ABSTRACT

THOMPSON, JOHN ALEXANDER III The Relationship Between Scaled Behavioral Ratings, Performance Dimension Ratings and Rankings Using United States Army Special Forces Soldiers. (Under the direction of Mark A. Wilson)

This study measured job performance of United States Army Special Forces (SF) soldiers using a Mixed Standard Rating Scale (MSRS) format, performance dimension ratings, and rankings. Each SF soldiers performance was examined with a three dimension conceptual model (i.e. soldiering skills, Special Forces specific skills, and team member skills). This model was tested on two different samples. The first sample was made up of 1,273 U.S. Army Special Forces Soldiers who were rated by officers. The second sample was comprised of 1,121 U.S. Army Special Forces Soldiers who were rated by noncommissioned officers. This study found that the three dimensions of performance were highly intercorrelated with the highest correlation being between the soldiering skills and SF specific dimensions. In addition to correlation analysis, several regression models were also used to analyze each of the three dimensions. A full model was used to regress each of the nine sub-dimensions with the three dimensions of performance. Three reduced models were also made by regressing three sub-dimensions with each of the three dimensions of performance. It was found that each dimension used significantly different regression models. As well, the three reduced models were found to be significant in explaining each dimension. However, the reduced models were unable to explain a significant amount of variance when compared to the full model. Additionally, this study concluded that including all the items (positive, neutral, negative) improved each dimensional model's multiple correlation coefficient, but only marginally.

The Relationship Between Scaled Behavioral Ratings, Performance Dimension Ratings and Rankings Using United States Army Special Forces Soldiers

by

JOHN ALEXANDER THOMPSON III

A thesis submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the Master Degree in Psychology

Raleigh

2001

APPROVED BY:

Mark A. Wilson, Chair

J.W. Cunningham

D. W. Drewes

DEDICATION

I would like to dedicate this dissertation to my Advisor, Dr. Mark A. Wilson. His interest and teaching style have given me countless opportunities to expand my skills and knowledge. I have learned a great deal form him about research and practice, and admire his rare ability to combine the two to address both theoretical questions and practical Human Resource problems. His support and patience at times when I wasn't finding the journey very rewarding are greatly appreciated. I want to thank him for his assistance both as an Advisor and a friend.

BIOGRAPHY

John A. Thompson (Jat), III was born on November 07, 1972 in Warner Robbins, GA. Jat's family resides in Lexington, KY. After completing high school in Lexington, he attended the University of Kentucky on a Cheerleading scholarship. He graduated with a BSA in Accounting and Management and a BBA in Finance in 1996.

After graduation he spent three years working as a consultant for Ernst & Young. In the Spring of 1999 he was accepted into the Industrial Organizational Vocational Psychology program and North Carolina State. He moved to Raleigh that summer and began working with the Special Forces and attending classes that fall. He has spent the last three years as a full time graduate student and a part time consultant with the Special Forces.

ACKNOWLEDGMENTS

I would like to thank, first and for most, my Lord and Savior for giving me the gifts to be where I am now. In reality he deserves all the credit. I would like to thank my family for supporting in my tough decision to leave work to return to school. I want to thank my advisor, Mark A. Wilson, for giving me the opportunity to attend North Carolina State University. I would also like to thank Dr. Sanders and the Special Forces for allowing me to work with them. They are truly an amazing organization and it has been my privilege to work so closely with so many honorable men. Finally, I would like to thank Eric Surface for mentoring me and making my transition into graduate school a smooth one.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

The Relationship Between Scaled Behavioral Ratings, Performance Dimension Ratings and Rankings Using United States Army Special Forces Soldiers.

One challenge that most organizations face is determining the value and utility each employee adds to the organization (i.e. employee performance). The prediction of an individual's performance for selection into an organization is as much at the forefront of industrial and organizational psychology today as it was one hundred years ago (Austin & Villanova, 1992; Schmidt & Hunter, 1998) However, understanding an individual's job performance is not only used for selection purposes. Performance measures can also be used for numerous other purposes. For example, these measures are frequently used as a basis of promotion and placement, as well, they have been used to allocate rewards or as a criterion for training programs (Kane & Lawler, 1978). Appraising an individual's performance has been viewed as a function of three interacting systems: the organizational setting within which the evaluation occurs, the rater's ability to process information, and the ratee's behavioral patterns (Ilgen & Feldman, 1983). Facets of each system can contribute to inaccurate and biased evaluations which have posed difficulties to both researchers and practitioners (Borman, White, Pulakos & Oppler, 1991). The challenge for both researchers and practitioners is to develop instruments that minimizes inaccuracies and biases. One way to improve these instruments is to develop well tested conceptual models which can then be used to provide a framework for the selection and organization of both predictors and criteria. This thesis did just that, by examining an existing three construct conceptual model of performance and evaluating rater accuracy when rating performance, using this model.

This thesis will examine several issues. First, it will define job performance and address a two factor, three factor and eight factor model of performance. It will also present relevant research findings that support each of these models. Second, it will discuss the Mixed Standard Rating Scale (MSRS) format. Included in this discussion will be a detailed description of the MSRS format and how it is used to measure performance. Relevant research findings will also be included to show the usefulness of this format. Finally, this thesis will discuss other performance ratings and performance rankings. It will explain how both ratings and rankings are used to measure an individual's job performance.

Performance Models

Performance can generally be defined as the actions people take which are relevant to an organization's goals. These goals are value judgments determined by individuals empowered to make such a decision. These individuals are usually experts who determine whether particular actions or behaviors are relevant for a particular goal (Campbell, 1990). Historically, performance has been viewed as a single construct (i.e. "Overall" Performance), but researchers now agree that job performance is a complicated multidimensional factor (Austin & Villanova, 1992; Campbell, 1990; Dunnette, 1963; Grant, 1996; Guion, 1991; Thayer, 1977; Thorndike, 1949; Wilson, 1983). Several salient models that express this concept include a two factor model that separates performance into a task dimension and a contextual dimension, Campbell's eight factor model of performance, and a three factor model of performance.

Katz & Kahn (1978) describe job performance as a two dimensional model made up of organizational behaviors that are prescribed to the actual task and those behaviors which are considered discretionary. Several other researchers have also separated an additional dimension from the core task requirements of the job, including Brief and Motowidlo's (1986) prosocial organizational behavior, Organ's (1988) organizational citizenship, and George and Brief's (1992) organizational spontaneity.

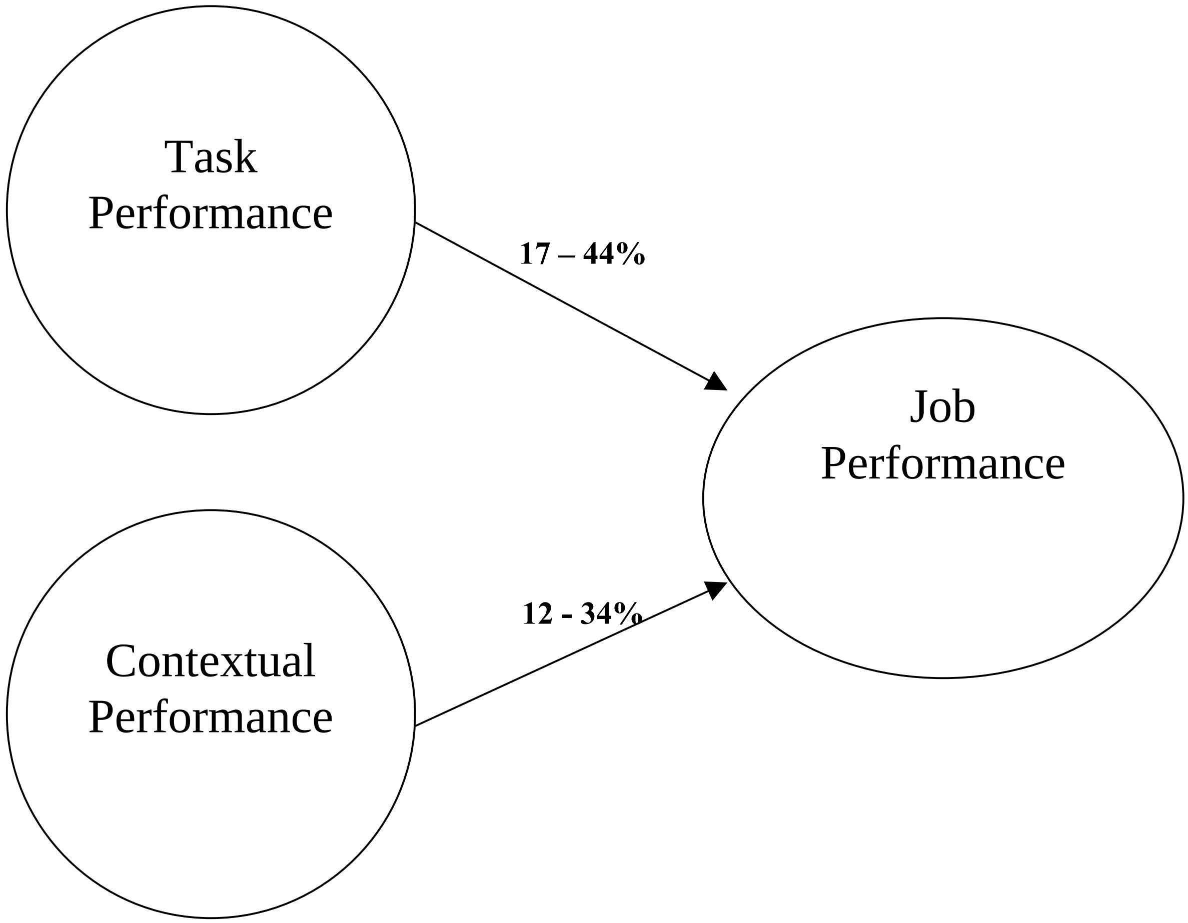

Figure 1. Motowidlo and Van Scotter's Two Factor Conceptual Model of Performance

Werner in his 1994 study separated performance into role – prescribed (in - role) characteristics and citizenship (extra - role) characteristics. He tested the extent to which experienced supervisors used both in-role and extra-role ratee characteristics when evaluating performance. The results supported the view that raters use both in - role and extra - role information when they make appraisal ratings (Werner, 1994). Motowidlo

& Van Scotter's (1994) study tested a group of U.S. Air Force mechanics on their task performance, contextual performance, and overall performance. They found that task performance explains from 17% to 44% of the variance in overall performance beyond the variance explained by contextual performance while contextual performance explains from 12% to 34% of the variance in overall performance beyond the variance explained by task performance (see Figure 1).

