

ANALYSIS OF LEADERSHIP PERCEPTIONS USING MULTIRATER FEEDBACK

Thomas P. Bradley, B.S., M.S.

Dissertation Prepared for the Degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS

May 2004

APPROVED:

Jeff Allen, Major Professor and Program Coordinator  
of Applied Technology, Training and  
Development

Joseph Huff, Minor Professor

Randall Schumacker, Committee Member

Jerry Wircenski, Committee Member

Jon I. Young, Chair of the Department of  
Technology and Cognition

M. Jean Keller, Dean of the College of Education

Sandra L. Terrell, Interim Dean of the Robert B.  
Toulouse School of Graduate Studies

## TABLE OF CONTENTS

LIST OF TABLES.....	iv
LIST OF ILLUSTRATIONS.....	v
CHAPTER	
1.INTRODUCTION.....	1
Background	
Need for the Study	
Theoretical Framework	
Purpose of the Study	
Research Questions	
Limitations: Outside the Researcher's Control	
Delimitations: Within Researcher's Control	
Definition of Terms	
Summary	
2. LITERATURE REVIEW.....	13
Introduction	
Role of Feedback in Leadership Development	
360-Degree Leadership Development Feedback	
Research on Rater Differences	
Contingency Theories of Leadership	
Perception of Leadership	
Summary	
3. METHODS.....	39

Research Design	
Data collection	
Population	
Sample	
Instrumentation	
Reliability Analysis	
Data Analysis	
4. ANALYSIS OF RESULTS.....	62
Data Preparation	
Descriptive Statistics	
Reliability	
MTMR Matrix	
Model Estimation, Fit, and Modification	
Hypothesis Testing	
5. DISCUSSION CONCLUSION AND RECOMMENDATIONS.....	90
Discussion	
Hypothesis Discussion	
Conclusions	
Recommendations	
APPENDIX.....	102
REFERENCES.....	122

## LIST OF TABLES

### Table

1. Organization Type.....	40
2. Ratee's Level in the Organization.....	41
3. Mean and Standard Deviation for Raw and Imputed Data.....	63
4. Cronbach Alpha for Observed Variables by Rater.....	65
5. Comparison of Model Fit Statistics.....	71
6. Fit Data for Modified Model.....	74
7. Factor Loading and Error Terms for Factor Diagrams.....	77
8. Fit Statistics for Factor Diagrams.....	79
9. Second-order Path Diagram Factor Loading.....	80
10. Fit Statistics for Second-order Factor Diagrams.....	83
11. Modified Model Factor Loading.....	85
12. Bootstrap Analysis of Second-order Factor Loading.....	86

## LIST OF ILLUSTRATIONS

### Figure

1. Model of rater perspective from Relationship to Results.....	6
2. Revised Model.....	70
3. Modified Model.....	73
4 Latent factor path diagrams for Relationship factors.....	76
5. Latent factor path diagrams for Results factors.....	78
6. Relationship second-order factor path diagram.....	81
7. Results second-order factor path diagram.....	82
8 . Second-order factor path diagram.....	84

## CHAPTER 1

### INTRODUCTION

#### Background

Organizations seek the means to improve the performance of their leaders. Improving performance usually begins with an assessment of the current level of performance. One assessment tool widely used today is the multitrait-multirater feedback instrument or, as it is commonly called, a 360-degree feedback instrument. Leader performance improvement has been linked with rater scores on multitrait-multirater assessments (L. Atwater, Roush, & Fischthal, 1995; Johnson & Ferstl, 1999; Walker & Smither, 1999).

Numerous interventions exist that could improve leadership and managerial performance. Identifying the intervention with the greatest opportunity to enhance performance is based upon an analysis of the gap between the current and the desired performance (Rossett, 1999). Correctly interpreting the meanings behind the 360-degree feedback may increase the chances of selecting the intervention with the greatest probability of success.

#### Need for the Study

The use of 360-degree leader performance feedback as a means of leadership development is growing (Harris & Schaubroeck, 1988; Penny, 2001). There are often differences in scores across raters. and these differences have been the focus of much research to identify the source of differences or look for agreement among raters (Harris & Schaubroeck, 1988; Maurer, Raju, & Collins, 1998; Mount, Judge, Scullen, Sytsma, & Hezlett, 1998; Shapira & Zevulun, 1989;

Tsui & Ohlott, 1988). Much of the research has looked for agreement between the various raters and rating groups (Harris & Schaubroeck, 1988). Often there is a difference in scores between rater groups. The rater groups are commonly identified as subordinates, boss, peers, and the individual leader being assessed. More often, agreement is found within rater groups. According to Borman (1974) "Raters at different organizational levels probably observe significantly different facets of a ratee's job performance in most organizations. If so their ratings ought to reflect these differences" (p. 105). A difference in rater assessment exists between rater groups but little has been concluded about the source of the between-group differences.

One research direction looked for antecedents to the rating differences (Antonioni & Park, 2001; Brutus & Fleenor, 1998; Cellar, Sidle, Gourdy, & O'Brien, 2001; Penny, 2001; Pulakos, Schmitt, & Chan, 1996). Another research path has suggested that the differences are the result of differences in the organizational level of the rater (Borman, 1974; Brutus & Fleenor, 1998; Pulakos et al., 1996).

This study has attempted to contribute to the knowledge of between-group differences by suggesting that the differences are the result of differences in observer perception. The perception is influenced both by differences between rater groups and leader behaviors toward members of a group. The organizational level of the observer influences the prototype or schema that the rater holds about leadership. That prototype then influences how the leader is perceived and rated by the observer (Lord, Brown, Harvey, & Hall, 2001).

Leadership contingency theories suggest that effective leaders exhibit different behaviors, depending upon the situation and the people involved (Fiedler, 1978; Hersey, Blanchard, & Johnson, 2001; House, 1971; Lord, Brown, & Freiberg 1999).

A second-order confirmatory factor analysis model was applied to a cross-section of 360-degree feedback reports obtained from a commercial database. It was assumed that observers of specific organizational levels would more consistently rate leaders on factors that contribute to the rater's perspective of leadership. Tsui and Ohlott (1988) contended that different organizational levels select criterion differences in judging performance. If a second-order factor is identified that shows consistency for a rater group, then this knowledge will increase the usefulness of multirater feedback. If no second-order factor is found, it is possible that a shared view of leadership does not exist within rater groups and that individual differences, perception, rater error, or some other unidentified variable causes observer differences.

