

1. Knowledge of the Job: Knowledge of evolutions, equipment, construction types, practices and procedures, and area assignment.

Tell exactly what a firefighter did that is a superior example of knowledge of the job. ________________________________

____

____

____

Tell exactly what a firefighter did that is an average or standard example of knowledge of the job. ________________

____

____

____

Tell exactly what a firefighter did that is an average or standard example of Care of Equipment. This firefighter would check and take care of his equipment when ordered.

Tell exactly what a firefighter did that is a poor or sub-standard example of Care of Equipment. This firefighter lost tools at a fire scene.

Please keep your behavioral examples as specific as possible. Do you notice anything wrong with the above three sentences. Do you notice anything correct in the above examples? DO NOT RECORD ANY NAMES of firefighters in your sentences. The incidents you describe, although they should have actually occurred, will not be used to either help or hurt any firefighters. Your sentences will be used only in the development of the performance evaluation system.

Remember, the sentences you write will be considered as potential sentences in the new performance evaluation check list. For this reason, try to write examples of each dimension that you yourself, as a firefighter (or lieutenant) would want to be rated on (or rate someone on). The behavioral statements you write as examples of each dimension should be behaviors that you think are important for firefighters to either perform or avoid performing.

Thank you for your cooperation.

Tell exactly what a firefighter did that is a poor or substandard example of knowledge of the job

2. Willingness to Work: performs work without a supervisor, willingness to help out or to perform disagreeable tasks, performs work with a good attitude.

Superior Performance:

Standard Performance:

Substandard Performance:

3. Physical Fitness: physically able to perform the required tasks on the job.

Superior Performance:

Standard Performance: ______________________________

____

____

____

Substandard Performance: ______________________________

____

____

____

4. <u>Teamwork</u>: ability to work in a group.

Superior Performance: ______________________________

____

____

____

Standard Performance: ______________________________

____

____

____

Substandard Performance: ______________________________

____

____

____

5. <u>Compatibility:</u> ability to get along with other people and treat fellow workers equally in responsibilities and duties.

Superior Performance: ______________________________

____

____

____

Standard Performance: ______________________________

____

____

____

Substandard Performance: ___________________________

____

____

____

6. Ability to perform under stressful conditions -

-remain calm and do one's work at emergency situations.

Superior Performance: _____________________________

____

____

____

Standard Performance: ______________________________

____

____

____

Substandard Performance: ___________________________

____

____

____

7. Responsibility - accept any required task or fill any position as required.

Superior Performance: _____________________________

____

____

____

Standard Performance: ____________________

____

____

____

Substandard Performance: ____________________

____

____

____

8. <u>Consistency in performance</u> -

regularly performs well.

Superior Performance: ____________________

____

____

____

Standard Performance: ____________________

____

____

____

Substandard Performance: ____________________

____

____

____

9. <u>Leadership</u> - ability to give and take orders and make proper decisions.

Superior Performance: ______________________________

____

____

____

Standard Performance: ______________________________

____

____

____

Substandard Performance: ___________________________

____

____

____

10. <u>Initiative</u> - a person who looks for opportunities to learn and practice other jobs.

Superior Performance: ______________________________

____

____

____

Standard Performance: ______________________________

____

____

____

Substandard Performance: ___________________________

____

____

____

11. Appearance - a fire fighter who is neat, in proper uniform and maintains his personal gear.

Superior Performance: ______________________________

Standard Performance: ______________________________

Substandard Performance: ______________________________

12. Public relations - ability to deal with the public.

Superior Performance: ______________________________

Standard Performance: ______________________________

Substandard Performance: ______________________________

Appendix G

Retranslation and Scaling Questionnaire

Sample Items

NEW FIREFIGHTER PERFORMANCE APPRAISAL RATING SCALE
RESEARCH PROJECT:

RETRANSLATION AND SCALING QUESTIONNAIRE

FORM A PART I

Instructions for Retranslation and Scaling

Today, you will be working on the third phase of developing a new performance evaluation system for firefighters. Phases One and Two were also completed by firefighters and lieutenants from the Chicago Fire Department. In Phase One, firefighters and lieutenants developed the twelve qualities that they believed are the essential skills required for proper firefighter job performance. A portion of your task today will be to become very familiar with these qualities of proper firefighter job performance and their definitions.

The second phase in the process of developing a firefighter job performance evaluation checklist involved having firefighters and lieutenants write sentences. These sentences were to be based on the definitions of proper firefighter job performance developed in Phase One. Another portion of your task today will be to read some of the sentences based on these twelve different qualities written by firefighters and lieutenants. The qualities and definitions are written on the left side of each page of this booklet, the sentences, on the right.

After reading the twelve qualities of proper firefighter performance, you will be asked to do two things for each of the definition-example sentences that you read:

1. Write down on the first blank line the definition number corresponding to the definition that you think the sentence was written about. Pick the definition related to each sentence. There are many examples of each definition.

2. Write down on the second blank line a number between one (1) and seven (7) corresponding to the quality of performance described in that sentence. Write the number:

   (1) If the sentence gives an example of <u>extremely poor performance</u> by a firefighter (the worst possible performance).
   (2) The sentence is an example of <u>very poor performance</u> by a firefighter.
   (3) The sentence is an example of <u>poor performance</u> by a firefighter.
   (4) The sentence is an example of <u>average performance</u> by a firefighter.
   (5) The sentence is an example of <u>good performance</u> by a firefighter.
   (6) The sentence is an example of <u>very good performance</u> by a firefighter.
   (7) The sentence is an example of <u>excellent performance</u> by a firefighter.

Instructions for Retranslation and Scaling

There are no right or wrong answers to these questions, so please don't copy any one else's work. The responses given by all the firefighters and lieutenants in Phase Three will be averaged and the best of the sentences will be chosen for the final performance evaluation checklist. A sentence will be considered "good" only if most firefighters and lieutenants can agree that the sentence is an example of only one, and not several, definitions.

Finally, we realize that this is a very difficult task. Reading the sentences and picking out the definition of which it is an example is a lot like taking a multiple-choice test with twelve (12) answers to choose from. However, please take your time and be careful to think about your responses. Try to answer each item and don't get discouraged. The final performance evaluation checklist can only be as good as the quality of work you do here today.

DEFINITIONS	Definition Number	Performance Level
1. Knowledge of the Job: Knowledge of evolutions, equipment, construction types, practices and procedures, and area assignment.	___	___
2. Willingness to Work: Performs work without a supervisor, willingness to help out or to perform disagreeable tasks, performs work with a good attitude.	___	___
	___	___
3. Physical Fitness: Physically able to perform the required tasks on the job.	___	___
4. Teamwork: Ability to work in a group.		
5. Compatibility: Ability to get along with other people and treat fellow workers equally in responsibilities and duties.	___	___
6. Ability to Perform Under Stressful Conditions: Remain calm and do one's own work at emergency situations.	___	___
7. Responsibility: Willingness to accept any required task or fill any position as required.	___	___
8. Consistency in Performance: Regularly performs well.	___	___
9. Leadership: A firefighter who has the ability to give and take orders and make proper decisions.	___	___
10. Initiative: A firefighter who looks for opportunities to learn and practice other jobs.	___	___
11. Appearance: A firefighter who is neat, in proper uniform, and maintains his personal gear.	___	___
12. Public Relations: Ability to deal with the public.	___	___
	___	___
	___	___

Performance Level Score	1 Extremely Poor	2 Very Poor	3 Poor	4 Average	5 Good	6 Very Good	7 Excellent

This firefighter performs well with coworkers and can be counted on to help in difficult situations.

This firefighter keeps himself in above average physical condition because of the physical strain connected with the job of firefighting.

This firefighter could be expected to almost come in clean every work day.

This firefighter could be expected to do work so it never has to be gone over.

This firefighter could be expected to foul up constantly.

This firefighter could be expected never to be around, and to have to be looked for when needed.

This firefighter is able to take up the slack of somebody who isn't doing his share of the work at fires or at the fire house.

This firefighter could be expected to be unable to get along with others; he feels better than others.

This firefighter could be expected to be unable to get his point across.

This firefigher could be expected to argue and fight with civilians.

This firefighter could be expected to refuse totally to do anything but his job.

This firefighter could be expected to have a little doubt about his own decisions.

This firefighter could be expected not to do anything.

This firefighter could be expected never to offer to lend anyone a hand.

Performance Level Score	1 Extremely Poor	2 Very Poor	3 Poor	4 Average	5 Good	6 Very Good	7 Excellent

This firefighter could be expected to be in very good condition, to take care of himself, and to perform any task well.

This firefighter not only knows his material thoroughly, but takes the initative to show the other men.

This firefighter could be expected to be rude and a smart alec; he shows no respect for others.

This firefighter can be put in any position or task and do it right.

This firefighter is a clean person.

This firefighter takes extra tactics courses on firefighting at college.

This firefighter could be expected to receive and obey orders directed to his team.

This firefighter does not have good safety habits and someone always has to look for him.

This firefighter never shaves before coming to firehouse and doesn't replace torn shirts or pants.

This firefighter uses good safety methods and keeps in close contact with his partner at all times.

This firefighter helps the public.

This firefighter could be expected to let subordinates make their own decisions.

This firefighter could be expected to go to check tools on the rig.

This firefighter could be expected to take a personal interest in people and their problems at fires.

Appendix H

Behaviorally Anchored Rating Scale for Firefighters

BEHAVIORALLY ANCHORED RATING SCALE

DIRECTIONS

Listed on the following pages are a set of examples of Firefighter job performance. These behavioral examples are organized under nine important definitions of proper Firefighter job performance. These nine definitions, as well as all of the behavioral statements, were written by Firefighters and Lieutenants in the Chicago Fire Department. Your task as a rater is to evaluate the job performance of a Firefighter along each of the following scales.