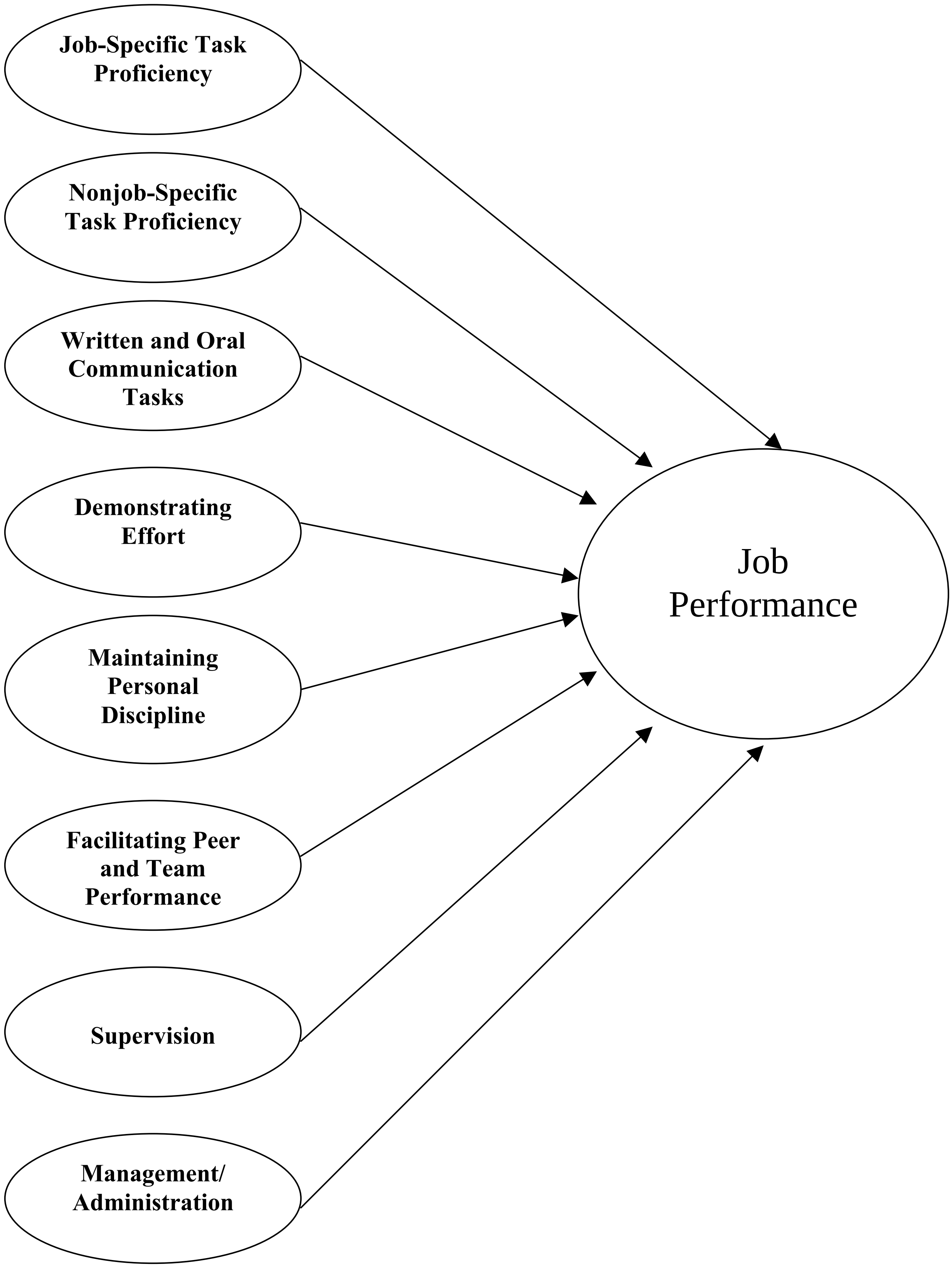

Figure 2. Campbell’s Eight Factor Model

Campbell (1990) uses a more detailed eight factor model to define job performance. The eight factors are explained by three underlying determinants. The first determinant, Declarative Knowledge, is defined as an individuals understanding of facts and information, in other words, knowing what to do. Procedural Knowledge, the second determinant, is the knowing how to do something. The final determinant is made up an individual's choice to expend effort, the choice to determine the level of effort, and the choice to persist at exerting that level of effort. These three choices are considered the motivation determinant (Campbell, 1990).

The determinants are used to explain the eight factors in Campbell's model (see Figure 2). The eight factors are Job-specific Task Proficiency, Non-job-specific Task Proficiency, Written and Oral Communication Tasks, Demonstrating Effort, Maintaining Effort, Maintaining Personal Discipline, Facilitating Peer and Team Performance, Supervision, and Management/Administration (Campbell, 1990). Campbell believes that the "eight factors are sufficient to describe the top of the latent hierarchy in all jobs in the Dictionary of Occupational Titles" (Campbell, 1990, p. 708).

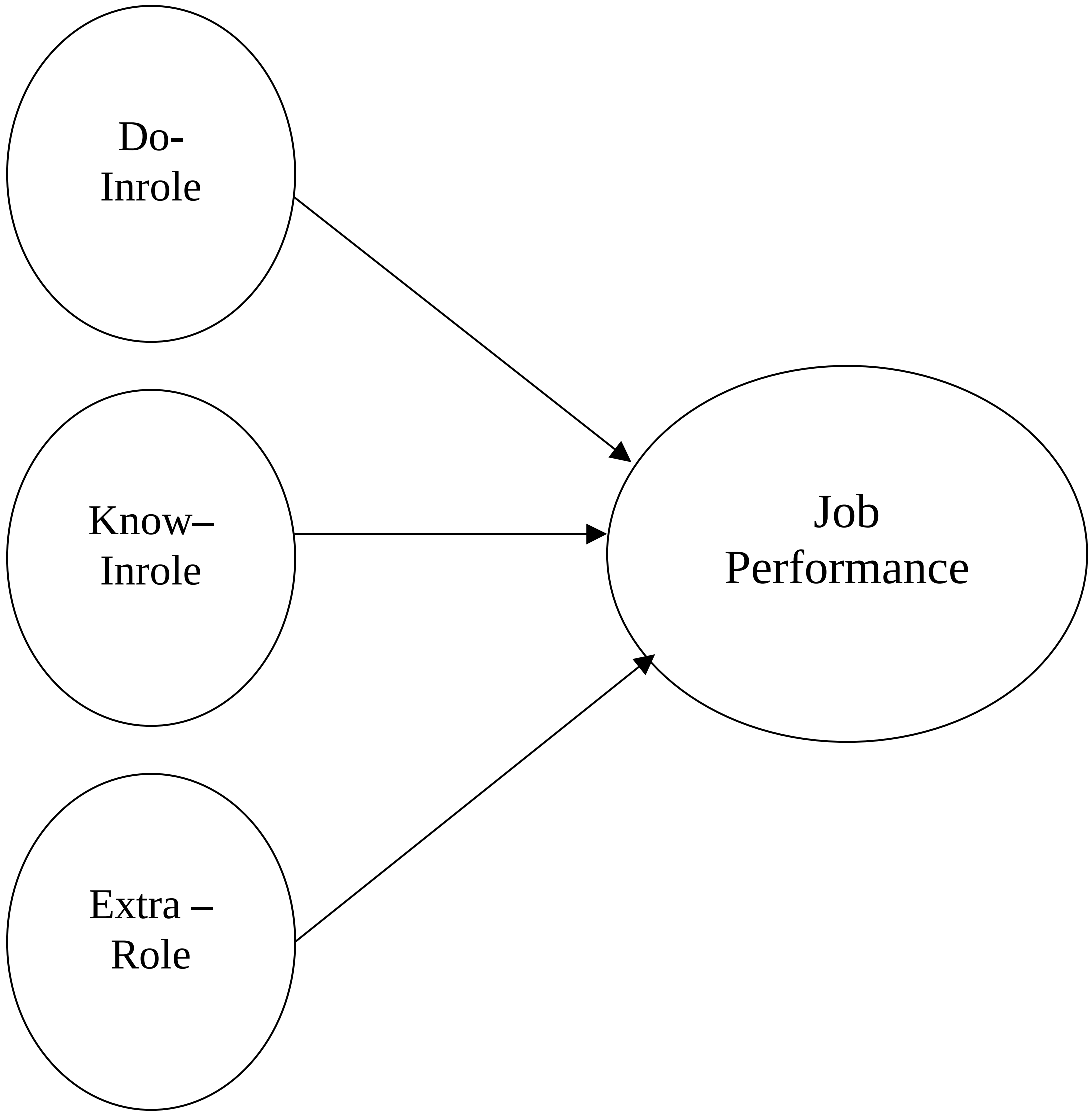

Figure 3. Grant's Three Factor Conceptual Model of Performance

Grant's (1996) study examined the latent structure of job performance using a large sample of state law enforcement officers. She was unable to explain her results with any of the existing performance models. Instead, she identified a new structural model to explain the latent structure of job performance. Her model included a Do-Inrole dimension, a Know-Inrole dimension, and an Extrarole performance dimension (see Figure 3), not unlike Campbell's three determinants of performance.

Grant suggests that the inrole factor, defined in the two factor model (Motowidlo & Van Scotter, 1994; Werner, 1994), corresponds to her Do-Inrole and Know-Inrole factors. In comparison to the eight factor model, the Do-Inrole is considered similar to Campbell's (1990) Job Specific factor, while the Know-Inrole dimension relates to the Non-Job Specific factor. The Extrarole performance dimension is similar to Motowidlo & Van Scotter's contextual performance dimension, Werner's extra-role factor, and a combination of Campbell's Maintaining Personal Discipline and Demonstrating Effort factors (Grant, 1996). Grant's three factor model of job performance also "corresponds to recent research on performance predictors, suggesting procedural knowledge, declarative knowledge, and motivation as the primary determinants of individual differences in job performance" (Grant, 1996, p. 69). Her research suggests that instead of using Campbell's eight performance dimensions for describing the latent structure of job performance her three factor model (Do-Inrole, Know-Inrole, Extrarole) should be considered.

Based on the research, it is clear that job performance can be conceptualized as multidimensional. Several studies have been able to explain job performance with two factors, while another study was able to explain job performance with three factors. It would be interesting to see if a similar three dimensional model could be replicated with a new, much larger, sample. In particular, if an instrument was designed, based on the three dimensional model (see Figure 3), and data was collected, would the results suggest a three dimensional model? Also, what is the relationship among the three constructs?

Mixed Standard Rating Scale Format

It is widely accepted that the most complete, readily available, and efficient measure of an individual's performance involves ratings of that performance by another individual (Borman, White, Pulakos & Oppler, 1991; Woehr & Miller, 1997). There have been several rating formats designed to use this technique. In particular, the MSRS is one format that offers several advantages over other formats used to measure performance. According to Blanz & Ghiselli (1972), the MSRS is reported to minimize the common rating errors of halo and leniency and also allows for several unique indices of reliability. Dickenson and Glebocki (1990) add that the MSRS is capable of identifying inconsistent raters, ratees, and dimensions. This format may assist in identifying raters who need additional training, ratees with inconsistent behaviors or who were not observed by the rater, or ambiguous items that need to be revised or replaced.

The MSRS format is designed to have the rater respond that the ratee's performance for each behavior is one of three alternatives: Better Than (+), Less Than (-), or Equal To (0). The rater's response to the three behaviors of a dimension are used to score that dimension on a 7 – point scale (Dickinson & Glebocki, 1990). Dimension ratings are based on the rater' s patterns of responses to the chosen standards for each dimension (Barnes-Farrell & Weiss, 1984).