### Theoretical Framework

A 360-degree leader development feedback instrument seeks input from a circle of observers: subordinates, boss, peers, and self. The subordinates are the subordinates who rate the leader behavior in an upward direction. The boss is the leader's supervisor, who rates the leader from a downward direction. Peers are persons who are employed at a similar organizational level and who rate the leader from a horizontal perspective. The leader also completes a self-rating from an inward perspective. All the raters provide feedback about various aspects of

the leader's leadership and management competencies (Antonioni & Park, 2001; Penny, 2001). Each group of observers evaluates the leader from their own perspective, and each group of observers typically has different scores for a given competency (Borman, 1974; Penny, 2001; Toegel & Conger, 2003).

Researchers have been searching for the meaning of differences in rater scores. Agreement is expected because the raters are each evaluating the same individual. However, a certain level of agreement has not been found.

Correlations between self and other ratings are modest at best (Harris & Schaubroeck, 1988). Numerous studies have attempted to explain the differences between rater groups (Antonioni & Park, 2001; Cellar et al., 2001; Penny, 2001; Pulakos, et al., 1996). These studies looked at leader attributes, rater characteristics and leadership style. Each are discussed in greater detail in chapter 2.

#### Why Rater Differences Exist

Perceptions of the leader and leadership behavior may influence rater scores. Leadership is very much in the eye of the beholder. Followers, not the leader or researchers, define leadership (Lord & Maher, 1991; Meindl, 1995). If leadership is defined from the perspective of the observer, then the perspective of the rater is important in understanding the 360-degree assessment. In fact, Mount et al. (1998) found that individual ratings were more valid than the aggregate score of all the raters. Individuals determine how to rate a supervisor based upon their own perspective and schema. Individuals in workgroups are

then likely to influence each other through shared observation, shared perceptions, and social contagion (Lord & Maher, 1991; Meindl, 1995).

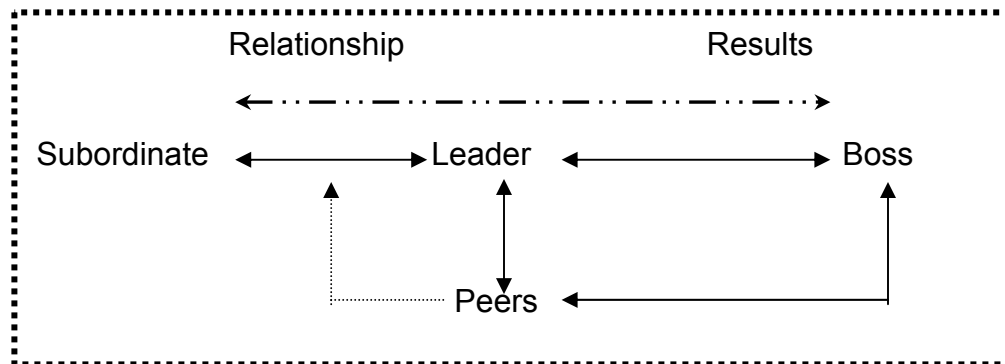
#### Model of Rater Perspectives

The model in Figure 1 depicts the dyadic interactions considered in this study. The model depicts interactions between subordinate and leader and also between the leader and the boss. From the perspective of the individual, to the right of each dyad the primary interest is in the relationship. The individual on the left of each dyad has a greater interest in results. The continuum from relationship to results above the dyads indicates the theory that relationship is more important to subordinates and results are more important to the boss.

The interactions are complex and are based upon how the individual raters perceive themselves and the leader. The subordinates' perception of the relationship they have with their leader influences their perception of the organization and subsequently their role in the organization (Lord et al., 2001). The leader is concerned with achieving results through the subordinates and in presenting the best image to the boss (Ulrich, Zenger, & Smallwood, 1999). At the same time, the leader recognizes the subordinates needs for a dyadic relationship (Lord et al., 1999). The leader fills a subordinate role when interacting with the boss and would seek to maintain a positive relationship.

Peers occupy a position where they too have an interaction with the boss and the leader. Peers have the unique opportunity to observe subordinate-supervisor relationships and interactions in a way that the boss does not see

(Zazanis, Zaccaro, & Kilcullen, 2001). Coupled with their understanding of organizational goals, peers are in a good position to provide feedback.



*Figure 1.* Model of rater perspective from relationship to results.

Note: Solid line indicates interactions. Dashed line indicates observation of an interaction. Broken line indicates the perspective of the rater, and dotted line around the model suggests an open system.

Organizations are open systems that engage in transactions with other systems (Nickols, 1989). The dotted line surrounding the model depicts an open system. The open system indicates that outside influences may interact with the observers and bias their ratings.

#### Purpose of the Study

Each group of raters observes the leader from their own perspective, schema, and bias (Borman, 1974). Raters at different organizational levels observe different facets of a leader's job performance. Members of the subordinate group perceive the leader from an interaction and relationship perspective, and they often interact on a more frequent basis with the leader than

other groups. The quality of that relationship is predictive of workgroup outcomes (Gerstner & Day, 1997). The relationship influences the follower's perception and schema of leadership. The schema of cognitive categories used by followers to define leadership has important implications concerning the level of influence the leader has over the subordinates (Lord et al., 1999). The nature of the subordinate supervisor relationship results in dependence on the part of the subordinate that affects perceived leadership. Lord et. al. suggested that the subordinate's observations are biased because of the relationship.

At the other end of the continuum is the leader's boss. Here, the subordinate-supervisor relationship is reversed. In this case the boss is in a position to observe more the results of the leader's efforts (Ulrich et al., 1999). Workgroup results are the traditional measure of leader performance and they are of great interest to the organization and the boss. Therefore, it is suggested that the boss will have a bias toward items that measure workgroup and leader Results.

Peers observe leader behavior from a different perspective. Peers have knowledge and expertise and responsibility for organizational results similar to that of the leader (Maurer et al., 1998). Peer ratings, therefore, may contain a bias toward organizational results. Peers also observe leader-subordinate interactions that the boss does not see. This perspective gives the peers a broader insight into leader behavior.

Self-ratings provided by the leader are likely to be biased to put the leader in a good light with the boss. Leaders may overrate their performance to appear

better than they are or they may underrate performance so as not to appear to be boasting (L. Atwater et al., 1995). Self-report data also tend to be unreliable (Podsakoff & Organ, 1986). Therefore, the self-ratings are not expected to show a strong bias in either results or relationship measures.

Understanding the perspective and bias of the rater can improve our interpretation of 360-degree feedback. This perspective can provide more insight into the cause of performance gaps within a workgroup and lead to more meaningful interventions. Tsui and Ohlott (1998) posited the following:

By isolating the sources of dissensus, researchers can identify when and to what extent interrater agreement is possible. By clarifying the bases for disagreement, managers can determine which raters would be appropriate for what evaluation purposes. (p 780)

This study examined homogeneity within the rating groups at different organizational levels and sought to expand our understanding of the perspective of the rating groups.