First, read the definition provided describing one aspect of proper Firefighter job performance. These definitions are written at the top of each page.

Second, place an "X" in any numbered box along the vertical rating scale line that you believe best represents the Firefighter's level of performance of that specific definition. ("1" represents the lowest level of performance and "7" represents the highest possible level of performance.)

This rating scale is a new format in the Chicago Fire Department under consideration for future Department wide use. The rating scale itself can be thought of as similar in style to a thermometer. The top of the scale means more of the quality (either heat in a thermometer or a performance quality in this rating scale). The top of the scale is good performance; the bottom, bad performance. The position of the "X" along the scale will decide how much of that quality the Firefighter possesses.

The behavioral statements written along side the scale are examples of the types of behavior that could be expected from a Firefighter performing at that level of performance. The sentences are presented in order to give you, the rater, an idea of what the numbers on the scales mean. Place an "X" on the scale at the point where you expect the Firefighter to perform. Remember to mark each scale by writing in only one of the seven numbered boxes. Do not mark outside of the boxes.

Rank of Rater ________________________

Name of Firefighter being rated (please print clearly)

____

Knowledge of the job - Knowledge of evolutions, equipment, construction types, practices and procedures, and area assignment.

7

This Firefighter, upon arriving at a fire scene, immediately has thorough knowledge of what is going on, vantage points, size up, etc.

This Firefighter knows his still districts; streets and avenues, building of a dangerous type, handicapped persons, one-way streets and dead end streets.

6

This Firefighter could be expected to know to a degree building construction and fire science.

5

4

This Firefighter could be expected to go on a roof, open a hole not directly over the fire or not make the hole large enough.

3

This Firefighter could be expected not to know how to wash down a room properly.

2

This Firefighter doesn't know how to tie knots, where to find tools on the apparatus, and is unable to perform the basic evolutions of the job.

This Firefighter has no idea of the streets and avenues in his still district and has no interest in remembering them.

1

Willingness to work - responsibility and consistency in performance.

7

6 — This Firefighter is first in line to volunteer, never complains and completes a job without supervision.

This Firefighter performs all required tasks in an outstanding display of workmanship as required of him as a Firefighter.

This Firefighter could be expected to do his work without being told what to do and then help others.

5

This Firefighter could be expected to perform his work in good fashion upon orders.

4

3

This Firefighter could be expected to have to be told to start assigned duties.

This Firefighter could be expected to do only required work and to sometimes "take a duck" when dirty work is required.

2

1

Physical fitness — physically able to perform the required tasks on the job.

7

6

This Firefighter could be expected to be able to chop, carry, and climb until the job is done.

This Firefighter could be expected to exercise and keep physically fit and to perform his job on his own.

This Firefighter could be expected to work out on a regular basis while at the fire station to be prepared for strenuous work (might do streching exercises etc.)

5

4

This Firefighter could be expected to be able to function on his own with little weight problem.

This Firefighter could be expected to perform his job well, but might get tired after a hard day and is not as alert as he could be.

3

This Firefighter doesn't seem to care about his physical condition.

2

1

Teamwork and Compatibility - the ability to work in a group, to get along with other people and treat fellow workers equally in responsibilities and duties.

7

6

This Firefighter could be expected to go all out to help, regardless of personalities.

This Firefighter works well as a leader or follower in team tasks with any team formed as well as his assigned team.

This Firefighter could be expected to have the ability to get along very well with his co-workers.

This Firefighter could be expected to be able to communicate with others and work with others.

5

This Firefighter could be expected to reflect spirit and effort in maintaining harmony between himself and fellow workers.

4

This Firefighter could be expected to be part of the crowd.

3

This Firefighter could be expected to be a loner, seldom seen talking with co-workers.

This Firefighter could be expected always to criticize others when they don't know what they're doing.

This Firefighter could be expected to be a person that would not remember that firefighting needs teamwork.

2

1

Ability to perform under stressful conditions - remain calm and do one's own work at emergency situations.

7

-

-

-

6

This Firefighter could be expected to remain calm and able to perform duties well under extreme stress.

This Firefighter seems to be in complete control at emergencies.

This Firefighter could be expected to keep a cool head and concentrate on the job in an emergency situation.

5

This Firefighter could be expected to get a little nervous when things get rough, but to get the job done.

-

-

-

4

-

-

-

3

This Firefighter could be expected to get nervous at fires and to be very unsure of himself.

This Firefighter cannot remain calm and sometimes does the wrong thing.

-

2

-

-

1

Leadership - a Firefighter who has the ability to give and take orders and make proper decisions.

7

This Firefighter could be expected to take charge and lead the company in any task.

6

This Firefighter has the knowledge and ability to take charge.

This Firefighter could be expected to give the right orders and make proper decisions when he is in charge.

This Firefighter is able to lead the company in a job.

5

4

This Firefighter does what he is told and has no leadership traits.

This Firefighter could be expected to make basic decisions, but won't fully try to lead.

3

This Firefighter could be expected to be unsure of himself and to look for others to make decisions.

This Firefighter could be expected not to want to be committed to a decision.

2

1

<u>Initiative</u> - a Firefighter who looks for opportunities to learn and practice other jobs.

7

6

This Firefighter could be expected always to be looking for different ways to improve.

This Firefighter could be expected to study and to practice what he is taught.

This Firefighter could be expected to take extra tactics courses on firefighting at college.

This Firefighter is always looking for something to do to make his job better and easier.

5

This Firefighter could be expected to ask questions about anything he doesn't know.

This Firefighter could be expected to ask questions when he needs to know something.

4

3

This Firefighter could not be expected to look for opportunities to learn his job.

This Firefighter could be expected to shy away from opportunities to learn which involve commitment on his part.

2

1

Appearance - a Firefighter who is neat, in proper uniform, and maintains his personal gear.

7

6

This Firefighter uses his clothing allowance check to replace worn or torn clothing.

This Firefighter could be expected to be well dressed, with proper fire gear and clothes.

5

This Firefighter could be expected to show uniformity in dress or work uniforms.

This Firefighter could be expected to have the prescribed clothing and fire clothing.

4

3

This Firefighter could not be expected to wear the prescribed uniform, to look unprofessional.

This Firefighter has a problem with personal hygiene or he never cleans his gear.

This Firefighter could be expected to have clothing which is not clean and fire clothing which is not organized to respond to a call.

2

1

Public relations - ability to deal with the public.

[7]

[6]

This firefighter likes to talk with people; he's the one that jumps up when a class of kids comes to the firehouse to learn how the Fire Department works.

This Firefighter could be expected to have the knowledge to talk to people in a hostile climate in order to calm them down.

[5]

This Firefighter helps the public with directions, automobile problems, and first aid.

This Firefighter could be expected to be able to deal with the public at all times.

This Firefighter could be expected to speak when spoken to, but not go out of his way to associate with the public.

[4]

[3]

This Firefighter could be expected not to like to deal with the public and to say it's not his job.

[2]

This Firefighter has constant arguments with neighbors and co-workers and does not help the public with directions or automobile problems.

[1]

Appendix I

Mixed Standard Rating Scale for Firefighters

MIXED STANDARD RATING SCALE

DIRECTIONS

Listed on the following pages are a set of twenty-seven (27) descriptions of Firefighter job performance. These sentences were written from examples provided by members of the Chicago Fire Department. Your task as a rater is to evaluate the job performance of a Firefighter by completing the following scale. First, read each of the examples of job performance. Second, determine whether the Firefighter you are rating performs his duties "better than," "the same as," or "worse than" the example.

If you believe that the Firefighter you are rating is "better than" the statement, write a "+" (plus) on the blank line to the left of the statement.

If you believe that the Firefighter you are rating is "the same as" or "exactly fits" the description, write a "0" (zero) in the space to the left of the statement.

If you believe that the Firefighter you are rating is "worse than" the statement, write a "—" (minus) on the blank line to the left of the statement.

Please do not skip any items. Make sure to write a "+", or a "0", or a "—" next to each of the statements. Please take your time and be as accurate as possible. Your check marks will decide this Firefighter's "performance" or "efficiency" marks.

Rank of Rater ______________________

Name of Firefighter being rated (please print clearly)

____

+ indicates that the Firefighter could be expected to perform better than the statement

0 indicates that the Firefighter could be expected to perform exactly the same as the statement

- indicates that the Firefighter could be expected to perform worse than the statement

<u>Rating</u>	<u>Statement</u>
1.________	1. This Firefighter could be expected to open a roof properly for fire control and ventilation or wash down charred ceiling joists properly.
2.________	2. This Firefighter could be expected to be a loner, seldom seen talking with co-workers.
3.________	3. This Firefighter is always on time and ready for work; he performs the tasks of a Firefighter with an outstanding display of workmanship and is looking to help others with their work.
4.________	4. This Firefighter could be expected to function as a Firefighter with little weight problem.
5.________	5. This Firefighter is able to lead the company in some tasks and make some basic decisions.
6.________	6. This Firefighter might be expected to have to be told to start assigned duties, do only required work or sometimes "take a duck" when dirty work is required.
7.________	7. This Firefighter might not be expected to remain calm in a stressful situation or might be unsure of himself and sometimes do the wrong thing when he is without close supervision.
8.________	8. This Firefighter could be expected to have difficulty working with people and sometimes might forget that firefighting needs teamwork, not criticism between team members.
9.________	9. This Firefighter could be expected to pride himself on his physical fitness to perform strenuous work such as chopping, carrying and climbing.
10.________	10. This Firefighter could be expected not to like to deal with the general public, saying "it's not his job."
11.________	11. This Firefighter could be expected to be looking for ways to improve himself by enrolling in formal classes, asking questions about things he doesn't know or looking for something to do to make conditions better and easier for his company.