Table 1. List of Possible Response Combinations and the Corresponding Derived Rating

Combination of Responses			
Combination			Derived Rating
I. High	II. Average	III. Low	
+	+	+	7*
+	+	0	6
+	+	-	5
+	0	+	6
+	0	0	5
+	0	-	4
+	-	+	5
+	-	0	4
+	-	-	3
0	+	+	6*
0	+	0	5
0	+	-	4
0	0	+	5
0	0	0	4
0	0	-	3
0	-	+	4
0	-	0	3
0	-	-	2
-	+	+	5*
-	+	0	4
-	+	-	3
-	0	+	4*
-	0	0	3
-	0	-	2
-	-	+	3*
-	-	0	2*
-	-	-	1*

Note 1: High = A behavior that has been scaled high on a BARS
Note 2: Average = A behavior that has been scaled average on a BARS
Note 3: Low = A behavior that has been scaled low on a BARS
Note 4: + = A 3 point rating for a soldier who “always” performs a high or average behavior or “never” performs a low behavior
Note 5: 0 = A 2 point rating for a soldier who “sometimes” performs a high, average, or low behavior
Note 6: - = A 1 point rating for a soldier who “always” performs a low behavior or “never” performs a high or average behavior
Note 7: Derived Rating = the derived rating score using the following formula:

(High behavior rating + Average behavior rating + Low behavior rating) – 2 = Derived Rating

Note 8: * = A logical pattern

For each dimension there are 27 possible combinations (response patterns) of scores, but only 7 are logically consistent (see Table 1). For example, a rater may respond 0, +, + to the high, average, and low behaviors, respectively. This would be a logical pattern for a ratee that performs at a high level. Logically, the rater's behavior would be rated better than the average and low performance behaviors and equal to the high performance behavior. An "illogical" (Benson & Dickinson 1983) response pattern would consist of a combination of responses that would describe the ratee as having an inconsistent set of behaviors. A rater responding 0, -, + to the high, average, and low behaviors, respectively, would be an example of an illogical response pattern. Here, the ratee's performance is rated as equal to the high performance yet less than the average performance which would not be a logical response pattern.

According to Dickinson and Glebocki (1990), the MSRS format reduces the tendency of raters to display leniency and halo when measuring performance. The MSRS format does not provide the rater with the performance dimensions or the effectiveness level of behavioral anchor statements. Rather than being asked to compare each ratee to a continuum of performance effectiveness for each factor, the rater is asked to compare the ratee' s performance to a series of behavioral examples representing varying levels of performance effectiveness and varying performance dimensions (Barnes-Farrell & Weiss, 1984). These behaviors are randomized which consequently disguises the dimensions from the raters. This is assumed to be the reason for the reduction in leniency and halo. Several studies that have evaluated the MSRS format have been mixed (Dickinson & Zellinger, 1980; Dickinson & Glebocki, 1990; Finley, Osborn, Dubin, & Jeanneret, 1977;

Saal, 1979; Saal & Landy, 1977). Blanz & Ghiselli (1972) state that "if there is too great a leniency error, it would suggest that the items contain descriptions of behavior that is too poor for an appropriate description of the ratees" (p. 197).

Blanz & Ghiselli (1972) proposed that the main advantage of the MSRS is that illogical raters can be identified. The number of illogical response patterns can be scored for each rater, ratee, and dimension. They proposed that the evaluations of illogical raters (raters who make a large number of illogical judgments) are lower in quality than the evaluations of logical raters. They further recommend that evaluations made by illogical raters should be discarded and illogical raters should be targeted for training.

Murphy, Garcia, Kerkar, Martin, and Balzer's (1982) study supported this assertion by finding the mean evaluation given to each ratee by illogical raters contained more error than the mean evaluation given by logical raters. Dickinson and Glebocki (1990) also supported Blanz and Ghiselli's assertion by finding that raters who have a greater number of illogical response patterns make fewer discriminations of individual performance and show more halo in their ratings. Dedrick, Dobbins, and Desselles (1990) found that evaluations of illogical raters are lower in quality than are evaluations of logical raters. Specifically, illogical raters were less accurate on differential elevation than were logical raters. Differential elevation is an important accuracy measurement for determining highest to lowest employee performance when distributing rewards and sanctions. An example of this would be promoting employees based on the highest overall mean performance (Murphy et al., 1982).

In summary, the MSRS is designed to minimize the common rating errors of halo and leniency. This is accomplished by the instrument design. The MSRS disguises the underlying dimensions which reduces the rater's ability to predict what is being measured. In deriving the MSRS dimensional scores it would be of interest to see if the neutral ratings make a significant contribution to dimensional scores. Likewise, it would be interesting to compare the MSRS derived dimensional scores to direct ratings of performance dimensions. The design of the MSRS also makes it capable to determine which raters are consistent (logical) and which raters are not consistent (illogical). Research has shown that evaluations made by logical raters are more accurate than evaluations made by illogical raters. Thus, illogical ratings are not as accurate at discriminating between the level of job performance as compared to ratings made by logical raters. Researchers that found no significant differentiation between logical and illogical raters assert that this may be a result of inaccurate descriptions of behavior for a particular job. The behavioral items used in this current study were taken directly from SF soldiers, in an earlier job analysis conducted by Russell, Crafts, Tagliareni, McCloy and Barkley (1996). It is assumed that the items should be accurate depictions of SF job behaviors. Whether or not the behavioral items used in measuring performance are accurate, it is assumed that the accuracy of the rater's responses may have an effect on the reliability of the derived dimensional scores. Accuracy is implied to mean raters with more logical response patterns will derive a score that correlates higher with the corresponding performance dimension's rating than raters with inconsistent response patterns. Following this train of thought it would be beneficial to compare illogical raters to logical raters to confirm previous research's conclusions that logical raters provide more accurate ratings.

Other Performance Ratings and Rankings

Performance ratings are the most common procedure for measuring performance, and in many work settings the only practical procedure. (Arvey & Murphy, 1998 ; Borman et. al, 1991) Ratings, or the use of expert judgment to assess performance, have received little respect due to their subjectivity. However, according to Campbell (1990) there is probably no such thing as an "objective" performance measure. Quantitative measures, like sales records, are subjectively weighted by some expert who determines the proportion each measure contributes to overall performance. According to Arvey & Murphy,(1998),

> We also find a trend in terms of an increased optimism regarding the use of supervisory ratings and other "subjective" appraisal instruments and formats. There is increased recognition that subjectivity does not automatically translate into rater error or bias and that ratings are most likely valid reflections of true performance and represent a low-cost mechanism (p. 163)

Not only are ratings considered valid reflections of the true performance but also their content can be directly linked to the measurement objectives by straightforward content validation methods. As well, their reliabilities which are between .55 and .65 are considered acceptable (Dickinson, Hassett & Tannenbaum, 1986). Similarly, Viswasvaran, Ones & Schmidt (1996) reported the results of a meta-analysis of the interrater and intrarater reliabilities of job performance ratings. Their findings indicated that supervisory ratings appear to have higher interrater reliability than peer ratings. They also found that intrarater estimates of reliability tend to be higher than interrater estimates.

Using more than one rater can increase the reliabilities of composite ratings. Using multiple raters also increases the predictability of performance ratings (Nathan &

Alexander, 1988; Schmidt, Gooding, Noe & Kirsch, 1984). The method of using two or more raters to assess multiple dimensions of performance is known as the multitrait-multimethod (MTMM) process (Kavanagh, MacKinney, & Wolins, 1971). This approach requires the use of two or more methods, raters, for evaluating two or more traits of work performance. The traits are the performance dimensions on which the ratees are rated.

A considerable amount of performance rating research has focused on the psychometric properties of the ratings (e.g. halo, central tendency, leniency) as an indicator of quality (Landy & Farr, 1980). Campbell (1990) states that, "the principal worry about ratings has always been their possible contamination with systematic variance unrelated to the performance of the person being assessed" (p. 714). Murphy, Jako, & Anhalt (1993) reviewed the definitions of rater halo error and noted that the major conceptions about this construct are either wrong or problematic. Their review of the evidence is that halo error is not pervasive, that inflated correlations among rating dimensions are not the norm, and that there are a number of contextual factors that influence when halo might be observed.

Borman, White & Dorsey's (1995) study, using a sample of U.S. soldiers, found that peers and supervisors differed slightly with regard to the different sources of information they use in providing performance ratings. Similarly, Chester Judy (1953) compared the supervisory rankings and peer rankings of 160 air force mechanics. The results indicated that supervisory rankings were significantly superior to peer rankings in their interrater reliability, freedom from bias, and practicality. As well, it was found that supervisory rankings were also more valid then peer ranking but not at a significant level.

It was also determined that there was no difference in reliability between supervisory and peer rankings when the group was comprised of more than 8 peers. Not surprisingly, supervisor's, as raters, tend to be the most widely used measure of job performance. Lent, Aurback, & Levin (1971) found that ratings were used as criterion in 948 out of 1506 validation studies and 879 were supervisors' evaluations.

With regards to rankings, in particular peer rankings, there has been very little research dealing with reliability and validity (Love, 1981). Several studies have shown that peer rankings tend to correlate higher with supervisory rankings than with ratings made by similar sources (Borg & Hamilton, 1956; Lawler, 1967; Tucker, Cline & Schmidt, 1967). This may indicate the rankings add unique variance to the evaluation process. Research that has tested the reliability of peer and supervisory rankings indicate that both provide acceptable levels of reliability (Love, 1981; Campbell, Dunnette, Lawler, & Weick, 1970). One study that provides some evidence to support these assumptions is Cathie Alderks' (1997) study of some 10,000 soldiers who attended the special forces and selection course over a 5.5 year period. She found that candidates who received the best peer rankings typically completed all the necessary requirements and were selected into the Special Forces in comparison to candidates who received poor peer rankings.

In summary, performance ratings are widely used mainly because they are considered reliable and cost effective. One advantage performance ratings provide is that they can be directly linked to an organization's measurement objectives. Using MTMM will increase the reliability and predictability of performance ratings. The greatest concern that researchers face, when using performance ratings, is systematic variance unrelated to actual performance. However, several researches believe that this is not as pervasive as was once thought. Using the ratings and rankings of a large sample of Special Forces soldiers may add significant value to the performance rating research. These measures could possibly provide a clearer understanding of how job performance is evaluated. As well, by comparing the rater's accuracy in predicting performance and the level of leniency each rater displays we may be able determine the magnitude that systematic error plays into performance ratings.

In conclusion, job performance can be conceptualized as multidimensional. However, it is still not clear as to the number of dimensions that explain job performance. Therefore, researchers have developed several conceptual models to explain job performance and are using various techniques to measure performance. The most common procedure for measuring performance is through performance ratings. Performance ratings are considered reliable and cost effective. Using MTMM will increase the reliability and predictability of performance ratings. The major concern facing performance ratings is the possibility of contamination through systematic variance that is unrelated to performance. One instrument that is designed to minimize rating errors of halo and leniency is the MSRS. However, studies that have evaluated the MSRS have been mixed. Therefore, it is not clear as to how accurate the MSRS is at measuring job performance.

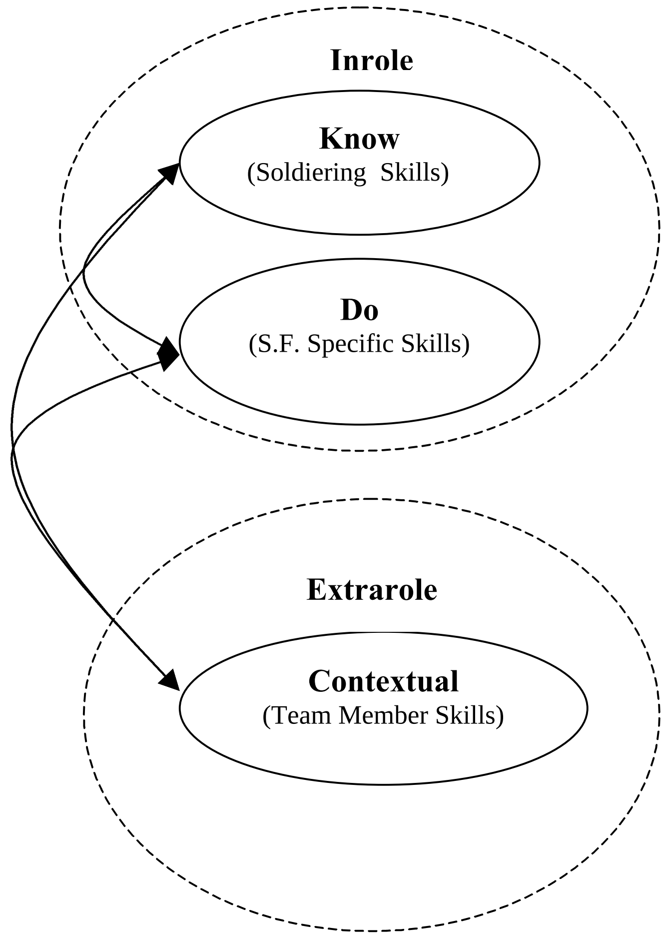

Figure 4. Correlation Between Three Factors Used in the SF Field Performance Study

Research Questions

There are four primary research questions addressed in this thesis. First, will there be significant correlation between the three dimensions used in the conceptual performance model (see Figure 4)? Hypothesis 1 is that there would be a significant correlation between each of the three conceptual constructs of job performance. As well, the highest correlation would be between the Know Inrole and Do Inrole dimensions. Even with high correlations, some of the previous research suggests that three dimensions of performance will each account for a uniquely significant proportion of variance in overall performance. Therefore, Hypothesis 2 proposes that three significantly different regression models will explain the three dimensions of performance.