### Research Questions

Each of the four rater groups, subordinates, peers, boss and self, has a different perspective and bias in observing leadership behaviors. The different raters observe, encode, store, and recall different information about the same leader (Tsui & Ohlott, 1988).

Subordinates perceive the leader from a common perspective. Their interests are more focused on the interpersonal relationship between themselves and the leader.

The boss has a greater interest in the performance and output of the leader and the workgroup. Therefore, the boss will best be in a position to rate the results achieved by the leader.

Peers occupy a unique position in that they can observe leader behaviors and workgroup results, and at the same time peers are cognizant of and responsible for similar organizational goals. Peers then, are in a position to observe behaviors related results.

Leaders themselves are responsible for attaining organizational results through the efforts of others. It is through the relationship behaviors that results are obtained. Therefore both factors are important. However, self-report data are unreliable and are not expected to show strong bias in either direction. These four perspectives suggest four research questions:

1. Factor loading for self-ratings will be statistically different, and Results factor loading will be greater than the Relationship factor loading.
2. Factor loading for boss-ratings will be statistically different, and Results factor loading will be greater than the Relationship factor loading.
3. Factor loading for subordinate-ratings will be statistically different, and Relationship factor loading will be greater than the Results factor loading.
4. Factor loading for peer-ratings will be statistically different, and Results factor loading will be greater than the Relationship factor loading.

#### Limitations: Outside the Researcher's Control

1. This research examined the perceptions of management and leadership behavior and makes no claims as to the effectiveness of those behaviors.

2. The research was limited by the constraints of the data set and instrument under analysis.

3. Perception was not the only factor influencing the differences in rater scores. Other variables and measurement error may have influenced the results obtained.

4. The assumption was that respondents made fair and accurate assessments of the leader based upon their observation and perception of the leader's behavior.

#### Delimitations: Within Researcher's Control

1. Demographics have been studied in several other works and are discussed in the literature review. This analysis did not consider demographic variables of either the rater or leader.

2. Agreement between rater groups may exist due to the fact that ratings are the result of leader behavior. However this study will not look for or attempt to define intergroup agreement.

3. Data were collected from a commercial database containing a cross-section of industries. It was not an intention of this research to separate results by industry. Therefore, results may generalize to many industries.

4. The distinction between leadership and management behaviors has been observed in other works. However, it was not desirable here to attempt to make the distinction when considering observer perception.

## Definition of Terms

360-degree feedback: A form of multitrait-multirater analysis that draws responses from four sources: (a) downward from the leader's boss, (b) upward from the leader's subordinates, (c) laterally from the leader's peers or customers, and (d) inwardly from the self (Penny, 2001).

Leadership: Leadership is in the eye of the beholder (Lord & Maher, 1991). It is a social cognitive process that depends upon both the follower and the leader (Lord et al., 1999; Meindl, 1995). Followers - not the leader and not researchers - define leadership (Meindl, 1995). From this perspective leadership cannot and does not exist without followers.

Multitrait-multirater analysis: The evaluation and analysis of multiple performance dimensions measured by multiple sources (Lawler, 1967).

Prototype or schema: The mental representations of leadership held by an observer based upon knowledge or understanding of leadership traits and behaviors or the linking of successful outcomes to a potential leader (Lord et al., 2001; Lord & Maher, 1991).

Relationship: Mutual trust, respect, and influence between a leader and follower (Howell, & Hall-Merenda, 1999).

Results: What to accomplish and how to accomplish it: achieving results means attaining targeted, expected, hoped for, and desired outcomes (Ulrich et al., 1999).

## Summary

Differences in scores among raters on multitrait-multirater assessment instruments have been an ongoing topic of research. The typical multirater assessment of 360-degree feedback instrument seeks input from the boss, peers, and subordinates and a self-rating from the leader. There have been weak correlations between the raters (Harris & Schaubroeck, 1988). Suggested reasons for the weak correlations are numerous and are important to understanding the results of the 360-degree feedback. The point of view of the rater group may be a contributing factor in the disagreement among observers. Subordinates, peers, and supervisor each perceive the leader from a different perspective (Borman, 1974). Perhaps there is agreement within the rater groups that would increase our ability to interpret 360-degree leadership performance feedback.

## CHAPTER 2

### LITERATURE REVIEW

#### Introduction

Feedback serves an important role in leadership development and motivation. However, there is some controversy concerning the actual impact of feedback on leader performance. In some cases feedback has no impact on performance, or it may reduce performance (DeNisi & Kluger, 2000; Kluger & DeNisi, 1996). Other research found specific cases in which feedback did improve leader performance (L. Atwater et al., 1995; Johnson & Ferstl, 1999). When a leader receives low to moderate ratings from observers this results in the greatest improvement in leader performance. Leaders with high ratings from others had little or no change in performance. These findings suggest that the leaders with the greatest need for improvement are the ones most impacted by multirater feedback (Walker & Smither, 1999).

Leadership is a multidimensional phenomenon; it cannot be measured by a single behavior or criterion (Lawler, 1967). This fact encourages the use of multitrait-multirater feedback in spite of the difficulty of rater differences. The ratings are, however, different enough to constitute different methods (Mount et al., 1998). Perhaps this explains why 360-degree feedback instruments are among most popular leader development tools in use today (Toegel, & Conger, 2003)

Raters from different organizational levels do not agree on leader performance ratings (Harris & Schaubroeck, 1988; Tsui & Ohlott, 1988). Research into the antecedents of disagreement between rater groups has not resolved rating differences or to explained the dissensus due to rater or leader characteristics (Antonioni & Park, 2001; Brutus & Fleenor, 1998; Cellar et al., 2001; Penny, 2001; Pulakos et al., 1996).

There are several reasons why the ratings may be different. Leadership itself is based upon contingency factors such as leader-follower relationship, leader behavior, and task requirements (Chemers, 1997; Lord et al., 1999). The quality of the relationship between the observer and the leader may impact rater perception, performance, and subsequent ratings (Lord et al., 2001; Howell & Hall-Merenda, 1999). Leadership contingency theories (Fiedler, 1978; Hersey et al., 2001; House, 1971) imply that differences in ratings are the result of effective leadership.