+ indicates that the Firefighter could be expected to perform better than the statement

0 indicates that the Firefighter could be expected to perform exactly the same as the statement

- indicates that the Firefighter could be expected to perform worse than the statement.

Rating	Statement
12.______	12. This Firefighter could be expected to get a little nervous when things get rough but to get the job done in some fashion.
13.______	13. This Firefighter could be expected to perform his work in good fashion upon orders, whether written or verbal.
14.______	14. This Firefighter has the knowledge and ability to take charge of a company, can be expected to give the right orders, and make proper decisions.
15.______	15. This Firefighter could be expected to have a problem with personal cleanliness, might not wear the prescribed uniform or might not have fire clothing organized for response to alarms.
16.______	16. This Firefighter could be expected to speak to the public when spoken to, but not go out of his way to associate with the public.
17.______	17. This Firefighter might not be able to know how to tie knots, where to find tools on the apparatus, is unable to perform basic evolutions of the job properly or has no interest in learning the still district.
18.______	18. This Firefighter performs his job at a minimal level, tiring easily because of his physical condition.
19.______	19. This Firefighter works well as a leader or follower in team tasks, with any team formed as well as his assigned team and could be expected to reflect spirit and effort in maintaining harmony between himself and fellow workers.
20.______	20. This Firefighter could be expected to look for opportunities to learn his job by studying and practicing what he has been taught.

+ indicates that the Firefighter could be expected to perform better than the statement

0 indicates that the Firefighter could be expected to perform exactly the same as the statement

- indicates that the Firefighter could be expected to perform worse than the statement

Rating	Statement
21. ________	21. This Firefighter could be expected to be well dressed at all times and have proper fire clothes by using his clothing allowance check to replace worn clothing.
22. ________	22. This Firefighter could be expected to have superior knowledge of his still district, including all streets, building constructions and dangerous buildings so that upon arriving at a fire scene, he immediately has knowledge of what is going on (e.g. vantage points, size-up, lead-outs, etc.).
23. ________	23. This Firefighter could be expected to remain calm and able to perform duties well under extreme stress.
24. ________	24. This Firefighter could be expected to shy away from opportunities to learn his job, especially when special committment would be involved (e.g. taking fire science courses, attending demonstrations at the fire academy, etc.)
25. ________	25. This Firefighter is unsure of himself and looks for others to make decisions.
26. ________	26. This Firefighter could be expected to show the prescribed uniformity in his dress or work clothes.
27. ________	27. This Firefighter likes to talk with people, he might take care of a class of children visiting the firehouse, help the public with their problems (e.g. give directions, help with auto difficulties, give first-aid) or could be expected to talk with people in a hostile climate and calm them down.

Appendix J

Summary of the Promotional Rank Order Criterion Regressions

BARS and MSS Total Score Regressions
On Promotional Rank Order Criterion

I. All Raters $N = 179$

Scale	R^2	r	B	BETA	F	P
BARS	.14092	-.375	-24.22	-.3754	13.45	.001
MSS	.21832	-.467	-26.14	-.4670	22.622	.001

II. Lieutenant Raters $N = 61$

Scale	R^2	r	B	BETA	F	P
MSS	.22555	-.47492	-27.254	-.4749	17.183	.001
BARS	.22099	-.47010	-32.839	-.4701	16.737	.001

III. Captain Raters $N = 22$

Scale	R^2	r	B	BETA	F	P
BARS	.01620	-.12727	- 6.745	-.12727	.329	NS
MSS	.19365	-.44006	-22.464	-. 4400	4.803	NS

Stepwise Regressions Showing Incremental Variance Accounted for by the BARS and MSS Subscales on the Promotional Rank Order Criterion

I. BARS $N = 83$ $F\ (9,73) = 2.88,$ $P < .001$

Subscales	Incremental R^2	R^2 Change	B	BETA	F	P
I*	.16552	.16552	-137.886	-.3047	16.378	.01
PSC	.20085	.03532	- 71.272	-.1829	3.495	NS
PR	.22511	.02426	82.141	.1949	2.400	NS
TAC	.23522	.01011	76.342	.1849	1.000	NS
L	.24361	.00839	- 54.180	-.1539	.823	NS
PF	.25692	.01331	- 45.029	.1118	1.317	NS
KOJ	.25916	.00224	- 38.149	-.0915	.221	NS
A	.26192	.00276	- 23.117	-.0592	.273	NS
WTW	.26223	.00031	- 11.145	-.0274	.031	NS
(constant)			1866.28			

II. MSS $N = 83$ $F\ (9,73) = 3.4259,$ $P < .01$

Subscale	Incremental R^2	R^2 Change	B	BETA	F	P
A	.22627	.22627	-100.47	-.3598	23.49	.01
KOJ	.27124	.04497	- 61.89	-.1912	4.67	.1
I	.28581	.01458	- 47.52	-.1349	1.51	NS
TAC	.28943	.00362	- 25.99	.0694	.38	NS
L	.29274	.00332	- 34.14	-.0995	.34	NS
WTW	.29547	.00272	25.39	.0695	.28	NS
PF	.29614	.00067	- 17.67	-.0459	.07	NS
PR	.29677	.00063	10.43	.0301	.06	NS
PSC	.29695	.00019	7.05	.0224	.02	NS
(constant)			1565.44			

* Subscale abbreviations are translated in Appendix E

Hierarchical Regressions Showing Unique Variance Accounted for by the Nine BARS and MSS Subscales on the Promotional Rank Order

I. BARS $N = 83$ $R = .51208$ $R^2 = .26223$

Subscales	Change R^2	r	B	BETA	F	P
I*	.04572	-.40685	-137.89	-.305	4.524	.05
PSC	.01128	-.36346	- 71.272	-.183	1.116	NS
PR	.01979	-.1199	82.141	.195	1.958	NS
TAC	.01524	-.1430	76.34	.185	1.508	NS
L	.00739	-.29133	- 54.180	-.154	.731	NS
PF	.00870	-.25227	- 45.029	-.112	.861	NS
KOF	.00240	-.3677	- 38.149	-.092	.237	NS
A	.00225	-.118	- 23.117	-.059	.223	NS
WTW	.00031	-.300	- 11.155	-.027	.030	NS

II. MSS $N = 83$ $R = .54493$ $R^2 = .29695$

Subscale	Change R^2	r	B	BETA	F	P.
A	.08172	-.476	-100.469	-.3598	8.485	.01
KOJ	.01331	-.40190	- 61.896	-0.191	1.382	NS
I	.01504	-.3006	- 47.520	-.135	1.562	NS
TAC	.00391	-.15011	25.991	.069	.406	NS
L	.00412	-.3397	- 34.144	-.099	.428	NS
WTW	.00262	-.2309	25.397	.069	.272	NS
PF	.00101	-.321	- 17.671	-.046	.105	NS
PR	.00064	-.208	10.435	.03009	.067	NS
PSC	.00019	-.33257	7.051	.02236	.019	NS

* Subscale Abbreviations are translated in Appendix E

Stepwise Regressions for Lieutenant Raters Showing the Incremental Variance Accounted for by the BARS and MSS Subscales for the Promotional Rank Order Criterion

I. BARS $N = 61$ $F(9,51) = 3.22$, $P < .01$

Subscales	R^2	Incremental R^2 Change	B	BETA	F	P
I*	.20856	.20856	-92.274	-.203	16.675	.01
PF	.26982	.06126	-97.542	-.244	4.898	.1
L	.30036	.03054	-82.741	-.225	2.442	NS
PR	.32533	.02497	74.540	.169	1.996	NS
WTW	.33636	.01103	-51.918	-.112	.882	NS
TAC	.35032	.01397	109.027	.219	1.117	NS
PSC	.35590	.00558	-68.032	-.160	.446	NS
A	.36047	.00457	-37.629	-.090	.335	NS
KOJ	.36212	.00164	-31.700	-.069	.131	NS
(constant)			2137.07			

II. MSS $N = 61$ $F(9,51) = 3.06$, $P < .01$

Subscale	R^2	Incremental R^2 Change	B	BETA	F	P
A	.27717	.27717	-125.89	-.462	21.77	.01
KOJ	.31984	.04267	- 92.18	-.251	3.35	NS
TAC	.33300	.01316	37.12	.099	1.00	NS
I	.33929	.00629	- 42.54	.123	.49	NS
PR	.34808	.00879	35.53	.098	.69	NS
PF	.34993	.00185	24.07	.059	.14	NS
L	.35026	.00033	- 10.99	-.029	.03	NS
PSC	.35050	.00024	11.23	.031	.02	NS
WTW	.35070	.00020	7.65	-.021	.02	NS
(constant)			1416.37			

---

* Subscale abbreviations are translated in Appendix E

Hierarchical Regressions for Lieutenants Showing Unique Variance Accounted for by the BARS and MSS Subscales on the Promotional Rank Order Criterion

I. BARS $N = 61$ $R = .58905$ $R^2 = .34698$

Subscales	R^2 Change	r	B	BETA	F	P
I*	.01747	-.457	-92.273	-.203	1.397	NS
PF	.03351	-.422	-97.542	-.243	2.679	NS
L	.02030	-.318	-82.741	-.225	1.623	NS
PR	.01514	-.165	75.540	.169	1.210	NS
WTW	.00652	-.387	-51.918	.112	.521	NS
TAC	.02240	-.144	109.027	.219	1.791	NS
PSC	.00853	-.406	-68.032	-.160	.682	NS
A	.00495	-.203	-37.629	-.090	.396	NS
KOJ	.00164	-.401	-31.700	-.069	.131	NS