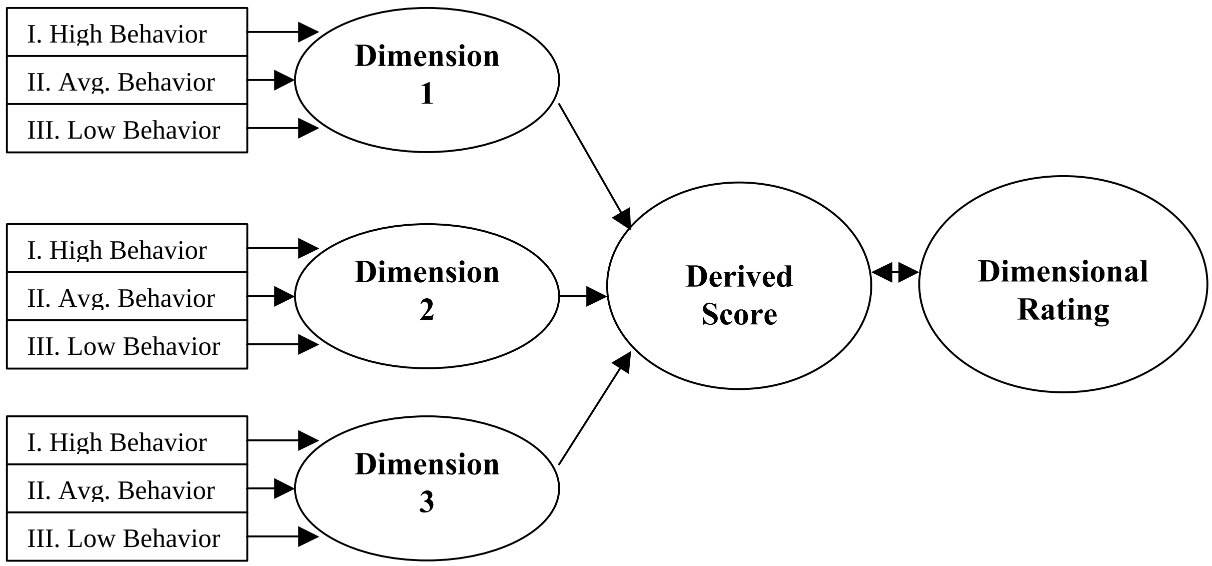

Figure 5. Study Model of Derived Score in Comparison to the Dimensional Rating

Second, will scores that are derived from a MSRS format correlate with a corresponding direct performance dimensional rating (see Figure 5)? The behaviors used on both rating forms were selected based on a three-factor model of job performance. Therefore, the derived scores should relate to their corresponding performance dimension rating. This leads to the third and fourth hypothesis. Hypothesis 3 suggests that each of the derived MSRS dimensional scores will significantly predict the corresponding dimensional rating (i.e. soldiering skills, SF specific skills, team soldiering skills). Hypothesis 4 would be supported if the three reduced models (the three sub-dimensions that theoretically predict the derived score) predict a significant proportion of variance in the full model (all nine sub-dimensions) for each dimension of performance .

Third, will raters that are more logical in their response patterns have a higher correlation between their derived MSRS scores and the corresponding performance dimension than raters who are less logical in their response patterns? Previous research indicated that logical raters have a higher correlation between derived MSRS scores and the corresponding performance dimension. Therefore, Hypothesis 5 suggests that an individual that receives a higher logical pattern score will have a higher multiple correlation coefficient, for each dimension of performance, than an individual that receives a lower logical pattern score.

Finally, what will be the effect if the neutral items (equal too critical incidents) are removed from the derived score? Neutral items tend to be ambiguous, in comparison to the low and high behaviors, which may confuse a rater. Raters who take the time to complete the form will, more than likely, make fewer mistakes with the neutral items. Therefore, the neutral items will distinguish the more accurate raters from the less accurate raters, resulting in an increase in the relationship between the derived MSRS scores and the corresponding performance dimension. Therefore, Hypothesis 6 states that by removing the neutral items (equal too statements) from the derived score the multiple correlation coefficient for each dimension of performance will be lower than the multiple correlation coefficient for each dimension when the neutral item remains in the derived score.

Method

Participants

The individuals who participated in this study all belong to an Occupational Detachment Alpha (ODA). ODA is the term given to a U.S. Army Special Force's team which includes a team leader, a team sergeant, and can have up to nine more members. The participants included 1273 U. S. Army Special Forces soldiers, assigned to an ODA, evaluated by officers. A second sample included 1121 U.S. Army Special Forces soldiers, assigned to an ODA, evaluated by noncommissioned officers. There were 740 SF soldiers who were rated by both an officer and a NCO, while 533 were rated only by an officer and 381 were rated only by a NCO.

Materials

The instruments used in this study were two rating forms that were specifically designed for the U.S. Army Special Forces (Appendix A & B). The major purpose of the rating forms were to provide field performance data to improve the selection process for Special Forces Assessment and Selection (SFAS) and Special Forces Qualification Course (SFQC).

In order to get the maximum utility from SF soldiers input, the field performance data needed to be standard, concise, and relevant. Researchers agreed that a standard form should be designed to capture company commanders, company sergeant majors, team leaders and team sergeants (SF leadership) performance input. As well, it was important to respect the soldiers time. Therefore the form was designed to be no longer than one page. Finally, the form should be created directly from the SF and approved by the SF. To achieve this, the initial form was designed using behaviors from Russell et al.'s (1996) study which identified critical incidents directly from the SF.

Rating Forms Constructs. Data provided by Russell et al.'s (1996) study was used to derive the dimensions that were applied in this study. A three factor model was decided upon after performing a Squared Multiple Correlation Promax Rotation factor analysis using SAS version 8.0 software. The three performance dimensions were Soldiering Skills, SF Specific Skills, and Team Member Skills (see technical report).

There were three sub-dimensions assigned to each of the performance dimensions. For each sub-dimension, three behaviors (items) were constructed to describe high, average, and low performance. These behaviors are stated in the form of critical incidents. All behaviors were then randomized in their order of presentation to the raters, and thus, the sub-dimensions were not explicitly identified and defined for the raters. The rating form was used to measure Special Forces (SF) soldiers performance within a framework of a three dimensional model, similar to Grant's three dimensional performance model. The three dimensions are a Know – Inrole dimension (Soldiering Skills), a Do – Inrole dimension (SF Specific Skills), and an Extra – Role performance dimension (Team Member Skills). Figure 6 lays out the three dimensional performance model that was used in this study.

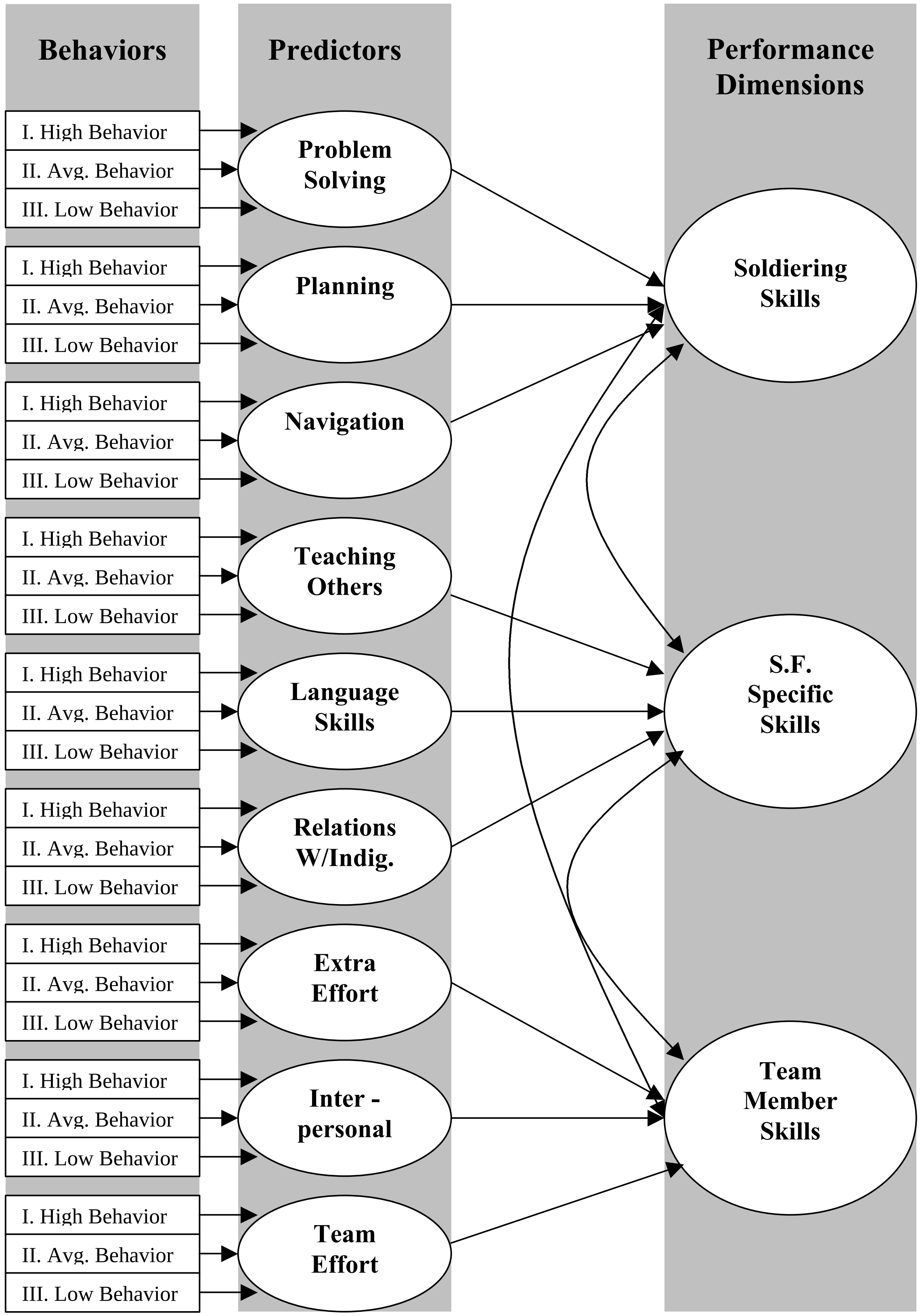

Figure 6. Three Dimensional Model Used in SF Field Performance

Two similar forms were designed using the three dimensional model described above. One form was developed for team leaders and team sergean to evaluate each of their team members. The other form was developed for company commanders and company sergeant majors to evaluate either team leaders or team sergeants that reported to them. Both forms were comprised of four sections. The following describes each of the four sections.