Leadership is also in part based upon the observer's perceptual processes. Perception involves the creation of a schema or prototype of leadership in the mind of the observer. A leader who matches a perceived schema is considered effective (Lord & Maher, 1991). The perceptions or prototypes may differ between organizational levels, based upon differences in observed leader behavior (Borman, 1974; Penny, 2001). External variables also influence the perceptions of the rater group (Lord et al., 2001).

## Role of Feedback in Leadership Development

### Impact on Leader Performance

Long-standing assumptions hold that 360-degree leadership development feedback improves performance (Johnson & Ferstl, 1999). DeNisi and Kluger (2000) and Kluger and DeNisi, (1996) found that this is not always the case. In approximately one third of the cases studied, feedback was found to reduce performance. Walker and Smither (1999) in a 5-year study of feedback found that there was little to no improvement in the mean score from year 1 to year 2. Gradual improvement was found in year 3 and continued over the 5-year period.

Other research shows that the difference between observer scores and leader self-rating scores predicts performance improvement or lack of improvement. Underraters, those leaders who rate themselves lower than other raters, do not improve from one rating period to the next. However, the underraters do tend to increase their self-ratings to agree with other raters. In-agreement raters, those leaders whose self-ratings are comparable to other ratings, tend neither to show performance improvement nor to adjust their scores (L. Atwater et al., 1995; Johnson & Ferstl, 1999).

Performance improvement does occur in the case of overraters. Overraters are those leaders who rate themselves higher than other raters. The overraters were also found to have the lowest levels of performance prior to feedback (L. Atwater et al., 1995; Johnson & Ferstl, 1999).

Walker and Smither (1999) found that low to moderate ratings from subordinates resulted in the greatest improvement. These findings suggest that

the leaders with the greatest need for improvement are the ones most impacted by multirater feedback. Johnson and Ferstl (1999) concluded that managerial performance will improve to a greater extent the more the self-ratings exceed the subordinate-ratings.

### Feedback and Performance Improvement

Maurer, Mitchell, and Barbeite (2002) found that feedback from subordinates and peers was significantly associated with the manager's attitude toward feedback. Another finding concerned the interaction between self and peer ratings. When discrepancies occurred between self and peer ratings, the leader's attitudes toward 360-degree feedback were more positive. The higher the ratings from subordinates and peers, the more positive the leaders reaction toward feedback interventions. This supports the researchers, proposal that subordinate and peer ratings would have greater influence than ratings from the boss. The reason for the greater influence is that feedback from subordinates and peers is more rare than feedback from the boss. Also, the leader relies on subordinates and peers to accomplish work goals. Ratings from these groups are typically averages of several subordinates and peers, so they may be perceived as more accurate. Subordinate and peer ratings hold greater confidentiality because they come from more than one source; therefore, they may be perceived as more valid.

Perhaps the depth of feedback is another consideration. Walker and Smither (1999) determined that leaders who held feedback meetings with subordinates improved more than other leaders and that those same leaders

improved more in years when the feedback meetings were held than in years when no meetings were held. These findings suggest that the value of feedback can be enhanced when the feedback is discussed.

L. E. Atwater, Ostroff, Yammarino and Fleenor (1998) suggested considering self-ratings and other ratings simultaneously in explaining managerial effectiveness. They found it important to consider the magnitude and direction of disagreement among raters. Their findings indicate that leader effectiveness is highest when self and other ratings are both high. Also, performance is high when other-ratings are substantially higher than self-ratings. Leader effectiveness is lowest when self-ratings are moderately high and other ratings are low. The final analysis comes down to understanding the entire rating including agreement and disagreement and level of rating between rater groups. The connection between rating agreement and performance is an important one since the most common reason for 360-degree feedback is to improve leader performance.

Observers should evaluate the leader on dimensions related to their level within the organization (Borman, 1974). Accordingly, the boss should evaluate the leader on dimensions of job success. Peers and subordinates should evaluate the leader on dimensions of performance deemed appropriate for the respective levels of peer or subordinate. For example, peers may be best able to evaluate conviviality, or sociability and subordinates may best be able to evaluate the ability to lead (Borman, 1974). It is practical to assume that the ratings from others would be of developmental value to the manager. Peer and subordinate

feedback are perceived as valuable because of their scarcity compared to feedback from the boss (Maurer et al., 2002).

### 360-Degree Leadership Development Feedback

#### Multidimensional Nature of Leadership

The case for 360-degree feedback derives from the idea that performance is a multidimensional phenomenon. The analysis of leader performance is complicated by this multidimensional attribute. The idea that a single criterion can measure managerial performance is unrealistic; performance is a function of a number of dimensions (Lawler, 1967). Shapira and Zevulun (1989) cited several researchers who agreed that performance must be evaluated from several perspectives:

Thorndike (1949), in discussing the notion of *ultimate criteria*, related the different aspects of a worker's performance on the one hand to organizational criteria on the other. Ghiselli (1960) suggested that different measures are used for different personal and organizational purposes and Smith (1976) pointed at the absence of one common factor to account for various performance criteria. . . . The strongest statement was made by Dunnette (1963) who said " . . . junk the criterion! Let us cease searching for single or composite measures of job success and proceed to undertake research which accepts the world of success dimensionally as it really exists." (p. 210)

Multidimensionality and popularity of subjective ratings have raised issues of the validity of the measures and procedures. Analysis of the multitrait-

multimethod matrix (Campbell & Fiske, 1959) is a way of analyzing the convergent and discriminate validity of performance data coming from several sources (Shapira & Zevulun, 1989).

#### Multiple Feedback Sources

Multitrait-multimethod feedback is not limited to the observations from one source. The 360-degree feedback process allows the leader to compare his or her self-ratings to the ratings of subordinates, peers, and the boss. It is important to include multiple rater groups in an analysis of leadership performance because raters at different levels capture unique rating variance. These ratings are different enough from those of other observers to constitute a different method (Mount et al., 1998).

Observer groups do not reach consensus in ratings. Low consensus may be caused by performance expectations, performance observed, observer recall, and observer cognition. Ratings of different observers may be equally valid or accurate. The lack of agreement suggests that multiple sources are necessary when assessing overall leadership effectiveness (Tsui & Ohlott, 1988).

An outcome of this feedback process is an increased self-awareness of the leader's skills and behaviors and the anticipated improvement in performance to reduce any discrepancy between ratings. (Maurer et al., 2002). Understanding of how others perceive us serves as a primary source of self-views (Lord et al., 1999). Feedback intervention theory implies that, when a negative discrepancy is detected between performance and standard, people become motivated to change their performance to match the standard (Kluger & DeNisi, 1996).