II. MSS $N = 61$ $R = .59220$ $R^2 = .35070$

Subscale	R^2 Change	r	B	BETA	F	P
A	.12037	-.526	-125.89	-.4617	9.455	.01
KOJ	.02382	-.440	- 92.18	-.251	1.871	NS
TAC	.00713	-.162	37.122	.099	.560	NS
I	.00988	-.366	- 42.535	-.123	.776	NS
PR	.00654	-.191	35.53	.097	.514	NS
PF	.00153	-.282	24.069	.059	.121	NS
L	.00042	-.323	- 10.99	-.029	.033	NS
PSC	.00036	-.332	11.23	.031	.028	NS
WTW	.00020	-.218	- 7.65	-.021	.015	NS

---

* Subscale abbreviations are translated in Appendix E

Stepwise Regressions for Captain Raters Showing the Incremental Variance Accounted for by the BARS and MSS Subscales for the Promotional Rank Order Criterion

I. BARS $N = 22$ $F(8,13) = .6583$, NS

Subscales	R^2	Incremental R^2 Change	B	BETA	F	P
PSC*	.08662	.08662	70.042	.225	1.553	NS
PF	.15058	.06395	180.98	.450	1.146	NS
WTW	.20365	.05307	194.06	.634	.951	NS
KOJ	.24430	.04065	-400.23	-1.213	.730	NS
I	.27211	.02718	-132.83	- .288	.498	NS
TAC	.30009	.02798	132.01	.423	.501	NS
A	.32443	.02434	- 84.98	- .262	.436	NS
PR	.33054	.00611	39.30	.110	.109	NS
(constant)			721.28			

II. MSS $N = 22$ $F(9,12) = .6076$, NS

Subscale	R^2	Incremental R^2 Change	B	BETA	F	P
PF	.17725	.17725	1.93	.006	3.096	NS
A	.19711	.01986	-96.46	-.324	.347	NS
L	.21443	.01732	-157.33	-.572	.302	NS
WTW	.24194	.02751	197.49	.539	.480	NS
PR	.28726	.04532	-93.63	-.313	.792	NS
I	.29707	.00980	-66.84	-.179	.171	NS
TAC	.30460	.00753	31.77	.087	.131	NS
PSC	.30727	.00268	34.18	.143	.047	NS
KOJ	.31305	.00578	-50.76	-.209	.101	NS
(constant)			1643.9			

---

* Subscale abbreviations are translated in Appendix E

Hierarchical Regressions for Captain Raters Showing Unique Variance Accounted for by the BARS and MSS Subscales on the Promotional Rank Order Criterion

I. BARS $N = 22$ $R = .57493$ $R^2 = .33054$

Subscales	R^2 Change	r	B	BETA	F	P
PSC*	.00547	-.294	70.092	.2256	.098	NS
PF	.13298	.276	180.944	.4501	2.383	NS
WTW	.07060	-.108	194.160	.634	1.266	NS
KOJ	.07127	-.285	-399.975	-1.212	1.278	NS
I	.04388	-.1985	-132.876	- .288	.787	NS
TAC	.03876	-.077	131.98	.423	.695	NS
A	.02926	.12560	- 84.961	- .262	.524	NS
PR	.00418	.02428	39.424	.110	.075	NS
L	.00000	-.21554	- .4085	- .001	.000	NS

II. MSS $N = 22$ $R = .55951$ $R^2 = .31305$

Subscale	R^2 Change	r	B	BETA	F	P
PF	.0001	-.4210	1.930	.006	.000	NS
A	.04363	-.28356	-96.458	-.324	.762	NS
L	.04103	-.403	-157.33	-.572	.717	NS
WTW	.07394	-.227	197.49	.539	1.292	NS
PR	.05060	-.239	- 93.63	-.3126	.884	NS
I	.01766	-.133	- 66.838	-.1787	.309	NS
TAC	.00642	-.109	31.768	.086	.112	NS
PSC	.00620	-.330	34.179	.143	.108	NS
KOJ	.00578	-.353	- 50.763	-.209	.101	NS

* Subscale abbreviations are translated in Appendix E

Appendix K

Summary of the Seniority

Criterion Regressions

BARS and MSS Total Score Regressions

on the Seniority Criterion

I. All Raters $N = 179$

Scale	R^2	r	B	BETA	F	P
BARS	.00846	-.093	-49.04	-.093	1.536	NS
MSS	.00767	-.087	-43.73	-.087	1.375	NS

II. Lieutenant Raters $N = 121$

Scale	R^2	r	B	BETA	F	P
BARS	.00916	-.09570	-58.82	-.0957	1.099	NS
MSS	.00964	-.09816	-53.79	-.09816	1.158	NS

III. Captain Raters $N = 57$

Scale	R^2	r	B	BETA	F	P
BARS	.00791	-.0889	-29.855	-.0889	.446	NS
MSS	.00032	-.0178	- 5.924	-.0178	.017	NS

Stepwise Regressions Showing Incremental Variance

Accounted for by the BARS and MSS Subscales

on the Seniority Criterion

I. BARS $\underline{N} = 166$ $\underline{F}\ (9,156) = 6.410,\ \underline{p} < .01$

Subscale	R^2	Incremental R^2 Change	B	BETA	F	P
L *	.08875	.08875	-129.32	-.089	18.965	.01
PF	.19313	.10438	623.67	.389	22.30	.01
KOJ	.22539	.03226	-393.24	.257	6.894	.01
I	.24353	.01814	386.78	.211	3.876	NS
WTW	.26184	.01831	-267.10	-.158	3.913	.05
PSC	.26587	.00403	-202.21	-.123	.861	NS
A	.26796	.00210	-116.99	-.073	.450	NS
PR	.26998	.00202	101.25	.062	.432	NS
(Constant)			12080.38			

II. MSS $\underline{N} = 166$ $\underline{F}\ (9,156) = 1.496$, NS

Subscale	R^2	Incremental R^2 Change	B	BETA	F	P
PSC	.01733	.01733	-458.88	-.180	2.94	NS
TAC	.03316	.01582	594.27	.190	2.68	NS
PR	.05794	.02478	-512.11	-.171	4.20	.05
I	.06848	.01054	375.86	.120	1.79	NS
PF	.07516	.00668	-379.74	.124	1.13	NS
KOJ	.07805	.00289	197.21	.070	.49	NS
A	.07862	.00056	73.80	.031	.09	NS
L	.07901	.00040	108.23	.036	.07	NS
WTW	.07946	.00044	- 91.13	-.027	.07	NS
(Constant)			11646.23			

* Subscale abbreviations are translated in Appendix E

Hierarchical Regressions Showing Unique Variance Accounted for by the BARS and MSS Subscales on the Seniority Criterion

I. BARS $N = 166$ $R = .51960$ $R^2 = .26998$

Subscale	R^2 Change	r	B	BETA	F	P
L *	.00265	-.298	-129.62	-.089	.566	NS
PF	.10257	.283	623.69	.389	21.92	.01
KOJ	.02250	-.282	-343.43	-.257	4.808	.05
I	.02334	.058	386.80	.211	4.988	.05
WTW	.01091	-.196	-268.35	-.158	2.33	NS
PSC	.00556	-.200	-202.71	-.123	1.19	NS
A	.00330	.049	-117.45	-.073	.71	NS
PR	.00190	-.064	100.51	.061	.41	NS
TAC	.00000	-.149	3.65	.002	-	NS

II. MSS $N = 178$ $R = .28188$ $R^2 = .07946$

Subscale	R^2 Change	r	B	BETA	F	P
PSC	.01258	-.132	-458.88	-.180	2.29	NS
TAC	.02687	.096	594.27	.190	4.90	.05
PR	.02144	-.120	-512.11	-.171	3.91	.05
I	.01195	.043	375.86	.120	2.18	NS
PF	.00701	-.129	-379.74	-.124	1.28	NS
KOJ	.00192	-.067	197.20	.070	3.50	NS
A	.00055	-.027	73.80	.031	.10	NS
L	.00060	-.040	108.23	.031	.11	NS
WTW	.00044	-.097	- 91.13	-.027	.08	NS

* Subscale abbreviations are translated in Appendix E

Stepwise Regressions for Captain Raters Showing Incremental Variance Accounted for by the BARS and MSS Subscales for the Seniority Criterion

I. BARS $\underline{N} = 57$ $\underline{F}$ (9,47) = 2.0588, NS

Subscale	R^2	Incremental R^2 Change	B	BETA	F	P
PF*	.11764	.11764	657.38	.379	7.71	.01
PSC	.22781	.11017	-346.56	-.211	7.22	.01
I	.25382	.02601	529.66	.251	1.70	NS
L	.26831	.01449	-117.67	-.081	.95	NS
PR	.27371	.00540	-241.72	-.136	.35	NS
WTW	.27727	.00355	-260.39	-.159	.23	NS
TAC	.28163	.00436	191.33	.112	.28	NS
KOJ	.28229	.00066	-100.24	-.063	.04	NS
A	.28276	.00047	46.45	.028	.03	NS
(Constant)			10342.9			

II. MSS $\underline{N} = 57$ $\underline{F}$ (9,47) = .5766, NS

Subscale	R^2	Incremental R^2 Change	B	BETA	F	P
PSC	.02226	.02226	-571.71	-.364	1.162	NS
A	.04737	.02510	263.52	.168	1.310	NS
PR	.07249	.02513	-332.73	-.183	1.311	NS
TAC	.08467	.01218	237.41	.110	.636	NS
KOJ	.09138	.00672	250.30	.137	.307	NS
I	.09816	.00678	195.74	.089	.354	NS
PF	.09864	.00048	84.50	.041	.025	NS
WTW	.09903	.00039	- 80.97	-.035	.023	NS
L	.09942	.00039	67.90	.035	.023	NS
(Constant)			11492.5			