Section I of the rating form was designed to collect the ratee social security number, Operational Detachment Alpha (ODA), and a minimal amount of rater identification data.

Section II was a Mixed Standard Rating Scale that used 33 behaviors. Twenty seven behaviors were taken directly from Russell et al.'s (1996) study. Appendix C lists these behaviors by dimension and includes the order in which they were listed on the rating form. The behaviors for each dimension are ordered high behavior (I), average behavior (II), and low behavior (III), respectively. Six additional behaviors were included on the form, after conducting several pilot studies and receiving feedback from subject matter experts. These behaviors are listed under the miscellaneous category.

The MSRS was scored using Saal's revised scoring procedure (1979). Desselles and Dobbins (1987) demonstrated that MSRS ratings calculated with Saal's algorithm were more accurate than those calculated with Blanz and Ghiselli's method (1972) (see Table 1).

Table 2. List of Possible Response Combinations without Neutral Items and the Corresponding Derived Rating

Combination of Responses			
Combination		1st Formula Derived Rating	2nd Formula Derived Rating
I. High	III. Low		
+	+	7*	7*
+	0	6*	5.5*
0	+	5	5.5
+	-	4*	4*
0	0	4*	4*
-	+	4	4
-	0	3	2.5
0	-	2*	2.5*
-	-	1*	1*

Note 1: High = A behavior that has been scaled high on a BARS
Note 2: Low = A behavior that has been scaled low on a BARS
Note 3: 1st Formula Derived Rating = High Rating + Low Rating
Where:
+ = A 4 point rating for a soldier who "always" performs a high behavior or a 3 point rating for "never" performing a low behavior
0 = A 2 point rating for a soldier who "sometimes" performs a high or low behavior
- = A 0 point rating for a soldier who "always" performs a low behavior or a 1 point rating for "never" performing a high behavior
Note 4: 2nd Formula Derived Rating = (((High Rating + Low Rating) – 2) * (3/2) + 1)
Where:
+ = A 3 point rating for a soldier who "always" performs a high or "never" performs a low behavior
0 = A 2 point rating for a soldier who "sometimes" performs a high or low behavior
- = A 1 point rating for a soldier who "always" performs a low behavior or "never" performs a high behavior

Note 5: * = A logical pattern

Two additional scoring methods were used to determine dimensional scores without the neutral items included (see Table 2). The first method allowed a + (high) rating for the high behavior to receive a score of 4 while the 0 (neutral) was given a 2 and a – (low) got a 1. A + (high) rating for the low behavior received a score of 3 while the 0 (neutral) received a 2 and a – (low) got a 0. The two scores were then summed together to get a final derived score. A second method assigned a 3 to the high score, a 2 to the neutral score and a 1 to the low score for both the low and high behaviors. The two scores were summed together to give you a range between 2 and 6. These summed scores were then standardized.

Section III of the rating form provided a dimensional rating, using a 7-point scale, for each of the three dimensions of performance.

Section IV was a forced ranking. Team leaders and team sergeants ranked each team member on an 11 point scale. Company commanders and company sergeant majors ranked team leaders or team sergeants using a 6 point scale.

<u>Procedures</u>

Once the initial form was constructed it was then piloted on three different groups of active duty SF soldiers. As well, the form was given to several SF senior leaders for review. The feedback provided by the three pilot studies and the SF command led to the final form design.

It was important that honest input was given from soldiers in the field. This was a daunting task due to the nature of this study. It was necessary to identify each soldier being rated so that their performance in the field could be linked to their SFAS and SFQC data. This made it critical for SF leadership to understand the data provided would not be used, in any way, to single out individuals. Rather, their input would be used to evaluate the SFAS and SFQC and to provide the SF command with an overall snapshot of the force.

To tackle this issue, researchers traveled to each of the Group HQs and several agreed upon locations to describe the purpose of this study and to answer questions. It was at this time that packets were distributed to each rater and a few extra were handed to each Group's point of contact (POC). Each packet included instructions (see Appendix

D), an addressed envelope to return to the researchers, and 11 rating forms for the team leaders or team sergeants. The company commander and company sergeant major's packet included a similar rating form with six copies enclosed. The POC was regularly updated on the number of forms returned and the number of forms outstanding.

Results

The original sample was comprised of 914 SF soldiers who were rated by a single rater and 740 SF soldiers who were rated by multiple raters. In order to keep soldiers who were evaluated by multiple raters from confounding the results, because they will have more than one evaluation in the data set, the sample was divided into two new samples. The first split sample was comprised of 1273 SF soldiers rated only by officers (officer sample) while the second sample was comprised of 1121 SF soldiers rated only by noncommissioned officers (NCO sample).

Table 3. Correlations Between the Three Dimensions of Performance

Officer Sample

Dimension	Mean	Standard Deviation
Soldiering Skills	5.70487	1.24963
SF Specific Skills	5.69556	1.26820
Team Member Skills	5.55089	1.43319

	Soldiering Skills	SF Specific Skills	Team Member Skills
Soldiering Skills	1		
SF Specific Skills	0.79591	1	
Team Member Skills	0.73525	0.75532	1

(table continues)

Table 3. Correlations Between the Three Dimensions of Performance (continued)

NCO Sample

Dimension	Mean	Standard Deviation
Soldiering Skills	5.70350	1.25986
SF Specific Skills	5.61404	1.28451
Team Member Skills	5.63735	1.44553

	Soldiering Skills	SF Specific Skills	Team Member Skills
Soldiering Skills	1		
SF Specific Skills	0.8258	1	
Team Member Skills	0.79193	0.78818	1

With regards to the first hypothesis, the correlations between the three performance dimensions (soldiering skills, SF specific skills and team member skills) for each sample are presented in Table 3. In both samples the correlations are high with the highest correlation between the soldiering skills dimension and the SF specific dimension. The lowest correlation in the officer sample is between the soldiering skills dimension and the team member skills dimension. The lowest correlation in the NCO sample is between the SF specific skills dimension and the team member skills dimension.

Table 4. Multivariate Dependent Regression Analysis Statistics and F Approximations

PROC REG: (mtest)

$Y_1\ Y_2\ Y_3 = b_0 + b_1A_1 + b_2A_2 + b_3A_3 + b_4A_4 + b_5A_5 + b_6A_6 + b_7A_7 + b_8A_8 + b_9A_9$

Test Y1 = Y2
Test Y1 = Y3
Test Y2 = Y3

Officer Sample

Multivariate Dependent Variable Regression Results:

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.71370583	25.76	18	2524	<.0001
Pillai's Trace	0.30648099	25.40	18	2526	<.0001
Hotelling-Lawley Trace	0.37285297	26.12	18	2099.6	<.0001
Roy's Greatest Root	0.26686476	37.45	9	1263	<.0001

NOTE: F Statistic for Roy's Greatest Root is an upper bound.
NOTE: F Statistic for Wilks' Lambda is exact.

Y1 = Y2

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.90415758	14.88	9	1263	<.0001
Pillai's Trace	0.09584242	14.88	9	1263	<.0001
Hotelling-Lawley Trace	0.10600190	14.88	9	1263	<.0001
Roy's Greatest Root	0.10600190	14.88	9	1263	<.0001

Y1 = Y3

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.81001438	32.91	9	1263	<.0001
Pillai's Trace	0.18998562	32.91	9	1263	<.0001
Hotelling-Lawley Trace	0.23454598	32.91	9	1263	<.0001
Roy's Greatest Root	0.23454598	32.91	9	1263	<.0001

Y2 = Y3

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.80813182	33.32	9	1263	<.0001
Pillai's Trace	0.19186818	33.32	9	1263	<.0001
Hotelling-Lawley Trace	0.23742188	33.32	9	1263	<.0001
Roy's Greatest Root	0.23742188	33.32	9	1263	<.0001

(table continues)

NCO Sample

Multivariate Dependent Variable Regression Results:

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.72021440	21.99	18	2220	<.0001
Pillai's Trace	0.30138857	21.90	18	2222	<.0001
Hotelling-Lawley Trace	0.35848024	22.09	18	1846.2	<.0001
Roy's Greatest Root	0.22541185	27.83	9	1111	<.0001

NOTE: F Statistic for Roy's Greatest Root is an upper bound.
NOTE: F Statistic for Wilks' Lambda is exact.

Y1 = Y2

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.87982183	16.86	9	1111	<.0001
Pillai's Trace	0.12017817	16.86	9	1111	<.0001
Hotelling-Lawley Trace	0.13659375	16.86	9	1111	<.0001
Roy's Greatest Root	0.13659375	16.86	9	1111	<.0001

Y1 = Y3

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.81962203	27.17	9	1111	<.0001
Pillai's Trace	0.18037797	27.17	9	1111	<.0001
Hotelling-Lawley Trace	0.22007457	27.17	9	1111	<.0001
Roy's Greatest Root	0.22007457	27.17	9	1111	<.0001

Y2 = Y3

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.83943808	23.61	9	1111	<.0001
Pillai's Trace	0.16056192	23.61	9	1111	<.0001
Hotelling-Lawley Trace	0.19127310	23.61	9	1111	<.0001
Roy's Greatest Root	0.19127310	23.61	9	1111	<.0001

Note 1: Y1 = Soldiering Skills Dimension
Note 2: Y2 = SF Specific Skills Dimension
Note 3: Y3 = Team Member Skills Dimension
Note 4: A1= Problem-solving Sub-dimension
Note 5: A2 = Planning and Preparing for Mission Sub-dimension
Note 6: A3 = Navigation Sub-dimension
Note 7: A4 = Teaching Others Sub-dimension
Note 8: A5 = Language Skills Sub-dimension
Note 9: A6 = Relationships with Indigenous Populations Sub-dimension
Note 10: A7 = Extra Effort Sub-dimension
Note 11: A8 = Interpersonal Sub-dimension
Note 12: A9 = Team Effort Sub-dimension

With regards to the second hypothesis, multivariate dependent variable regression analysis was performed using SAS v.8 software. The results in Table 4 show that there is a significant difference between the three dimensions of performance using the full model (all nine sub-dimensions), for both samples. A between model comparison was also carried out to get a more detailed understanding of where these differences may be. Again, the results in Table 4 show that, for both samples, there are significant F-probabilities at the .0001 level for all three dimensions. These results show that the three dimensions of performance are significantly different from each other, which indicates that each dimension is using a significantly different regression model.