Evaluations from multiple sources may reduce problems that arise from inaccurate self-assessments. If an individual fails to perceive a weakness, there is little stimulus or motivation to change behavior. Persons with inaccurate self-ratings have been found to be poorer performers than those with accurate self-ratings, and an individual's reaction to feedback is influenced by the level of self-assessment (L. Atwater et al., 1995).

A study conducted by L. Atwater et al., (1995) found that subordinate feedback had positive effects on self-rating and performance for leaders who tended to rate themselves higher than the average subordinate rating. On the other hand leaders who underrated themselves relative to average subordinate scores tended to increase their self-ratings to be in agreement with subordinate scores. The underrating supervisors did not show a change in performance as measured by subordinate feedback. Managers who rated themselves in-agreement with subordinates did not change either their subsequent ratings or their performance. These results imply that leaders seek to establish a match between their self-rating and the ratings of others. The two methods of achieving that match are through adjusting self-ratings or improving performance.

#### Concerns Over MTMR Assessment

Questionnaires used to evaluate leadership measure more than just leader behavior. Observers not only recall behaviors but also actively transform the behavioral information into more general and useful constructs (Lord & Maher, 1991). Therefore, a 360-degree feedback instrument will measure more

than leader behavior; it will contain the constructs of leadership developed by the observers.

Self-report questionnaires, at best, provide soft data. These data are perhaps better than opinion or no data at all but are still inferior to most other types of data. However, many organizational research issues resist other approaches to data collection. Some of the problems associated with self-report data are common method variance, consistency motif, and social desirability. Common method variance occurs when data on two or more variables are collected from one source and correlations are interpreted between them. Consistency motif is the tendency of the respondent to maintain a consistent response in a series of answers. Social desirability is the tendency of respondents to select a response that puts themselves or the person being rated in a favorable light (Podsakoff & Organ, 1986).

No simple statistical procedure has adequately eliminated the problems of same-source variance, and laboratory studies often do not replicate certain behaviors found in field studies such as leader-member interactions. The practical utility of self-report measures means that they will continue in organizational research (Podsakoff & Organ, 1986).

### Research on Rater Differences

#### Differences in Ratings

Previous research in multitrait-multirater feedback has tried to reconcile the differences between rater groups. Studies have compared scores of the boss and subordinates; the subordinates and peers; and peers and boss. A number of

studies have tried to determine the effects of self-ratings compared to the ratings of others (Atkins & Wood, 2002; Harris & Schaubroeck, 1988). Because each rater group is evaluating the same individual leader, it is expected there would be some agreement. However, strong correlations between groups have not been found (Harris & Schaubroeck, 1988; Tsui & Ohlott, 1988).

Correlations between self and other ratings are modest at best. In a meta-analysis of self-supervisor, self-peer and peer-supervisor found moderate agreement between self-peer ( $\rho = .36$ ) and self-supervisor ( $\rho = .35$ ) (Harris & Schaubroeck, 1988). They found higher agreement between peers and supervisors ( $\rho = .62$ ). Understanding the reasons for low agreement is limited (Tsui, & Ohlott, 1988). The lack of agreement has continued as a subject of research and has raised questions concerning the effects of individual differences between raters and between raters and leaders.

#### Antecedents to Rating Differences

*Affect and leniency.* Rater characteristics are one possible source of variation. Antonioni and Park (2001) examined interpersonal affect in a study that specifically investigated "the relationship between the raters interpersonal affect toward leaders and the leniency of ratings form 360-degree feedback" (p 480). In simple terms, does liking a manager influence the level of rating scores? Liking the leader may influence what behaviors the raters observe, and it may influence their subsequent perceptions of leader behavior. This suggests that raters with strong feelings about the leader may pay particular attention to certain behaviors that confirm their feelings. The results suggest that interpersonal affect does

influence rater scores but that the influence is not the same for individual raters or groups. Interpersonal affect exhibits greater influence on subordinates and peers than on supervisors (Antonioni & Park, 2001). The researchers wrote: "We believe that each source of 360-degree feedback may be influenced differently by the same factors: they may even be influenced by different factors" (Antonioni & Park, 2001, p 490).

*Race, gender and level effects.* Pulakos et al., (1996) investigated leader race and gender as well as level effects. The primary purpose of their research was twofold. First, it investigated the differences in rating due to rater race and gender subgroups. Second, it investigated the differences for peer versus boss ratings. They found no statistically significant differences in scores based upon leader demographic differences of race or gender. The research results did suggest that the raters from different organizational levels weigh factors differently. Pulakos et al., (1996) proposed that the disagreement between rater levels is a result of different roles, orientations, and perceptions. They suggest that "raters from different organizational levels may consider different factors in formulating their ratings of leaders. Future research may need to focus on and explore factors other than those investigated in this research to understand the determinants of peer ratings" (p. 117).

Differential item functioning was used to determine whether leader gender or rater group exhibited differences in the measurement itself (Penny, 2001). The study proposed that raters may interpret the text of an item differently from another group of raters and that the resulting differences may be caused both by

differences in observation and interpretive differences of the items. Therefore, groups of raters from different environments may give different ratings.

Only one item was found with differential item functioning due to leader gender, while 55 of the 106 items exhibited anomalous functioning related to rater groups. The effect in each case was in the small to medium range. Further analysis suggests that the differential item functioning due to environmental complexity and leader contingency behavior is a naturally occurring phenomenon and that the interpretation of 360-degree feedback data should take these factors into account (Penny, 2001). This analysis further illustrates the role of perception in observer ratings. Individual items may be interpreted differently, depending upon the observer's organizational level. Complexity and contingency behavior may contribute to these differences in that leaders may exhibit different behaviors, depending upon the situation.

*Leader style, gender and subordinate personality* Leader style, leader gender, and subordinate personality are the variables under consideration in a study of leader evaluations and subordinate motivation by Cellar et al., (2001). This study looked for interactions between rater agreeableness, leader style (autocratic or democratic), and leader gender. Rater agreeableness interacted with leader style to affect leader ratings. Agreeable participants rated democratic leaders as more effective than autocratic leaders. Leader gender had no effect on ratings of leader effectiveness. These results, as well as some of the others above, suggest that a leader's individual differences have little effect on leader ratings. There does seem to be some indication of differences for different

organizational levels of the observers. but what determines those differences is not clear.

*Differences across industries.* Brutus and Fleenor (1998) investigated 360-degree feedback across several industries to determine whether the variance between industries was greater than within industry ratings. Potential differences could include cultural differences, goals, technology, organizational structures, and management systems. The results indicated small but significant differences between organizational types. Leniency bias, for example, was stronger for educational institutions. Peer ratings also were sensitive to leniency effect in public sector organizations. Interrater agreement was lowest in government agencies and highest in education and manufacturing. Within-rater agreement, however, did not differ among various types of industries. Brutus and Fleenor concluded that organization type did affect agreement between rater groups, but not within a rater group.