* Subscale abbreviations are translated in Appendix E

Stepwise Regressions for Lieutenant Raters Showing Incremental Variance Accounted for by the BARS and MSS Subscales for the Seniority Criterion

I. BARS $N = 109$ $F\ (9,99) = 5.6088,\ p < .01$

Subscale	R^2	Incremental R^2 Change	B	BETA	F	P
KOJ*	.10620	.10620	-566.20	-.374	15.87	.01
PF	.22180	.11559	579.69	.382	17.28	.01
WTW	.25168	.02988	-385.80	-.221	4.47	.05
I	.29231	.04063	453.22	.271	6.07	.05
L	.30461	.01231	-229.26	-.156	1.84	NS
PR	.31927	.01466	347.63	.224	2.19	NS
A	.33028	.01101	-213.25	-.135	1.64	NS
TAC	.33510	.00481	-135.06	-.079	.72	NS
PSC	.33770	.00261	-135.38	-.082	.301	NS
(Constant)			13656.1			

II. MSS $N = 109$ $F\ (9,99) = 1.13$, NS

Subscale	R^2	Incremental R^2 Change	B	BETA	F	P
PF	.03108	.03108	-613.46	-.178	3.39	NS
TAC	.05459	.02352	705.26	.205	2.57	NS
PR	.07262	.01803	-689.20	-.194	1.97	NS
I	.08884	.01622	513.78	.150	1.77	NS
PSC	.09001	.00117	-242.54	-.079	.13	NS
L	.09182	.00181	221.06	.064	.19	NS
WTW	.09312	.00131	-176.76	-.047	.14	NS
KOJ	.09335	.00022	75.17	.023	.02	NS
(Constant)			11906.48			

* Subscale abbreviations are translated in Appendix E

Hierarchical Regressions for Lieutenant Raters Showing Unique Variance Accounted for by the BARS and MSS Subscales on the Seniority Criterion

I. BARS $\underline{N}$ = 109 $\underline{R}$ = .58112 $\underline{R}^2$ = .33770

Subscale	R^2 Change	r	B	BETA	F	P
KOJ*	.05881	-.326	-566.20	-.374	8.79	.01
PF	.08852	.254	579.70	.382	13.23	.01
WTW	.02655	-.220	-385.81	-.221	3.97	NS
I	.03306	.006	453.22	.271	4.94	.05
L	.01105	-.324	-229.26	-.156	1.65	NS
PR	.02519	-.05504	347.63	.224	3.76	NS
A	.01121	-.00268	-213.25	-.135	1.68	NS
TAC	.00313	-.19819	-135.06	-.078	.468	NS
PSC	.00261	-.16784	-135.38	-.082	.390	NS

II. MSS $\underline{N}$ = 121 $\underline{R}$ = .30557 $\underline{R}^2$ = .09337

Subscale	R^2 Change	r	B	BETA	F	P
PF	.01353	-.176	-624.75	-.182	1.66	NS
TAC	.03092	.106	700.86	.203	3.78	NS
PR	.02673	-.137	-691.57	-.195	3.27	NS
I	.01632	.019	509.54	.148	1.99	NS
PSC	.00232	-.099	-241.78	-.079	.28	NS
L	.00201	-.019	217.77	.063	.24	NS
WTW	.00121	-.117	-173.74	-.046	.14	NS
KOJ	.00023	-.064	75.49	.023	.03	NS
A	.00002	-.057	17.76	.007	.00	NS

* Subscale abbreviations are translated in Appendix E

Hierarchical Regressions for Captain Raters Showing Unique Variance Accounted for by the BARS and MSS Subscale on the Seniority Criterion

I. BARS $\underline{N} = 57$ $\underline{R} = .53175$ $\underline{R}^2 = .28276$

Subscale	R^2 Change	r	B	BETA	F	P
PF*	.09965	.34299	657.38	.379	6.53	.05
PSC	.01304	-.22922	-346.56	-.211	.85	NS
I	,03257	.15873	529.67	.251	2.13	NS
PR	.00824	-.071	-241.72	-.136	.54	NS
L	.00092	-.248	-117.68	-.081	.06	NS
WTW	.00659	.154	-260.39	-.160	.43	NS
TAC	.00431	.057	191.33	.112	.28	NS
KOJ	.00054	-.202	-100.24	-.063	.03	NS
A	.00047	.119	46.45	.028	.03	NS

II. MSS $\underline{N} = 57$ $\underline{R} = .31532$ $\underline{R}^2 = .09942$

Subscale	R^2 Change	r	B	BETA	F	P
PSC	.05011	-.149	-571.71	-.364	2.61	NS
A	.01416	.099	263.52	.168	.74	NS
PR	.02257	-.098	-332.73	-.183	1.18	NS
TAC	.00803	.068	237.41	.110	.42	NS
KOJ	.00460	-.034	250.31	.137	.24	NS
I	.00560	.088	195.00	.090	.29	NS
PF	.00077	.020	84.50	.041	.04	NS
WTW	.00065	-.008	- 80.97	-.035	.03	NS
L	.00039	-.035	67.90	.035	.02	NS

* Subscale abbreviations are translated in Appendix E

APPENDIX L

Summary of the Education

Criterion Regressions

BARS and MSS Total Score Regressions on the Education Criterion

I. All Raters $N = 178$

Scale	R^2	r	B	BETA	F	P
BARS	.07667	-.2769	-1.782	-.277	14.697	.01
MSS	.01994	-.1412	- .977	-.141	3.580	

II. Lieutenant Raters $N = 121$

Scale	R^2	r	B	BETA	F	P
BARS	.05076	-.2253	-1.475	-.225	6.365	.01
MSS	.00496	-.0704	- .517	-.070	.593	NS

III. Captain Raters $N = 57$

Scale	R^2	r	B	BETA	F	P
BARS	.11129	-.3336	-2.062	-.334	7.379	.01
MSS	.0570	-.2388	-1.499	-.239	3.385	

Stepwise Regressions Showing Incremental Variances Accounted for by the BARS and MSS Subscales on Education Criterion

I. BARS $N = 179$ F (8,170) = 4.1668, $p < .01$

Subscale	R^2	Incremental R^2 Change	b	BETA	F	P
WTW*	.10997	.10997	-.297	-.283	22.361	.01
L	.12494	.01497	.082	-.092	3.044	NS
I	.14019	.01524	.168	.152	3.099	NS
PF	.14685	.00666	.140	.141	1.354	NS
A	.15502	.00818	-.133	-.132	1.663	NS
KOJ	.16255	.00752	-.127	-.138	1.529	NS
PR	.16360	.00105	.046	.046	.213	NS
PSC	.16394	.00034	-.309	-.031	.069	NS
(Constant)				2.654		

II. MSS $N = 178$ F (8,169) = 2.587, $p < .01$

Subscale	R^2	Incremental R^2 Change	b	BETA	F	P
KOJ	.06271	.06271	-.206	-.249	11.896	.01
TAC	.08712	.02442	.170	.185	4.632	.1
PR	.09409	.00696	-.106	-.121	1.320	NS
I	.10336	.00927	.086	.094	1.758	NS
PSC	.10584	.00248	-.059	-.078	.470	NS
A	.10783	.00200	.048	.070	.379	NS
PF	.10891	.00108	-.041	-.046	.205	NS
L	.10910	.00019	.171	.019	.036	NS
(Constant)			1.365			

* Subscale abbreviations are translated in Appendix E

Hierarchical Regressions Showing Unique Variance Accounted for by the Nine BARS and MSS Subscales on the Education Criterion

I. BARS $\underline{N}$ = 179 $\underline{R}$ .40489 $\underline{R}^2$ = .16394

Subscale	R^2 Change	r	B	BETA	F	P
WTW*	.03496	-.33162	-.297	-.283	7.067	.01
L	.00280	-.2882	-.082	-.092	.566	NS
I	.01172	-.1013	.168	.152	2.369	NS
PF	.01347	.0036	.140	.142	2.723	NS
A	.01072	-.1517	-.133	-.132	2.167	NS
KOJ	.00639	-.2788	-.127	-.138	1.292	NS
PR	.00105	-.1575	.047	.046	.212	NS
PSC	.00033	-.2279	-.031	-.030	.067	NS
TAC	.00000	-.2387	-.001	-.001	-	NS

II. MSS $\underline{N}$ = 178 $\underline{R}$ = .33036 $\underline{R}^2$ = .10914

Subscale	R^2 Change	r	B	BETA	F	P
KOJ	.02445	-.250	-.207	-.250	4.611	.1
TAC	.02539	.077	.170	.185	4.788	.1
PR	.01039	-.082	.105	-.119	1.959	NS
I	.00743	.039	.087	.095	1.401	NS
PSC	.00233	-.204	-.058	-.078	.439	NS
A	.00278	-.011	.048	.070	.524	NS
PF	.00085	-.135	-.039	-.043	.160	NS
L	.00022	-.136	.019	.023	.041	NS
WTW	.00004	-.096	.008	-.008	.007	NS

*Subscale abbreviations are translated in Appendix E

Stepwise Regressions for Lieutenant Raters Showing Incremental Variance Accounted for by the BARS and MSS Subscales on the Education Criterion