Table 5. Full Model Parameter Estimates and t-values

	Officer Sample			NCO Sample		
	y1	y2	y3	y1	y2	y3
a1	- .01044 (.7135)	- .03369 (.2327)	.01779 (.5549)	.03739 (.2118)	.05466 (.0821)	.08729 (.0079)
a2	.27034 (.0001)	.19508 (.0001)	.14821 (.0001)	.29293 (.0001)	.12383 (.0004)	.06829 (.0622)
a3	.26860 (.0001)	.10242 (.0026)	.04039 (.2646)	.19579 (.0001)	.07343 (.0396)	.05521 (.1381)
a4	.18434 (.0001)	.22183 (.0001)	.10691 (.0004)	.17653 (.0001)	.20601 (.0001)	.14518 (.0001)
a5	.04233 (.0734)	.17007 (.0001)	.05446 (.0298)	.03702 (.1388)	.15449 (.0001)	.02033 (.4581)
a6	.00739 (.8078)	.10401 (.0006)	- .00254 (.9371)	.02079 (.4703)	.13161 (.0001)	.05690 (.0715)
a7	.15382 (.0001)	.10710 (.0010)	.28334 (.0001)	.16750 (.0001)	.13591 (.0001)	.30331 (.0001)
a8	.01349 (.5395)	.01582 (.4687)	.25825 (.0001)	- .00822 (.7089)	.02701 (.2426)	.17890 (.0001)
a9	.09971 (.0002)	.13571 (.0001)	.27007 (.0001)	.11751 (.0001)	.12321 (.0001)	.23796 (.0001)

Note 1: y1 = Soldiering Skills Dimension
Note 2: y2 = SF Specific Skills Dimension
Note 3: y3 = Team Member Skills Dimension
Note 4: a1= Problem-solving Sub-dimension
Note 5: a2 = Planning and Preparing for Mission Sub-dimension
Note 6: a3 = Navigation Sub-dimension
Note 7: a4 = Teaching Others Sub-dimension
Note 8: a5 = Language Skills Sub-dimension
Note 9: a6 = Relationships with Indigenous Populations Sub-dimension
Note 10: a7 = Extra Effort Sub-dimension
Note 11: a8 = Interpersonal Sub-dimension
Note 12: a9 = Team Effort Sub-dimension

Multivariate regression analysis was performed on each of the three dimensions of performance, for both samples, to get the parameter estimates for each of the significantly different regression models (see Table 5). When regressing the full model, on each dimension of performance, the parameter estimates for the three sub-dimensions that should theoretically predict the performance dimension (reduced model) are significant for two of the three dimensions. The one exception is the problem-solving sub-dimension (a1) that does not have a significant parameter estimate for the soldiering skills performance dimension for both samples. The results in Table 5 also show that the highest parameter estimate for each dimension, in both samples, is the same sub-dimension. The planning and preparing for missions sub-dimension (a2), has the highest parameter estimate for predicting the soldiering skills dimension. The teaching others sub-dimension (a4) has the highest parameter estimate for predicting the SF specific dimension. The extra effort sub-dimension (a7) has the highest parameter estimate for predicting the team member dimension.

Table 6. Regression Models for the Three Dimensions of Performance

Regression Equation	Reduced Model R^2 and F-prob.	Full Model R^2 and F-prob.	% Full explained by Reduced	Independent Variable Standardized Estimate(t-prob)								
				a1	a2	a3	a4	a5	a6	a7	a8	a9
Officer Sample												
y1=a1a2a3	0.4801 (.0001)	0.5408 (.0001)	88.78%	.0911 (.0005)	.40007 (.0001)	.30955 (.0001)	-	-	-	-	-	-
y2=a4a5a6	0.4891 (.0001)	0.5617 (.0001)	87.07%	-	-	-	.38367 (.0001)	.21105 (.0001)	.21773 (.0001)	-	-	-
y3=a7a8a9	0.5829 (.0001)	0.6096 (.0001)	95.62%	-	-	-	-	-	-	0.30977 (.0001)	0.27391 (.0001)	0.33565 (.0001)
NCO Sample												
y1=a1a2a3	0.5127 (.0001)	0.5821 (.0001)	88.08%	.17426 (.0001)	.40049 (.0001)	.26018 (.0001)	-	-	-	-	-	-
y2=a4a5a6	0.4792 (.0001)	0.5538 (.0001)	86.53%	-	-	-	.39675 (.0001)	.18224 (.0001)	.24042 (.0001)	-	-	-
y3=a7a8a9	0.5907 (.0001)	0.6191 (.0001)	95.41%	-	-	-	-	-	-	.36056 (.0001)	.21646 (.0001)	.32921 (.0001)

Note 1: y1 = Soldiering Skills Dimension
Note 2: y2 = SF Specific Skills Dimension
Note 3: y3 = Team Member Skills Dimension
Note 4: a1= Problem-solving Sub-dimension
Note 5: a2 = Planning and Preparing for Mission Sub-dimension
Note 6: a3 = Navigation Sub-dimension
Note 7: a4 = Teaching Others Sub-dimension
Note 8: a5 = Language Skills Sub-dimension
Note 9: a6 = Relationships with Indigenous Populations Sub-dimension
Note 10: a7 = Extra Effort Sub-dimension
Note 11: a8 = Interpersonal Sub-dimension
Note 12: a9 = Team Effort Sub-dimension

With regards to the third hypothesis, multiple regression analysis was performed using the reduced models for each dimension of performance. Table 6 shows the parameter estimates and t – probabilities for each regression model in both samples. Also included in Table 6 are the R^2s and F- probabilities for both the reduced models and the full model. The three regression models, in both samples, have significant R^2s and parameter estimates. The problem-solving sub-dimension (a1) in the reduced model is significant at the .0005 level for the officer sample and the .0001 level for the NCO sample, when regressed on the soldiering skills dimension.

For the officer sample, the reduced model is able to explain 48.01% of the variance in the soldiering skills dimension, which is 88.78% of what the full model accounts for in this dimension. The reduced models account for 48.91% of the variance in the SF specific skills dimension and 58.29% in the team member skills dimension, which is 87.07% and 95.62% of what the full model is able to explain in these dimensions, respectively.

For the NCO sample, the reduced model is able to explain 51.27% of the variance in the soldiering skills dimension, which is 88.08% of what the full model accounts for in this dimension. The reduced models account for 47.92% of the variance in the SF specific skills dimension and 59.07% in the team member skills dimension, which is 86.53% and 95.41% of what the full model is able to explain in these dimensions, respectively.

Table 7. Testing the Reduced Model to the Full Model

```
model y1 = a1 a2 a3 a4 a5 a6 a7 a8 a9;
test a4=0, a5=0, a6=0, a7=0, a8=0, a9=0;

Officer Sample:                                            Mean
                   Source              DF          Square    F Value    Pr > F

                   Numerator            6        19.65752      27.81    <.0001
                   Denominator       1263         0.70691

NCO Sample:                                                 Mean
                   Source              DF          Square    F Value    Pr > F

                   Numerator            6        20.83174      30.78    <.0001
                   Denominator       1111         0.67687

model y2 = a4 a5 a6 a1 a2 a3 a7 a8 a9;
test a1=0, a2=0, a3=0, a7=0, a8=0, a9=0;

Officer Sample:                                            Mean
                   Source              DF          Square    F Value    Pr > F

                   Numerator            6        24.29401      34.88    <.0001
                   Denominator       1263         0.69657

NCO Sample:                                                Mean
                   Source              DF          Square    F Value    Pr > F

                   Numerator            6        23.09251      30.97    <.0001
                   Denominator       1111         0.74557

model y3 = a1 a2 a3 a4 a5 a6 a7 a8 a9;
test a1=0, a2=0, a3=0, a4=0, a5=0, a6=0;

Officer Sample:                                            Mean
                   Source              DF          Square    F Value    Pr > F

                   Numerator            6        11.38270      14.34    <.0001
                   Denominator       1263         0.79353

NCO Sample:                                                Mean
                   Source              DF          Square    F Value    Pr > F

                   Numerator            6        11.21533      13.80    <.0001
                   Denominator       1111         0.81280
```

With regards to the fourth hypothesis, the test procedure in multivariate regression analysis, using SAS v.8 software, was performed to determine if the reduced models were significantly different from the full model (see Table 7). The results show that the three reduced models, for both samples, are significantly different from the full model at the

.0001 level. This would indicate that, even though the reduced models explain a high percentage of the variance explained by the full model in each dimension, there is still a significant difference between the two models. Therefore, quantitatively speaking, the full model is explaining significantly more variance in each dimension when compared to the reduced model.

Table 8. Logical Pattern Results for the Three Dimensions of Performance in the Officer Sample

	Officer Sample		**NCO Sample**	
	Sample Size	Adjusted R^2	Sample Size	Adjusted R^2
Predicting Soldiering Skills				
0-5 Logical Patterns	96	0.5782	117	.4523
6 Logical Patterns	144	0.4926	117	.5951
7 Logical Patterns	265	0.5116	256	.4591
8 Logical Patterns	395	0.4470	347	.5653
9 Logical Patterns	373	0.5169	284	.5938
Predicting SF Specific Skills				
0-5 Logical Patterns	96	0.6012	117	.4420
6 Logical Patterns	144	0.6066	117	.5888
7 Logical Patterns	265	0.5446	256	.3715
8 Logical Patterns	395	0.4607	347	.5306
9 Logical Patterns	373	0.5182	284	.6126
Predicting Team Member Skills				
0-5 Logical Patterns	96	0.6418	117	.6043
6 Logical Patterns	144	0.5607	117	.6840
7 Logical Patterns	265	0.5182	256	.4510
8 Logical Patterns	395	0.5200	347	.5433
9 Logical Patterns	373	0.5674	284	.6057

To answer the fifth hypothesis, both samples were further divided into smaller samples based on their logical pattern scores. Table 8 shows the sample size and adjusted

R^2 for each logical pattern within each dimension. The lowest five logical patterns were combined (for both the officer and NCO sample) to increase the sample size. With regards to the officer sample, the highest adjusted R^2s for the soldiering skills dimension and team member skills dimension are with the lowest logic pattern samples. In other words, the soldiers who receive the lowest logical patterns have the highest adjusted R^2s for the soldiering skills dimension and team member skills dimension. The sample with six correct logical patterns has the highest adjusted R^2 for the SF specific dimension. With regards to the NCO sample, the highest adjusted R^2s for the soldiering skills dimension and team member skills dimension are with the samples that have six logical pattern scores. The sample with all nine logical patterns correct has the highest adjusted R^2 for the SF specific skills dimension.

Table 9. Correlation Coefficients for Predicting Each Dimension using both the Three Behavior and Two Behavior Derived Rating

	Officer Sample Adjusted R^2			NCO Sample Adjusted R^2		
	3 Items	2 Items*	2 Items**	3 Items	2 Items*	3 Items**
Soldiering Skills	.5376	.5212	.5176	.5787	.5701	.5763
SF Specific Skills	.5586	.5522	.5463	.5502	.5512	.5513
Team Member Skills	.6068	.5895	.5891	.6160	.5967	.6058

Note 1: 2 Item* = Derived Rating = High Rating + Low Rating
Note 2: 2 Item ** = Derived Rating = (((High Rating + Low Rating) – 2) * (3/2) + 1)

With regards to the sixth hypothesis, the adjusted R^2s are lower for the three performance dimensions, in both samples, when the neutral items are removed from the sample (see Table 7). However, the differences between the adjusted R^2s are relatively small for each dimension.

Post Hoc Results

Although not part of the original study, having two samples allows for comparison of the relative order of the results between the samples. These comparisons, though only descriptive in nature, may provide a different perspective to the data, which could help to explain the quantitative results. As well, these comparisons may give rise to additional research questions.

With regards to results shown in table 6, only the a1 sub-dimension (trouble shooting and solving problems) for the officer sample has a t -probability at the .0005 level; all other sub-dimensions are at the .0001 level. Both samples have the a2 sub-dimension (planning and preparing for missions) as the highest parameter estimate for predicting soldiering skills. As well, both samples have the a4 sub-dimension (teaching others) as the highest parameter estimate for predicting SF specific skills. These results are similar to the full model regression results shown in Table 5. The highest parameter estimate, in the officer sample, for predicting team member skills is the a9 sub-dimension (contributing to team effort and moral), while the NCO sample's highest parameter estimate is the a7 sub-dimension (showing initiative and extra effort). The sub-dimensions in the reduced model explain the most variance in the team member skills dimension for both samples. However, in the officer sample, the sub-dimensions explain the least amount of variance in soldiering skills dimension while, in the NCO sample, the sub-dimensions explain the least amount of variance in the SF specific skills dimension.

With regards to the results shown in Table 9, the highest adjusted R^2, in the officer sample, is with the team member skills dimension, followed by SF specific skills dimension, and finally the soldiering skills dimension. In the NCO sample, the highest adjusted R^2 is with the team member skills dimension, followed by the soldiering skills dimension, and finally the SF specific dimension. With both samples, there is no change in the relative order when the neutral items are removed from the derived score.