*Informational difference, rating error, and criteria.* Tsui and Ohlott (1988) considered three explanations for low rater agreement: informational differences, rating error tendency, and variations in performance criteria. They argued that the most plausible source of dissensus lies in the criteria used to judge the leader. The differences lie in both the type of criteria used and importance the observer attributes to the criteria. They also contended that there may be common criteria associated with leader effectiveness.

The Tsui and Ohlott (1988) results suggested that all three are plausible causes of rater dissensus and confirm the lack of consensus in overall ratings

between the rater groups. They did not find differences in the criteria associated with overall judgment of effectiveness. Rather, the results implied that the weight assigned to a given criterion by a rater could be a cause for dissensus (Tsui & Ohlott, 1988).

### Contingency Theories of Leadership

Leader behaviors are predicted to be different for different observers based upon leadership contingency theories. Lord et al., (1999) pointed out that 40 years of leadership research confirm that leadership is contingent in nature and that no systematic framework for understanding the boundaries of the various leadership theories has been developed. Howell and Hall-Merenda (1999) noted that "leader-focused research implicitly assumes a relationship of some sort between leader and follower, and that implied relationship is fundamental to the link between leader behavior and follower response." (p. 680). Results from the Howell Hall-Merenda study showed a significant relationship between the quality level of leader-follower relationship and higher levels of follower performance.

The concept of leadership is defined by the relationship between the leader and the followers (Chemers, 1997). Followers define leadership when they choose to follow someone. That someone may be in a formal position of authority, or that someone may be a peer (Lord & Maher, 1991; Meindl, 1995). The important factor is that the followers have the choice of whom they follow.

Path goal theory (House, 1971), contingency theory (Fiedler, 1978), and situational leadership theory (Hersey et al., 2001) imply that differences in ratings

are the result of good leadership. Different situations, groups and individuals require different considerations from the leader. Thus, different behaviors are observed by various groups in the normal course of leadership behavior.

Leader-initiating structure describes the degree to which leaders initiate psychological structures such as assigning tasks, specifying procedures, clarifying expectations, and scheduling work. Leader consideration describes the degree to which leaders create a supportive environment, are friendly and approachable, look out for the personal welfare of the group, and giving advance notice of change (House, 1971).

Fiedler's (1978) contingency model considered three variables that influenced workgroup performance. The three variables are listed in order of importance; leader member relations, task structure, and position power. Leader member relations are of paramount concern. A leader who maintains high-quality interpersonal relationships is more likely to successfully influence followers. Clarity of task goals is the second consideration and the third variable is the amount of authority the leader held due to position power.

Situational leadership theory contends that there is no one best way to lead people. The theory states that the amount of direction or task behavior that a leader uses and the amount of support or relationship behavior that a leader provides is dependant upon the readiness level of the follower (Hersey et al., 2001).

Research in contingency theories of leadership "reveals that effective leadership is strongly affected by the match or fit between the leader's

orientation, inclinations, and skills and the demands of the leadership situation. The leader's relative emphasis on task versus relationship concerns appears to be an important characteristic (Chemers, 1997, p. 42).

According to Chemers (1997) the effects of contingency leadership manifest themselves in both group productivity and performance as well as in the thoughts and emotions of the group members.

### Perception of Leadership

Cognitive concepts such as leadership are mentally represented in social information-processing theories of leadership. The mental representations are called prototypes, implicit theories, schema, and so on. Leadership is an emerging social process influenced by a number of factors, including context, task, group histories, and the personal qualities of the leader and follower (Lord et al., 2001).

Leadership prototypes vary both within an individual and across a group of observers.

Consistency in leadership perceptions across multiple perceivers may occur when external factors impose very similar constraints in multiple individuals, leading to the creation of similar leadership prototypes within dyads, groups, or other relevant organizational aggregates and variability in prototypes across these higher-level entities (Lord et al., 2001, p 322).

External variables such as culture, task, values, norms, affect, and goals influence group perception when they are stronger than individual internal

constraints. National culture may create a group preference for certain leadership behavior. Constraints on leader prototypes may also emerge from follower characteristics, values, norms, affect, and goals. Over time, group identities and a shared group prototype can become increasingly important in defining leadership. The shift to group level constraints reduces individual differences in leadership prototypes. Self-identity may also influence level of analysis. Wide variability in self-identity will increase the variability of in-group constraints. However, if leaders have considerable influence on subordinate self-identity consistency is likely for subordinates of the same leader (Lord et al., 2001).

Leadership is also defined as a social process that depends upon both the follower and the leader (Lord et al., 1999; Meindl, 1995). Leaders are a vital link between individuals and the organization. The relationship between the subordinate and supervisor determines how the subordinate views the work environment. In turn, the relationship affects subordinates' self-concept and is critical to determining social and organizational processes. Leaders influence subordinates by activating self-concepts or creating new self-schemas (Lord et al., 1999; Gerstner & Day, 1997). According to Chemers (1979), "Effective leaders must build a relationship with subordinates that results in highly motivated, mission-oriented, and goal-directed team members" (p.155).

Leadership may be recognized in any member of an organization. It results from a combination of cognitive, affective and social processes that involve both the observer and the perceived leader (Lord et al., 1999).

## Perception of Observers

Leadership is based on two perceptual processes of the observer: recognition-based and inferential perception. The two processes work together to define the leadership prototype or schema in the mind of the observer.

Recognition-based perception relies on knowledge of leadership traits and behaviors. Identifying a leader involves matching behaviors to the prototype of leadership held by the observer. Inferential perception is formed by linking organizational outcomes to the leader. An observer connects positive outcomes to effective leadership. The prototypes vary according to each individual observation, experience, and perception. The prototypes are shared among coworkers, who subsequently influence each other's prototype (Lord & Maher, 1991; Meindl, 1995).

Little research was found comparing the differences in perception of the leader for the 360-degree feedback circle of subordinates, peers, boss, and self. One interesting parallel study by Yukl and Tracey (1992) examined the consequences of influence tactics on subordinates, peers, and the boss. The study indicated that different influence tactics had different effects based upon the level of the target of the influence. Ingratiation and exchange were moderately effective with subordinates and peers. Exchange has often been found to be effective in obtaining support and assistance from peers. Rational persuasion was more effective with the boss. Ingratiation and exchange are likely to be viewed as manipulative in an upward direction. These results suggest that there is a difference in the relationship transactions, depending upon the

organizational level of the influence target, and hint at the differences in the type of relationship between the leader and the observer groups. The differences in influence will likely affect perceptions and ratings provided by the boss, subordinates, and peers.