I. BARS $\underline{N} = 121$ $\underline{F}\ (9,111) = 2.09,\ \underline{p} < .06$

Subscale	R^2	Incremental R^2 Change	b	BETA	F	P
WTW*	.08298	.08298	-.306	-.287	10.774	.01
PF	.09270	.00972	.147	.160	1.262	NS
A	.10579	.01309	-.213	-.218	1.699	NS
PSC	.11670	.01091	-.146	-.149	1.417	NS
I	.13646	.01976	.222	.224	2.566	NS
KOJ	.14287	.00640	.110	-.124	.831	NS
PR	.14424	.00138	.042	.045	.179	NS
TAC	.14485	.00061	.039	.037	.079	NS
L	.14511	.00026	.213	-.024	.034	NS
(Constant)			2.800			

II. MSS $\underline{N} = 121$ $\underline{F}\ (8,112) = 1.46$, NS

Subscale	R^2	Incremental R^2 Change	b	BETA	F	P
KOJ	.02809	.02809	-.16760	-.229	3.475	NS
TAC	.07256	.04447	.15778	.215	5.501	.05
PSC	.07840	.00584	-.073	.099	.722	NS
WTW	.08319	.00478	-.101	-.064	.591	NS
I	.08773	.00455	-.001	.798	.563	NS
PR	.09361	.00587	-.050	-.075	.726	NS
PF	.09431	.00071	-.091	-.041	.088	NS
A	.09464	.00032	.003	.016	.040	NS
(Constant)				.874		

* Subscale abbreviations are translated in Appendix E

Hierarchical Regression for Lieutenant Raters Showing Unique Variance Accounted for by the Nine BARS and MSS subscales on the Education Criterion

I. BARS $\underline{N}$ = 121 $\underline{R}$ = .38094 $\underline{R}^2$ = .14511

Subscale	R^2 Change	r	B	BETA	F	P
WTW*	.04415	-.288	-.306	-.287	5.732	.05
PF	.01560	-.018	.147	.160	2.025	NS
A	.02873	-.189	-.213	-.218	3.730	NS
PSC	.00818	-.179	-.146	-.149	1.062	NS
I	.02214	-.084	.223	.224	2.874	NS
KOJ	.00613	-.197	-.109	-.124	.796	NS
PR	.00096	-.135	.042	.045	.125	NS
TAC	.00069	-.176	.039	.037	.089	NS
L	.00026	-.182	-.021	-.024	.033	NS

II. MSS $\underline{N}$ = 121 $\underline{R}$ = .30764 $\underline{R}^2$ = .09464

Subscale	R^2 Change	r	B	BETA	F	P
KOJ	.03482	-.168	-.230	-.290	4.269	NS
TAC	.04805	.158	.215	.254	5.891	.05
PSC	.00638	-.073	.098	.131	.782	NS
WTW	.00277	-.101	-.064	-.070	.340	NS
I	.00668	-.001	.080	.095	.819	NS
PR	.00515	-.050	-.074	-.086	.631	NS
PF	.00097	-.091	-.041	-.048	.119	NS
A	.00031	.003	.016	.024	.038	NS
L	.00001	-.054	.003	.004	.001	NS

* Subscale abbreviations are translated in Appendix E

Stepwise Regressions for Captain Raters Showing Incremental Variance Accounted for by the Nine BARS and MSS Subscales on the Education Criterion

I. BARS $N = 57$ $F\ (6,50) = 2.0359$, NS

Subscale	R^2	Incremental R^2 Change	b	BETA	F	P
KOJ*	.13067	.13067	-.277	-.315	8.130	.01
PSC	.14787	.01720	-.212	-.281	1.070	NS
A	.16248	.01461	.136	.180	.909	NS
PR	.18389	.02141	-.147	-.168	1.332	NS
L	.19271	.00882	.119	.130	.549	NS
TAC	.19634	.00363	.073	.071	.226	NS
(Constant			2.571			

II. MSS $N = 58$ $F\ (8,49) = 1.78$, NS

Subscale	R^2	Incremental R^2 Change	b	BETA	F	P
L	.17813	.17813	-.377	-.365	11.268	.01
PF	.19708	.01894	.171	.155	1.198	NS
WTW	.21028	.01320	-.222	-.215	.835	NS
PSC	.21961	.00933	.192	.184	.590	NS
KOJ	.22288	.00327	-.114	-.113	.207	NS
I	.22434	.00146	.054	.040	.092	NS
A	.22473	.00038	-.038	.036	.024	NS
PR	.22539	.00067	.043	.037	.042	NS
(Constant)			2.249			

*Subscale abbreviations are translated in Appendix E

Hierarchical Regression for Captain Raters Showing Incremental Variance Accounted for by the Nine BARS and MSS Subscales on the Education Criterion

I. BARS $N = 58$ $R = .47483$ $R^2 = .22546$

Subscale	R^2 Change	r	B	BETA	F	P
L*	.01872	-.422	-.339	-.367	1.184	NS
PF	.01652	.050	.169	.154	1.045	NS
WTW	.01111	-.375	-.214	-.208	.702	NS
PSC	.00997	-.283	.191	.183	.631	NS
KOJ	.00166	-.383	-.110	-.109	.105	NS
I	.00093	-.119	.057	.042	.059	NS
A	.00069	-.100	-.036	-.034	.044	NS
PR	.00073	-.187	.045	.040	.046	NS
TAC	.00007	-.300	-.015	-.014	.004	NS

II. MSS $N = 57$ $R = .44325$ $R^2 = .19647$

Subscale	R^2 Change	r	B	BETA	F	P
KOJ	.02536	-.36148	-.283	-.321	1.578	NS
PSC	.03021	-.355	-.214	-.282	1.880	NS
A	.01664	-.020	.137	.182	1.035	NS
PR	.01859	-.126	-.145	-.166	1.157	NS
L	.00579	-.235	.125	.136	0.360	NS
TAC	.00308	-.058	.025	.068	.192	NS
I	.00000	.093	.001	.001	-	-
PF	.00007	-.199	.012	.012	.004	NS
WTW	.00009	-.078	-.014	-.013	.006	NS

* Subscale abbreviations are translated in Appendix E

Appendix M

Summary of the Promotion Related
Experience Criterion Regressions

BARS and MSS Total Score Regressions
on Promotion Related Experience Criterion

I. All Raters N = 155

Scale	R^2	r	B	BETA	F	P
BARS	.05226	.2286	.089	.229	8.441	.01
MSS	.04335	.2082	.076	.208	6.841	.01

II. Lieutenant Raters N = 101

Scale	R^2	r	B	BETA	F	P
BARS	.03997	.19991	.074	.200	4.121	.05
MSS	.00754	.08681	.029	.087	.752	NS

III. Captain Raters N = 51

Scale	R^2	r	B	BETA	F	P
BARS	.08031	.28339	.102	.283	4.279	.05
MSS	.16107	.40134	.138	.401	9.408	.01

Stepwise Regressions Showing Incremental Variance Accounted for by the BARS and MSS Subscales on the Promotion Related Experience Criterion

I. BARS $N = 155$ $F\ (7, 147) = 3.9512$ $p < .05$

Subscale	R^2	Incremental R^2 Change	b	BETA	F	P
L*	.12669	.12669	.575	.284	4.422	.01
TAC	.13813	.01145	-.322	-.138	1.536	NS
KOJ	.14632	.00819	.331	.157	1.512	NS
PF	.15625	.00993	-.267	-.119	1.975	NS
PSC	.15764	.00139	.131	.057	.217	NS
WTW	.15802	.00038	.068	.029	.068	NS
PR	.15836	.00033	-.053	-.023	.058	NS
(constant)			.928			

II. MSS $N = 153$ $F\ (9, 143) = 3.938$ $p < .05$

Subscale	R^2	Incremental R^2 Change	b	BETA	F	P
PSC	.14711	.14711	.643	.365	9.573	.01
I	.16229	.01518	-.267	-.127	2.286	NS
TAC	.17315	.01085	-.296	-.140	2.577	NS
PR	.18486	.01172	.270	.139	2.538	NS
L	.18878	.00391	.136	.068	.381	NS
PF	.19339	.00461	-.198	-.095	.742	NS
KOJ	.19594	.00255	.155	.081	.444	NS
A	.19765	.00170	-.092	-.058	.334	NS
WTW	.19865	.00100	-.090	.040	.179	NS
(constant)			1.127			

* Subscale abbreviations are translated in Appendix E.

Hierarchical Regressions Showing Unique Variance accounted for by the nine BARS and MSS Subscales on the Promotion Related Experience Criterion

I. BARS N = 155 R = .39753 R^2 = .15803

Subscale	R^2 Change	r	b	BETA	F	P
L*	.02535	.35593	.576	.284	4.367	.01
TAC	.00872	.11332	-.324	-.139	1.503	NS
KOJ	.00852	.31226	.330	.157	1.467	NS
PF	.01004	-.05566	-.274	-.123	1.730	NS
PSC	.00124	.25970	.135	.058	.214	NS
WTW	.00031	.18850	.062	.026	.053	NS
A	.00002	-.01028	.013	.006	.004	NS
PR	.00036	.13275	-.061	-.027	.062	NS
I	.00001	.12377	.013	.005	.002	NS

II. MSS N = 153 R = .44570 R^2 = .19865

Subscale	R^2 Change	r	b	BETA	F	P
PSC	.05364	.38355	.643	.365	9.573	.01
I	.01281	-.05016	-.267	-.127	2.286	NS
TAC	.01444	-.03926	-.296	-.140	2.577	NS
PR	.01422	.07505	.270	.139	2.538	NS
L	.00214	.27528	.136	.068	.381	NS
PF	.00416	.11872	-.198	-.098	.742	NS
KOJ	.00249	.30570	.155	.080	.444	NS
A	.00187	.02540	-.092	-.058	.334	NS
WTW	.001	.13904	.090	.041	.179	NS

* Subscale abbreviations are translated in Appendix E.