Discussion

As expected, the intercorrelations between the three dimensions are high for both samples. Though this indicates that the three dimensions share considerable variance and are not independent of each other it is difficult to determine why the intercorrelations are so high. One possible explanation for the high correlations is that SF soldiers who demonstrate one dimension, such as a high score on soldiering skills, may also demonstrate the same level of competence for the other two dimensions, a high score on team member skills and SF specific skills. Another possibility is that raters find it hard to distinguish between the three dimensions. For example, a soldier may have high SF specific skills and soldiering skills but may not be at a high level in team member skills. The rater may evaluate such a SF soldier low on both SF specific and soldiering skills because he is unable to see past the lack of team member skills. Another example would be a SF soldier who puts out the extra effort and helps team members but lacks both soldiering and SF specific skills. A rater may find it hard to attribute low scores to the SF specific and soldiering dimensions for a soldier who works hard at being a good team member. When visually comparing the two samples, the NCO sample appears to have higher correlations on all three dimensions. This may suggest that the NCOs were less effective at distinguishing between the three dimensions of performance when compared to the officer sample. Future research should look to see if the differences in correlations, between samples, are significant. If significant differences are found between the two samples, it may be due to the rater's frame of reference. Officers may have a different perspective as to what is considered high performance because of the position they hold on the team compared to the position and responsibilities held by the NCO. Another possibility to potential differences in raters may be attributed to the time a rater has spent on a team. Typically the team sergeant (NCO) has spent more time on an ODA than the team leader (officer). Raters with more time on a team may have a clearer understanding of what is considered good performance, as well as, more opportunities to observe a SF soldiers performance. However, the latter should not be the case, in most circumstances, since the nature of the job requires both raters to spend considerable amounts of time with each team member. There should be ample opportunities for both raters to observe each team member's performance.

Though the differences between the correlations, within each sample, are small the relative order from highest to lowest may offer some insight. As predicted the correlation between the SF specific skills and the soldiering skills is the highest for both samples. This would indicate that SF soldiers who have high declarative knowledge (know what to do) also have high procedural knowledge (know how to do it) and SF soldiers who lack declarative knowledge also lack procedural knowledge. The soldiering skills dimension and team member skills dimension have the lowest intercorrelation in the officer sample. This may suggest that officers attribute contextual performance more to what a SF soldier does than what he knows. The lowest intercorrelation, in the NCO sample, is between SF specific skills and team member skills, which suggests that NCOs attribute contextual performance more to what a SF soldier knows than to what he does. Future research should look to see if the raters position on the team (officer or NCO) is a moderator to performance ratings. The difference in the relative order of the intercorrelations between the two samples suggests that there may be.

The results of the between model comparison indicate that even though the three dimensions are highly correlated, SF soldiers are being evaluated on significantly different regression lines for each of the three dimensions of performance. Future research should factor analyze the current data to see if a three dimensional model, similar to the one used in this study, can be replicated. Using the data in this study, it may be found that two, four, or even more than four dimensions are more adequate at explaining SF soldiers performance. Though still inconclusive, the between model comparison results support the three dimensional performance model by indicating that SF soldiers are being evaluated significantly different in their declarative knowledge, procedural knowledge, and contextual performance.

The results section indicate that troubleshooting and problem solving, planning and preparing for missions, and navigating in the field are all significant predictors of soldiering skills. However, the three sub-dimensions, in the reduced model, are not able to significantly account for the variance in the soldiering skills dimensions when compared to all nine sub-dimensions (the full model). The same can be said for the SF specific skills and team member skills dimensions for both samples. Teaching others, using and enhancing language skills, and building effective relationships with indigenous populations are all significant predictors of the SF specific skills dimension. Likewise, showing initiative and extra effort, handling interpersonal relationships, and contributing to the team effort and morale are significant predictors of team member skills. However, the three sub-dimensions together are not able to explain a significant proportion of variance in the corresponding dimension when compared to the full model. However, each of the reduced models explains a considerable proportion of variance in each dimension when compared to the full model. This would suggest, from a practical standpoint, that the reduced models are a cost effective way to account for most of the variance in each dimension that would be accounted for by full model. However, there may be a better fitting reduced model. It would be useful to know if the three theoretical sub-dimensions, for each dimension, are the best three predictors for that dimension. As well, future studies, using these samples, should determine not only the best predictors but, as well, the least number of predictors that can be used to significantly account for all nine sub-dimensions. It would be beneficial to see if the two samples produce similar predictive models for each dimension.

The relative positions of the standardized estimates, in the reduced models, offer another perspective when compared across samples. The relative orders of the standardized estimates, for both samples, are the same for soldiering skills and SF specific skills. This would suggest that a SF soldiers ability to trouble shoot and solve problems, plan and prepare for missions, and navigate in the field are all significant predictors in determining his soldiering performance. As well, teaching others, using and enhancing language skills, and building effective relationships with indigenous populations are all significant at predicting a soldiers SF specific performance. Not only are these behaviors significant in predicting performance but, as well, both samples agree on the relative position each behavior has in predicting performance when compared to the other behaviors for the specific dimension. This is not the case for the team member skills dimension. The relative positions of the standardized estimates for the team member skills model are different in each sample. Though showing initiative and extra effort, handling interpersonal situations, and contributing to team effort and morale are all significant in predicting the team member skills dimension the two samples differ in the relative orders of the weights each behavior contributes to this dimension. The officer sample has the highest standardized estimate on the contributing to team effort and morale behavior while the NCO sample has the highest standardized estimate on showing initiative and extra effort. Though it is not for certain that the difference between the two samples are significant, the difference in the behaviors for the team member skills dimension's relative order may suggest, once again, that the raters view performance differently. Future research should compare the performance ratings made by different raters to see if there is a significant difference in the way raters evaluate team member skills performance.

Based on the results of this study, it is difficult to draw any concise conclusion as to why the team member skills model carried the highest multiple correlation coefficient for both samples. It may be that a SF soldier's team member behaviors are easier to evaluate in comparison to SF specific behaviors and soldiering behaviors. Another possibility is that the behaviors used to describe the team member dimension carried a more accurate depiction of actual performance in comparison to the other two dimensions' behaviors, which made it easier to evaluate. The lowest multiple correlation coefficient for the officer sample is with the soldiering skills dimension while the lowest multiple correlation coefficient, in the NCO sample, is with the SF specific dimension. These results may add more insight to the previous assertion made concerning the intercorrelations between dimensions. It may be that officers are attributing contextual performance more to what a soldier does (SF specific skills) than what he knows (soldiering skills). The low multiple correlation coefficient for the soldiering skills dimension, in comparison to the team member and SF specific dimensions, may suggest something else. It may be that, in the officer sample, a ratee's soldiering skills behavior is more difficult to evaluate. Future research should look to see if officers are having a difficult time distinguishing high and low soldiering skills behavior or if the ratee's soldiering skills behavior is genuinely a more difficult construct to evaluate. The opposite can be said of the NCO sample. One assumption made in the low intercorrelation between the team member skills and SF specific skills dimension is that NCOs may attribute contextual performance more to what a soldier knows (soldiering skills) than what he does (SF specific skills). However, it may be that NCOs are having a difficult time determining what is considered high and low SF specific skills behavior. Future research should look to see if this is a valid assumption.

Removing the neutral items from the regression equation reduced the adjusted multiple correlation coefficients for all three dimensions in both the NCO sample and the officer sample. As well, this procedure had no effect in changing the relative orders of the adjusted multiple correlation coefficients. This may suggest that the neutral items improve the derived scores overall predictability of each dimension. However, the differences between the adjusted multiple correlation coefficients are very small. This could also suggest that removing the neutral items would only minimally reduce the amount of variance explained. It may be more practical to not include the neutral items in the form to shorten the time needed to complete the evaluation. Reducing the amount of time needed to complete the form may increase the R^2 by reducing rater fatigue. Future research should look to see if removing the neutral items, all together from the form, would increase the correlation coefficient when compared to another sample using the exact same form with the exception that neutral items are included.

The logical pattern results make it difficult to draw any concise conclusions. The highest adjusted multiple correlation coefficients are with SF soldiers who received five or six logical patterns, except for the SF specific dimension in the NCO sample. Conversely, the lowest adjusted multiple correlation coefficients were SF soldiers who received seven or eight logical patterns correctly except for the soldiering skills dimensions, in the NCO sample. These results are difficult to explain and open the door to many questions. It would seem that the more logical a SF soldier is evaluated the more variance would be explained in the model. This, however, may be the problem. One possibility is that a number of raters did not differentiate at the behavioral level, which is where the sub-dimensions were derived. Instead, these raters may have evaluated all SF soldiers exactly the same. Whereas, at the dimensional level, the raters differentiated their ratings. Having only three judgments to make on performance (soldiering, SF specific, and team member skills) made it easier for the officers to compare between soldiers. On the other hand, officers who took the time to evaluate each behavior made more mistakes but differentiated behaviors enough to more fully explain the performance dimension.

Future research should explore possible explanations for these results. Some research possibilities include removing influential outliers from the sample. If there are several influential outliers with low logical patterns then this may skew the results in favor of lower logical patterns. As well, future research should look at rater patterns. It may be useful to remove raters who did not differentiate behavioral ratings between the SF soldiers they rated. If, as expected, a majority of these raters were perfectly logical in their behavioral ratings then removing them may potentially increase the adjusted multiple correlation coefficient.

Future research could also look at removing items that were consistently missed by a majority of raters in each of the samples. Comparing the missed items between samples may also add some insight to the reason for such disparity in the results. It would be interesting to see if the most frequently missed items in the NCO sample were similar to the most frequently missed items in the officer sample. Another avenue for research to pursue is removing the neutral items from the sample and comparing the newly constructed logical patterns to see if SF soldiers with high logical scores are able to explain more of the variance in the model in comparison to SF soldiers with lower logical scores. It would also be beneficial to determine which sub-dimensions produced the highest number of illogical scores. By identifying these sub-dimensions and removing their frequently missed items then rerunning a new regression analysis using a combination of sub-dimensions that have all the items included and some that have a frequently missed item removed may produce higher adjusted multiple correlation coefficients for each performance model.

In conclusion this study seems to deliver more questions than answers, however, what was discovered seems to be a good starting place for future research. The three dimensions of performance are highly intercorrelated with the highest correlation between the soldiering skills and SF specific dimensions. However, even with the high intercorrelations, significantly different regression equations are used to predict each dimensional model. As well, the three sub-dimensions that theoretically make up each dimension are all significant in explaining their corresponding dimension. Nonetheless, the three sub-dimensions are not able to explain a significant amount of variance in comparison to the full model. The logical pattern findings are inconclusive. The highest adjusted multiple correlation coefficients are SF soldiers with lower logical pattern scores, in the officer sample, while the highest adjusted multiple correlation coefficients were mixed, in the NCO sample. Finally, including the neutral items in the sample increase the multiple correlation coefficients for each dimensional model, but only marginally.

References

Alderks, C. E. (1997). Special Forces: Assessment and Selection (SFAS) Course: Similarities and Differences of Candidates Based on Phase Performance.: Army Research Institute.

Arvey, R. D., & Murphy, K. R. (1998). Performance evaluation in work settings. Annual Review of Psychology, 49, 141-168.

Austin, J. T., & Villanova, P. (1992). The criterion problem: 1917-1992. Journal of Applied Psychology, 77(6), 836-874.

Barnes Farrell, J. L., & Weiss, H. M. (1984). Effects of standard extremity on Mixed Standard Scale performance ratings. Personnel Psychology, 37(2), 301-316.