In an earlier study by Borman (1974), persons at different organization levels rated 41 secretaries. The results suggested that interrater agreement might be higher in dimensions where raters share meaningful information about ratees. The small sample and the specific nature of secretarial work may not be generalizable to larger populations or may not indicate similar results for leaders. However, Borman speculated that

if different organizational levels really do observe different samples of ratee behavior, and then disagreement in ratings across levels seems likely to be caused by a kind of bias which, if properly analyzed, could prove informative. That is the "bias" comes at least partially from honest differences in orientation and perspective. (p. 118)

There is evidence that members from different levels will observe and perceive different leadership behaviors. How then will the differences be quantified? What factors will help explain the underlying perceptions of a leader's behavior?

The leader's effectiveness is, in part, measured by the ability to achieve objectives through the efforts of others. Achieving organizational results has been shown to imply leadership performance (Lord & Maher, 1991). Research evidence also suggests that the relationship between the leader and observer

affect observer perception. Results and relationship are suitable variables for further analysis.

### Perception of Rater Groups

Ratings from multiple sources are particularly useful because they provide different perspectives of leader skill or behavior (Maurer et al., 1998). It is believed that within rater groups there is a common perspective of the leader (Penny, 2001). Members of the group will influence each other in what has been termed social contagion (Meindl, 1995). Social contagion is expected to be stronger among the subordinate group because they have more interactions with each other and the leader than with the peer group or boss. This perspective should result in homogeneity within the subordinate group. Identifying and building upon the understanding of rater perspective will result in a more complete understanding of 360-degree feedback.

Shapira and Zevulun (1989) found that there is regularity in the way people in different positions evaluate performance. The regularity suggests that there is consistency within rater groups although they acknowledge that each group may observe different aspects of performance. Maurer et al., (1998) used confirmatory factor analysis and item response theory to determine whether there is a measurement equivalence between peer and subordinate ratings. Their analysis explored only one factor, team-building skills. They concluded that the calibration of the observed skill rating scale did not differ significantly across rater groups. This suggests that peers and subordinates recognize the underlying characteristics of team building skills being measured as similar.

*Subordinate's perspective.* The perspective of the rater group is predicted along the lines of rater interest and interaction with the leader. The subordinates, for example, are likely to have a perspective greater in interpersonal relationship with the leader than with peers or the boss. "Subordinates are the direct targets of the leader's behavior and thus can provide feedback to the manager about his or her behaviors from first hand experience" (L. Atwater et al., 1995, p.36). Perhaps subordinate ratings reflect the relative importance of the manager in their work lives, raises, development, assignments, and other outcomes. Subordinates may monitor manager behavior more closely in this regard (Maurer et al., 1998).

The subordinate looks to the leader for vision, consideration, and stimulation (Avolio & Bass, 1988). Many leadership behaviors occur to which the subordinates are the only observers (L. Atwater et al., 1995; Lawler, 1967). Subordinates have a unique position of dyad participant and observer. Therefore, they should more consistently rate leaders on items related to the relationship elements of performance.

*Boss's perspective.* At the other end of the spectrum is the boss. Traditionally, the boss is considered to have the best overall perspective of the leader. The perspective of interest to the boss relates to job performance as measured by organizational results achieved by the leader and subordinate workgroup efforts (Lawler, 1967; Ulrich et al., 1999). This job performance perspective is in contrast to that of the subordinates' and therefore should result

in differences between the scores for the two groups. The boss is most likely to consistently rate the leader on items corresponding to organizational results.

*Peer's perspective.* Peers have a singular opportunity to observe the behavior of the leader as both a rival and a colleague. Peers are witnesses to leader behavior that is not observed by the boss. It is believed that the boss's presence elicits maximum performance. Peers see the leader's behavior when the boss does not and may observe aspects of behaviors of which the boss is unaware (Lawler, 1967; Zazanis et al., 2001). Peers understand the importance of subordinate leader relationships as well as appreciate and are responsible for similar organizational results. Peers have a greater knowledge and expertise relevant to the leader's job level than do subordinates. They may then possess a higher standard or expectation than subordinates for leader performance (Maurer et al., 1998). Therefore, it is expected that peers will consistently rate the leader on items related to organizational results.

*Self-perspective.* Self-ratings may be good predictors of future behaviors, and managers do know more about their own behavior than anyone else (Lawler, 1967). However, self-ratings have numerous measurement problems and tend to be unreliable (Podsakoff & Organ, 1986). Self-raters are categorized relative to the ratings of others as in-agreement, over-estimators, or under-estimators. In-agreement means that the self-rating is similar to the rating of others. Over-estimators rate themselves higher than others, and under-estimators rate themselves lower than other raters (L. Atwater et al., 1995).

Leaders are measured and rewarded by the organization based upon results, but those results are typically achieved through the efforts of others. This fact may influence the manager's self-rating to emphasize results and consider the importance of leader-follower relationship. Due to these issues self-ratings are not expected to load strongly on either relationship or results. That is not to say that self-ratings are unimportant but rather that in-depth interpretation should be left to other research.

### Summary

Leadership performance is a multidimensional phenomenon. No single criterion can measure all the aspects of leader performance (Lawler, 1967). Multi-source feedback utilizes multiple perspectives to reduce some of the error associated with single-source feedback.

The selected review of research into agreement between observer levels reveals that there are differences between rater groups. There is little support for intergroup differences due to individual traits. A number of studies have tried to determine the effects of self-ratings compared to the ratings of others (Atkins & Wood, 2002). Correlations were modest for cross-group ratings and has inspired research into antecedents of the rating differences (Harris & Schaubroeck, 1988).

Leadership is a cognitive concept that is mentally represented as prototypes, implicit theories, or schemas, and these prototypes vary across perceivers. Consistency across perceivers exists when external factors impose similar constraints on members of a group or dyad (Lord et al., 2001).

Leadership research assumes a relationship between the leader and follower, and the relationship impacts both leader behavior and follower response (Howell & Hall-Merenda, 1999). Further, the leader-follower relationship has an impact upon workgroup results. When the coordinated efforts of the leader and the subordinates are appropriate to the situation, positive results in terms of productivity, efficiency and effectiveness are realized (Chemers, 1997).