Stepwise Regressions for Lieutenant Raters Showing Incremental Variance Accounted for by the BARS and MSS Subscales on the Promotion Related Experience Criterion

I. BARS $N = 101$ $F(8,92) = 2.163$ $p < .05$

Incremental R^2

Subscale	R^2	Change	b	BETA	F	P
KOJ*	.10343	.10343	.684	.315	4.525	.01
L	.12351	.02008	.641	.305	4.212	.01
WTW	.14063	.01712	-.265	-.104	.709	NS
PSC	.15003	.00940	-.399	-.167	1.218	NS
PR	.15480	.00477	-.136	-.060	.217	NS
A	.15653	.00173	-.147	-.064	.294	NS
PF	.15813	.00160	-.118	.054	.192	NS
I	.15830	.00017	-.047	-.019	.019	NS
(constant)			.859			

II. MSS $N = 101$ $F(9,91) = 2.457$ $p < .05$

Incremental R^2

Subscale	R^2	Change	b	BETA	F	P
PSC	.11269	.11269	.887	.466	10.620	.01
A	.15644	.04375	-.214	-.133	1.045	NS
I	.17534	.01890	-.391	-.192	2.864	NS
PR	.18028	.00494	.179	.089	.649	NS
WTW	.18365	.00337	.282	.132	1.124	NS
PF	.18931	.00566	-.260	-.125	.780	NS
L	.19465	.00534	-.210	-.103	.628	NS
KOJ	.19507	.00042	.069	.035	.061	NS
TAC	.19550	.00043	-.050	-.024	.049	NS
(constant)			1.562			

* Subscale abbreviations are translated in Appendix E.

Hierarchical Regressions for Lieutenant Raters Showing Unique Variance Accounted for by the BARS and MSS Subscales for the Promotion Related Experience Criterion

I. BARS $N = 101$ $R = .39789$ $R^2 = .15831$

Subscale	R^2 Change	r	b	BETA	F	P
KOJ*	.04140	.32161	.684	.315	4.476	.01
L	.03770	.32078	.640	.305	4.076	.01
WTW	.00599	.06528	-.267	-.105	.648	NS
PSC	.01077	.15366	-.400	-.167	1.164	NS
PR	.00188	.12929	-.138	-.061	.203	NS
A	.00264	-.0265	-.148	-.065	.286	NS
PF	.00175	.00576	.118	.054	.189	NS
I	.00016	.13785	-.046	-.019	.017	NS
TAC	.00000	.11345	.006	.002	.000	NS

II. MSS $N = 101$ $R = .44214$ $R^2 = .19549$

Subscale	R^2 Change	r	b	BETA	F	P
PSC	.09389	.33569	.887	.466	10.620	.01
A	.00924	-.09637	-.214	-.133	1.045	NS
I	.02532	-.07931	.391	-.193	2.864	.05
PR	.00574	-.00750	.179	.089	.649	NS
WTW	.00994	.15043	.282	.132	1.124	NS
PF	.00690	.02553	-.260	-.125	.780	NS
L	.00553	.12527	-.210	-.103	.626	NS
KOJ	.00054	.19104	.069	.035	.061	NS
TAC	.00043	-.04339	-.050	-.024	.049	NS

---

* Subscale abbreviations are listed in Appendix E.

Stepwise Regressions for Captain Raters Showing Incremental Variance Accounted for by the BARS and MSS Subscales on the Promotion Related Experience Criterion

I. BARS $N = 52$ $F\ (9,42) = 2.2955$, $p < .05$

Subscale	R^2	Incremental R^2 Change	b	BETA	F	P
PSC *	.17835	.17835	.845	.372	2.553	NS
PF	.24368	.06533	-.616	-.255	2.918	NS
TAC	.25796	.01428	-.925	-.399	3.518	NS
WTW	.30966	.05170	.740	.334	1.708	NS
A	.31572	.00606	.176	.078	.241	NS
L	.32097	.00525	.602	.298	.737	NS
KOJ	.32730	.00633	-.441	-.204	.352	NS
I	.32880	.00150	.185	.063	.133	NS
PR	.32971	.00091	-.108	-.044	.057	NS
(Constant)			.837			

II. MSS $N = 51$ $F\ (9,41) = 3.5202$ $P < .01$

Subscale	R^2	Incremntal R^2 Change	b	BETA	F	P
L	.24763	.24763	1.057	.531	6.184	.01
PR	.29045	.04371	.633	.337	5.464	.01
TAC	.37700	.08655	-.824	-.367	6.464	.01
PSC	.39829	.02129	.247	.148	.566	NS
WTW	.41767	.01938	-.533	-.221	1.986	NS
A	.43458	.01692	.276	.172	1.109	NS
KOJ	.43543	.00085	.097	.052	.043	NS
PF	.43575	.00032	-.053	-.025	.018	NS
I	.43589	.00014	-.032	-.014	.010	NS
(Constant)			-.154			

*Subscale abbreviations are translated in Appendix E.

Hierachical Regressions for Captain Raters Showing Unique Variance Accounted for by the BARS and MSS Subscales for the Promotion Related Experience Criterion

I. BARS $\underline{N} = 52$ $\underline{R} = .57420$ $\underline{R}^2 = .32971$

Subscale	R^2 Change	r	b	BETA	F	P
PSC *	-.04075	.42232	.845	.372	2.553	NS
PF	.04658	-.16013	-.616	-.255	2.918	NS
TAC	.05614	.10766	-.925	-.399	3.518	NS
WTW	.02726	.34693	.740	.334	1.708	NS
A	.00384	.01945	.176	.078	.241	NS
L	.01177	.41298	.602	.298	.737	NS
KOJ	.00562	.32727	-.441	-.204	.744	NS
I	.00213	.11401	.185	.063	.133	NS
PR	.00091	.14069	-.108	-.044	.057	NS

II. MSS $\underline{N} = 51$ $\underline{R} = .66022$ $\underline{R}^2 = .43589$

Subscale	R^2 Change	r	b	BETA	F	P
L	.08509	.49672	1.057	.530	6.184	.01
PR	.07517	.19520	.633	.337	5.464	.01
TAC	.08894	-.08894	-.824	-.367	6.464	.01
PSC	.00779	.45818	.247	.148	.560	NS
WTW	.02733	.12132	-.533	-.221	1.986	NS
A	.01524	.20737	.276	.172	1.108	NS
KOJ	.00060	.46758	.097	.052	.043	NS
PF	.00025	.26119	-.053	-.025	.018	NS
I	.00014	.00265	-.032	-.014	.010	NS

* Subscale abbreviations are translated in Appendix E.

Appendix N

Summary of the Practical Performance Test Criterion Regressions

BARS and MSS Total Score Regressions
on the Practical Performance Test Criterion

I. All Raters N = 178

Scale	R^2	r	B	BETA	F	P
BARS	.00111	.0334	.2508	.0334	.197	NS
MSS	.00534	.0731	.591	.0731	.945	NS

II. Lieutenant Raters N = 121

Scale	R^2	r	B	BETA	F	P
BARS	.00032	-.0180	-.130	-.0180	.038	NS
MSS	.0000	-.0050	-	-.0050	-	-

III. Captain Raters N = 57

Scale	R^2	r	B	BETA	F	P
BARS	.0146	.1209	.9456	.1209	.831	NS
MSS	.0483	.2198	1.745	.2198	2.792	NS

Stepwise Regressions Showing Incremental Variance Accounted for by the BARS and MSS Subscales on the Practical Performance Test Criterion

I. BARS $\underline{N}$ = 134 $\underline{F}$ (10, 123) = .6645, NS

Subscale	R^2	Incremental R^2 Change	B	BETA	F	P
Test Order	.00134	.00134	-.010	-.034	.174	NS
WTW*	.00979	.00844	.122	.132	1.095	NS
KOJ	.02695	.01716	-.127	-.152	2.225	NS
PSC	.03146	.00452	.129	.141	.586	NS
L	.03859	.00713	-.116	-.144	.925	NS
TAC	.04265	.00406	.108	.121	.525	NS
PR	.04685	.00420	-.092	-.101	.545	NS
PF	.04935	.00250	-.069	-.076	.324	NS
I	.05109	.00175	.056	.054	.227	NS
A	.05162	.00053	.025	.029	.069	NS
(constant)			-.101			

II. MSS $\underline{N}$ = 133 $\underline{F}$ (9, 123) = 1.4748, NS

Subscale	R^2	Incremental R^2 Change	B	BETA	F	P
Test Order	.00086	.00086	-.0166	-.055	.117	NS
A	.03655	.03570	.175	.276	4.865	.1
PSC	.06917	.03261	-.126	-.187	4.444	.1
L	.08101	.01184	.169	.210	1.613	NS
KOJ	.09154	.01053	-.117	-.155	1.435	NS
I	.09438	.00283	-.052	-.063	.386	NS
TAC	.09636	.00198	-.046	-.054	.270	NS
PR	.09694	.00058	.237	.032	.079	NS
PF	.09740	.00046	-.024	-.030	.063	NS
(constant)			.227			

* Subscale abbreviations are translated in Appendix E.