Benson, P. G., & Dickinson, T. L. (1983). Mixed standard scale response inconsistencies as reliability indices. Educational and Psychological Measurement, 43(3), 781-789.

Blanz, F., & Ghiselli, E. E. (1972). The mixed standard scale: A new rating system. Personnel Psychology, 25(2), 185-199.

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

Borman, W. C., White, L. A., & Dorsey, D. W. (1995). Effects of Ratee Task Performance and Interpersonal Factors on Supervisor and Peer Performance Ratings. Journal of Applied Psychology, 80(1), 168-177.

Borman, W. C., White, L. A., Pulakos, E. D., & Oppler, S. H. (1991). Models of supervisory job performance ratings. Journal of Applied Psychology, 76(6), 863-872.

Brief, A. P., & Motowidlo, S. J. (1986). Prosocial Organizational Behaviors. Academy of Management Review, 11, 710-725

Campbell, J. P. (1990). Modeling the performance prediction problem in Industrial and Organizational Psychology. In M. D. D. L. M. Hough (Ed.), Handbook of Industrial and Organizational Psychology (2 ed., Vol. 1, pp. 687-732). Palo Alto, CA: Consulting Psychologist Press.

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

Dedrick, E. J., Dobbins, G. H., & Desselles, M. L. (1990). The Accuracy of Evaluations Made by Logical and Illogical Raters Using a Mixed Standard Scale Format. Educational and Psychological Measurement, 50, 411-419.

Desselles, M. L., & Dobbins, G. H. (1987). A Comparison of Mixed Standard Scale Ratings Calculated with Blanz and Ghiselli's and Saal's Scoring Algorithms. Educational and Psychological Measurement, 47(3), 799-806.

Dickinson, T. L., & Glebocki, G. G. (1990). Modifications in the Format of the Mixed Standard Scale. Organizational Behavior and Human Decision Processes, 47, 124-137.

Dickenson, T. L., Hassett, C. E., & Tannenbaum, S. I. (1986). Work Performance Ratings: A Meta-Analysis of Multitrait - Multimethod Studies. San Antonio: Texas Maxima Corp.

Dickinson, T. L., & Zellinger, P. M. (1980). A Comparison of the Behaviorally Anchored Rating and Mixed Standard Formats. Journal of Applied Psychology, 65(65), 147-154.

Dunnette, M. D. (1963). A Modified Model for Test Validation and Selection Research. Journal of Applied Psychology, 47, 317-323.

Finley, D. M., Osborn, H. G., Dubin, J. A., & Jeanneret, P. R. (1977). Behaviorally Based Rating Scale: Effects of Specific Anchors and Disguised Scale Continua. Personnel Psychology, 30, 659-669.

George, J. M., & Brief, A. P. (1992). Feeling Good - Doing Good: A Conceptual Analysis of the Mood at Work - Organizational Spontaneity Relationship. Psychological Bulletin, 112, 310-329.

Grant, L.D. (1996). A comprehensive examination of the latent structure of job performance. Doctoral Dissertation, North Carolina State University, Raleigh.

Guion, R. M. (1991). Personnel Assessment, Selection, & Placement. In D. L. Hough (Ed.), Handbook of Industrial and Organizational Psychology (2 ed., Vol. 2). Palo Alto, CA: Consulting Psychologist Press.

Ilgen, D. R., & Feldman, J. M. (1983). Performance Appraisal: A Process Approach. Research in Organizational Behavior, 5, 141-197.

Judy, C. J. (1953). A comparison of supervisory rankings with peer rankings. USAF Human Resources Research Center Research Bulletin, 53(43), 36.

Kane, J. S., & Lawler, E. E. III, (1978). Methods of Peer Assessment. Psychological Bulletin, 85, 555-586.

Katz, D., & Kahn, R. L. (1978). The Social Psychology of Organizations (2nd ed.). New York: Wiley.

Kavanagh, M.J., MacKinney, & Wolins, (1971). The content issue in performance appraisal: A review. Personnel Psychology, 24, 653-668.

Landy, F. J., & Farr, J. L. (1980). Performance Rating. Psychological Bulletin, 87(1), 72-107.

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

Lent, R. H., Aurback, H. A., & Levin, L. S. (1971). Prediction, criteria, and significant results. Personnel Psychology, 24, 519-533.

Love, K. G. (1981). Emprical Recommendations for the Use of Peer Rankings in the Evaluation of Police Officer Performance. Public Personnel Management Journal, 25-32.

Motowidlo, S. J., & Van Scotter, J. R. (1994). Evidence that Task Performance Should Be Distinguished from Contextual Performance. Journal of Applied Psychology, 79(4), 475-480.

Murphy, K. R., Garcia, M., Kerkar, S., Martin, C., & Balzer, W. K. (1982). Relationship Between Observational Accuracy and Accuracy in Evaluating Performance. Journal of Applied Psychology, 67, 320-325.

Murphy, K. R., Jako, R. A., & Anhalt, R. L. (1993). Nature and Consequences of Halo Error: A Critical Analysis. Journal of Applied Psychology, 78(2), 218-225.

Nathan, B.R. & Alexander, R.A. (1988). A comparison of criteria for test validation: A meta analytic investigation. Personnel Psychology, 41, 517-536.

Organ, D. W. (1988). Organizational Citizenship Behavior: The Good Soldier Syndrome. Lexington, MA: Lexington Books.

Russell, T.L., Crafts, J.L., Tagliarini, F.A., McCloy, R.A., Barkley, P. (1996). Job Analysis of Special Forces Jobs (ARI Research Note 96-76). Alexandria, VA: U.S. Army Research Institute of Behavioral and Social Sciences.

Saal, F. E. (1979). Mixed Standard Rating Scale: A Consistent System for Numerically Coding Inconsistent Response Combinations. Journal of Applied Psychology, 64, 422-428.

Saal, F. E., & Landy, F. J. (1977). The Mixed Standard Rating Scale: An Evaluation. Organizational Behavior and Human Performance, 18, 19-35.

Schmidt, N. Gooding, R.Z., Noe, R.A. & Kirsch, M. (1984). Meta analysis of validity studies published between 1964 and 1982 and the investigation of study characteristics. Personnel Psychology, 37, 407-422.

Schmidt, F.L., & Hunter, J.E. (1998) . The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. Psychological Bulletin, 124, 262-274.

Thayer, P. W. (1977). Somethings Old, Somethings New. Personnel Psychology, 30(4), 513-524.

Thorndike, R. L. (1949). Personnel Selection. New York: John Wiley.

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

Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative Analysis of the reliability of Job Performance Ratings. Journal of Applied Psychology, 81(5), 557-574.

Werner, J. M. (1994). Dimensions that Make a Difference: Examining the Impact of In-Role Behaviors on Supervisory Ratings. Journal of Applied Psychology, 79, 98-107.

Wilson, M. A. (1983). Opposing Models of Employee Selection Data: A Structural Equation Analysis Dissertation., Ohio State University, Columbus.

Woehr, D. J., & Miller, M. J. (1997). Distributional Ratings of Performance: More Evidence for a New Rating Format. Journal of Management, 23(5), 705-720.

Appendix A

A Sample of the team member performance measurement form used in this study.

Appendix B

A Sample of the Team Leader/Sergeant performance measurement form used in this study.

Appendix C

List of Special Forces Job Performance Behaviors Categorized by Dimension

Item Order on Form	I = + II = 0 III = -	Description
Trouble Shooting and Solving Problems		
23	I	Makes the most of resources at hand thinks of novel ways to use available materials; invents or fabricates needed items from seemingly useless materials.
5	II	Uses available resources to resolve problems and to construct needed items; may occasionally overlook some resources that might have been useful.
14	III	Lacks resourcefulness; may simply give up if needed tools are not available or may rely excessively on others to find a way to accomplish a task.
Planning and Preparing for Missions		
4	I	Develops plans that are technically sound, well-coordinated, and likely to lead to mission accomplishment; plans are so well-formed that the briefback is readily accepted.
13	II	Develops workable mission plans that are likely to be successful, although some modification may be needed.
22	III	Develops plans that have critical flaws or that fail to consider second & third order effects of actions; prepares mission analysis that is incomplete or insufficient.
Navigating in the Field		
20	I	Gets from place to place without errors and on time; without having access to a map, correctly uses terrain features and distances traveled to determine approximate location.
2	II	Usually arrives at destination on time; notices and takes into account map or environmental details to facilitate moving to targets.
11	III	Gets lost and arrives very late to destination or not at all; becomes geographically disoriented or confused when not navigating in daylight conditions (e.g., in darkness, rough or unfamiliar terrain, etc.)
Teaching Others		
7	I	Creates novel approaches to capture and hold audience attention or to increase audience interest and involvement; incorporates real-life examples into training
16	II	Uses techniques to maintain attention of the audience during presentations
25	III	Loses control of the training environment or loses audience attention; may read to audience directly from notes or training materials

(table continues

Item Order on Form	I = + II = 0 III = -	Description
Using and Enhancing Language Skills		
26	I	Picks up languages readily; uses language skillfully; translates adeptly, rarely, if ever, miscommunicating information; catches errors in others' translations; may create tools (such as a dictionary) for others to use to communicate more effectively.
8	II	Can communicate sufficiently in most situations, even though language skills are not at a conversational level; uses gestures appropriately to enhance communication; uses a dictionary to aid in communication when needed.
17	III	Lacks language skills; frequently misunderstands, miscommunicates, or cannot communicate. May simply give up or not try to communicate or learn.
Building Effective Relationship with Indigenous Populations		
18	I	Discovers the needs and desires of HN/G personnel and takes steps to satisfy them, provides special skills and services that enhance HN/G respect for and rapport with SF.
27	II	Helps indigenous persons; provides effective services when asked or when the need is obvious; fixes weapons and provides first aid or other assistance to gain HN/G rapport.
9	III	Overlooks or avoids opportunities to build relations with locals, may fail to assist HN/G when rapport could have been built.
Showing Initiative and Extra Effort		
1	I	Puts in whatever time and effort is needed to get the job done; fulfills commitments to multiple projects or missions; overcomes obstacles or unusual difficulties to complete a task or mission
10	II	Completes tasks assignments up to standard in a timely manner
19	III	Leaves work undone to pursue personal interests; procrastinates before starting tasks; fails to follow through on or complete tasks once started.
Handling Interpersonal Situations		
12	I	Deals with others constructively, with tact and diplomacy; is highly adept to persuading others to go along with ideas rather than pushing or forcing own way.
21	II	Is usually polite and courteous toward others; deal effectively with most conflict situations.
3	III	Is inappropriately argumentative and confrontational, often creating tension and worsening conflict situations.

(table continues)

Item Order on Form	I = + II = 0 III = -	Description
Contributing to the Team Effort and Morale		
15	I	Devotes personal time and effort to train team members; teaches unique personal skills to team members to improve their readiness or effectiveness.
24	II	Makes an effort to motivate other team members through actions or words; teaches technical skills in own areas of expertise to team members to ensure team readiness.
6	III	Puts self - interest and priorities above team welfare; avoids or overlooks opportunities to apply personal or technical skills to benefit the team.
Miscellaneous		
28	I	Seeks out opportunity to be cross trained in another SF MOS.
29	III	Regularly subject to disciplinary actions.
30	III	Makes promises or commitments to HN/G that he cannot deliver.
31	I	When required, he would engage and destroy the enemy in accordance with a legal order and established ROE's.
32	I	Refuses to give up despite pain, uncertainties, and adversity.
33	I	Is proficient in performing the duties of his SF MOS.

Appendix D

A Sample of the directions used in this study.