Hierarchical Regression Showing Unique Variance Accounted for by the BARS and MSS Subscales on the Practical Performance Test Criterion

I. BARS $\underline{N}$ = 134 $\underline{R}$ = .22720 $\underline{R}^2$ = .05162

Subscale	R^2 Change	r	B	BETA	F	P
WTW*	.00736	.09486	.122	.132	.954	NS
KOJ	.00808	-.04465	-.127	-.152	1.048	NS
PSC	.00788	.046	.129	.141	1.023	NS
L	.00737	-.043	-.116	-.144	.956	NS
TAC	.00628	.087	.108	.121	.814	NS
PR	.00533	-.031	-.019	-.101	.691	NS
PF	.00394	-.014	-.069	-.076	.511	NS
I	.00162	.048	.056	.054	.210	NS
A	.00053	.037	.025	.029	.069	NS

II. MSS $\underline{N}$ = 133 $\underline{R}$ = .31211 $\underline{R}^2$ = .09741

Subscale	R^2 Change	r	B	BETA	F	P
A	.04315	.18976	.17512	.27574	5.83	.05
PSC	.01346	-.095	-.126	-.186	1.82	NS
L	.01814	.063	.168	.209	2.45	NS
KOJ	.00862	-.079	-.118	-.156	1.16	NS
I	.00319	-.002	-.052	-.064	.43	NS
TAC	.00202	.034	-.046	-.054	.27	NS
PR	.00074	.061	.023	.031	.10	NS
PF	.00047	.007	-.025	-.032	.06	NS
WTW	.00002	.02063	.005	.005	.00	NS

* Subscale abbreviations are translated in Appendix E.

Stepwise Regressions for Lieutenant Raters Showing Incremental Variance Accounted for by the BARS and MSS Subscales on the Practical Performance Test Criterion

I. BARS $N = 88$ $F\ (9, 78) = .9877$, NS

Subscale	Incremental R^2	R^2 Change	B	BETA	F	P
Test Order	.00043	.00043	-.016	-.054	.037	NS
KOJ*	.03052	.03009	-.217	-.237	2.61	NS
WTW	.05561	.02509	.148	.141	2.22	NS
PF	.06540	.00979	-.209	-.224	.851	NS
A	.07928	.01388	.141	.152	1.21	NS
PSC	.08841	.00913	.179	.178	.793	NS
L	.10101	.01261	-.138	-.154	1.096	NS
TAC	.10213	.00111	.048	.049	.096	NS
PR	.10231	.00019	-.017	-.018	.016	NS
(constant)			.539			

II. MSS $N = 88$ $F\ (10, 77) = 1.511$, NS

Subscale	Incremental R^2	R^2 Change	B	BETA	F	P
Test Order	.00043	.00043	-.0215	-.0709	.0396	NS
KOJ	.07957	.07914	-.3025	-.3786	7.2897	.01
A	.12183	.04226	.0891	.1357	3.893	NS
I	.13310	.01127	.1187	.1432	1.038	NS
L	.14502	.01193	.1830	.2091	1.099	NS
PSC	.15806	.01304	-.1273	-.1627	1.201	NS
PR	.16099	.00293	.0538	.0667	.270	NS
TAC	.16306	.00207	-.0481	.0558	.191	NS
WTW	.16373	.00067	-.0365	-.0429	.062	NS
PF	.16406	.00033	.0223	.0284	.030	NS
(constant)			.4896			

* Subscale abbreviations are translated in Appendix E.

Hierarchical Regressions for Lieutenant Raters Showing Unique Variance Accounted for by the BARS and MSS Subscales on the Practical Performance Test Criterion

I. BARS $N = 88$ $R = .31987$ $R^2 = .10231$

Subscale	R^2 Change	r	B	BETA	F	P
KOJ*	.02711	-.1715	-.2173	-.2375	2.325	NS
WTW	.01191	.0590	.1484	.1414	1.022	NS
PF	.03011	-.1209	-.2095	-.2237	2.583	NS
A	.01528	.0902	.1415	.1521	1.311	NS
PSC	.01390	-.0479	.1795	.1776	1.192	NS
L	.01245	-.1190	-.1386	-.1542	1.068	NS
TAC	.00125	.0101	.4851	.0493	.107	NS
PR	.00018	-.0570	-.1716	-.0181	.015	NS
I	.00000	-.0716	.0002	.0002	-	NS

II. MSS $N = 88$ $R = .40504$ $R^2 = .16406$

Subscale	R^2 Change	r	B	BETA	F	P
KOJ	.06526	-.278	-.302	-.379	6.011	.05
A	.00874	.076	.089	.136	.081	NS
I	.01391	.069	.119	.143	1.28	NS
L	.01960	-.023	.183	.209	1.80	NS
PSC	.01109	-.207	-.127	-.163	1.02	NS
PR	.00331	.059	.054	.067	.305	NS
TAC	.00214	-.026	-.048	-.056	.197	NS
WTW	.00096	-.104	-.036	-.043	.009	NS
PF	.00033	-.051	.022	.028	.034	NS

* Subscale abbreviations are translated in Appendix E.

Stepwise Regressions for Captain Raters Showing Incremental Variance Accounted for by the BARS and MSS Subscales on the Practical Performance Test Criterion

I. BARS $\underline{N} = 46$ $\underline{F}$ (10,35) = .6606, NS

Incremental R^2

Subscale	R^2	Change	B	BETA	F	P
Test Order	.00348	.00348	.0328	.1103	.1448	NS
I*	.06180	.05832	.3006	.2828	2.4227	NS
A	.08649	.02469	-.1593	-.1986	1.0272	NS
PF	.10146	.01497	.0923	.1071	.6228	NS
TAC	.11783	.01637	.2634	.3205	.6811	NS
PR	.13555	.01772	-.1323	-.1513	.7373	NS
WTW	.14170	.00615	-.0979	-.1222	.2559	NS
PSC	.14811	.00640	.1642	.2042	.2663	NS
L	.15834	.01024	-.1878	-.2631	.4260	NS
KOJ	.15878	.00044	.0551	.0722	.0183	NS
(constant)			-1.7003			

II. MSS $\underline{N} = 45$ $\underline{F}$ (10,34) = 1.379, NS

Incremental R^2

Subscale	R^2	Change	B	BETA	F	P
Test Order	.00140	.00140	.0225	.0757	.0669	NS
A	.13158	.13017	.3288	.5418	6.221	.05
I	.20142	.06985	-.2233	-.2717	3.338	NS
PR	.21963	.01821	-.1065	-.1597	.870	NS
WTW	.23723	.01760	.1390	.1480	.841	NS
PSC	.26197	.02475	-.2002	-.3410	1.183	NS
KOJ	.27851	.01654	.1663	.2637	.790	NS
PF	.28453	.00602	-.0776	-.0975	.288	NS
TAC	.28655	.00202	-.0532	-.0650	.096	NS
L	.28858	.00203	.0670	.0930	.097	NS
(constant			-.0753			

---

* Subscale abbreviations are translated in Appendix E

Hierarchical Regressions for Captain Raters Showing Unique Variance Accounted for by the BARS and MSS Subscales on the Practical Performance Test Criterion

I. BARS $\underline{N} = 46$ $\underline{R} = .39848$ $\underline{R}^2 = .15878$

Subscale	R^2 Change	r	B	BETA	F	P
I*	.03687	.244	.301	.283	1.534	NS
A	.02147	-.041	-.159	-.199	.893	NS
PF	.00687	.144	.092	.107	.286	NS
TAC	.02490	.182	.263	.321	1.036	NS
PR	.00926	.001	-.132	-.151	.385	NS
WTW	.00303	.136	.020	-.122	.126	NS
PSC	.01134	.166	.164	.204	.472	NS
L	.00637	.043	-.188	-.263	.265	NS
KOJ	.00044	.114	.055	.072	.018	NS

II. MSS $\underline{N} = 45$ $\underline{R} = .53719$ $\underline{R}^2 = .28858$

Subscale	R^2 Change	r	B	BETA	F	P
A	.14711	.36131	.3288	.542	7.031	.01
I	.04793	-.10799	-.2233	-.272	2.291	NS
PR	.01493	.06058	-.1065	-.159	.713	NS
WTW	.01134	.2503	.1390	.148	.542	NS
PSC	.03732	.0183	-.2002	-.341	1.784	NS
KOJ	.00883	.2027	.1663	.237	.422	NS
PF	.00390	.1055	-.0776	-.097	.186	NS
TAC	.00265	.1307	-.0532	-.065	.127	NS
L	.00203	.1721	.0670	.930	.097	NS

* Subscale abbreviations are listed in Appendix E.

Appendix O

Summary of the Expectancy-Value Theory Component Regressions

Expectancy-Value Theory Component

Regressions

I. Intention N = 59 F (2,56) = 2.715, p .1

VARIABLE	R	R^2	r	B	BETA	F	P
SUM NBX MC	.287	.082	.287	.008	.291	5.31	.05
SUM BX EB	.290	.084	.018	.001	.043	.12	NS
SUM BX EB	.0182	.000	.000	.001	.043	.021	NS
SUM NB XMC	.290	.084	.084	.008	.291	5.40	.05

II. Attitude-Behavior Scale Preference

VARIABLE	R^2	r	B	BETA	F	P
SUM BX EB	.1141	.3378	.0152	.3378	7.598	.01

III. Subjective Norm

VARIABLE	R^2	r	B	BETA	F	P
SUM NB XMC	.001	.034	.0012	.034	.098	NS

APPROVAL SHEET

The thesis submitted by Robert L. Holmgren has been read and approved by the following committee:

Dr. John D. Edwards
Associate Professor, Psychology, Loyola

Dr. Homer Johnson
Professor, Psychology, Loyola

Mr. Robert T. Joyce, M.A.
Deputy Commissioner of Personnel
City of Chicago Department of Personnel

The final copies have been examined by the director of the thesis and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the thesis is now given final approval by the Committee with reference to content and form.

The thesis is therefore accepted in partial fulfillment of the requirements for the degree of Master of Arts.

April 24, 1981 — Date

John D. Edwards — Director's Signature