Perhaps there is an explanation for the observer differences based upon the rater's organizational level. Rater perception is suggested as one possible source of the difference. Observers at different levels observe different behavior and therefore are likely to have a different bias from other observers (Borman, 1974). Perhaps the disagreement between rater levels is a result of different roles, orientations, and perceptions. Observers from different levels may consider different factors in formulating their ratings (Pulakos et al., 1996). This research explored the perception of two factors, relationship and results, in order to further the understanding of determinants in observer ratings

Therefore, the following hypotheses are proposed. Subordinates have a greater self-interest and therefore are more likely to rate Relationship factors over Results. Therefore, hypothesis 1 states: Second-order factor loading for subordinate-ratings will be statistically different, and Relationship factor loading will be greater than the Results factor loading.

$$\begin{array}{l}
 H_1: \\
 \quad H_a : \nu_{1,1} - \nu_{2,2} > 0 \\
 \quad H_0 : \nu_{1,1} - \nu_{2,2} \leq 0
 \end{array}
 \tag{2.1}$$

Peers have responsibility similar to the leader and will therefore rate Results higher than Relationship. Therefore, hypothesis 2 states: Second-order factor loading for peer-ratings will be statistically different, and Results factor loading will be greater than the Relationship factor loading.

$$\begin{aligned}
 H_2: \quad & H_a : \nu_{3,1} - \nu_{4,2} < 0 \\
 & H_0 : \nu_{3,1} - \nu_{4,2} \geq 0
 \end{aligned}
 \tag{2.2}$$

The boss has the organization's goals in mind. The boss is expected to rate the leader higher on factors relating to Results. Therefore, hypothesis 3 states: Second-order factor loading for boss-ratings will be statistically different, and Results factor loading will be greater than the Relationship factor loading.

$$\begin{aligned}
 H_3: \quad & H_a : \nu_{5,1} - \nu_{6,2} < 0 \\
 & H_0 : \nu_{5,1} - \nu_{6,2} \geq 0
 \end{aligned}
 \tag{2.3}$$

The leaders performing the self-rating would most likely rate themselves in the best possible light. Since the organization has an interest in results, the leaders will rate themselves higher on factors related to Results. Therefore, hypothesis 1 states: Second-order factor loading for self-ratings will be statistically different, and Results factor loading will be greater than the Relationship factor loading.

$$H_4: \begin{aligned} H_a &: v_{7,1} - v_{8,2} < 0 \\ H_0 &: v_{7,1} - v_{8,2} \geq 0 \end{aligned} \tag{2.4}$$

## CHAPTER 3

### METHODS

#### Research Design

This research has evaluated an existing commercial data set in an attempt to find a possible cause of the differences in rater scores in multitrait-multirater feedback. The method consisted of: (a) data analysis to identify potential problems that may interfere with the results; (b) an internal consistency reliability study to determine the reliability of the variable measures; (c) a multitrait-multirater matrix analysis to identify trait and method validity; and (d) a second-order confirmatory factor analysis to evaluate possible differences in rater perception.

#### Data Collection

An existing database was utilized in this analysis. The data are from a commercially available 360-degree feedback instrument called the CheckPoint 360° Competency Feedback System™ (used with permission). The data set includes measures dating from January, 2002 to October 2003. The database contains 6,021 ratings of leaders and managers from a variety of organizations and industries (Profiles International, 2001).

#### Population

The population for this study includes all those leaders and managers who have participated in the CheckPoint 360° Competency Feedback System from January 2002 to October 2003. Because these persons are members of organizations that choose to use a leader development instrument, it could be

assumed that there may be differences between these organizations and organizations that choose not to use multirater instruments for developmental feedback.

Table 1

<i>Organization Type</i>	
Type of organization	%
Education	4.8
Military of Government	3.3
Other Public Sector	2.5
Social Services	0.7
Other Non-Public Sector	4.8
Financial Services	9.7
Manufacturing	14.8
Sales	14.9
Health Care	16.1
Utilities, Transportation and Communication	11.9
Consulting	16.4

A wide range of industries, including financial services, manufacturing, health care, retail, and wholesale sales as well as public sector organizations, have utilized the CheckPoint 360° Competency Feedback System. A breakdown by organization type is shown in Table 1. Each rater self-reported his or her

organization type. The percentage column indicates the percent of total responses for that type of organization.

### Sample

The sample comes from an existing commercial data set consisting of over 6,021 sets of feedback ratings. Each rating set includes feedback results from up to 10 raters. The 10 raters include one boss and one self-rating; the remaining 8 are a combination of peers and subordinates. Peer and subordinate scores are aggregate scores based upon simple averages. Leaders observed self-identified their organization level in one of six categories. Table 2. shows the categories and percentage of responses for each category

Table 2

*Ratee's Level in the Organization*

<u>Organization level</u>	<u>%</u>
Senior Executive	15
Executive	14
Mid Management	33
Supervisor	22
Lead Person	3
Other Staff	13

Initial data analysis using Lisrel 8.5<sup>®</sup> Simplis and Prelis (Jöreskog & Sörbom, 2000a, 2000b) was applied to the total data base sample of 6,021 ratings. Reliability and the MTMM matrix also used the entire 6,021 rating

sample. The second-order confirmatory factor analysis was done using Lisrel 8 (Jöreskog & Sörbom, 1993) on successive random samples of 400 ratings involving bootstrap methodology.

### Instrumentation

The CheckPoint 360° Competency Feedback System was originally developed in 1992 for a major oil company. Three industrial/organizational psychologists and a cross-functional focus group defined the critical competencies needed to sustain competitive advantage. The team was charged with defining competencies that were measurable, observable behaviors that are subject to modification or change. Inherent in the definition is the reason to measure competency at all, it is to identify gaps and modify behavior to close performance gaps for the good of the individual and organization. In 1994 an expanded design team completed extensive research into the utility, implementation, theory, and pertinent literature published on 360° instruments that resulted in improvements to the original instrument (Profiles International, 2001).

The resulting instrument is the CheckPoint 360° Competency Feedback System (Profiles International, 2001). The intended purpose of the instrument is to provide feedback from observers on the strengths and developmental needs of the leader. The instrument is administered on-line, and confidentiality is assured for peer and subordinate raters.

The CheckPoint 360° Competency Feedback System is a 70-item survey. The 70 items are scored on a 5-point Likert scale. The rater scores the item by

indicating the frequency of observed leader behavior. A score of 1 indicates that the behavior is never observed. A score of 5 indicates the behavior is always observed. The 70 items are broken down to 17 skills. The 17 skills then define 8 core competencies. The 8 core competencies are communication, leadership, adaptability Relationships, task management, production, development of others, and self-development (Profiles International, 2001).

Profiles International (2001) has defined each of the eight core competencies as follows:

Communication - Including the skills of listening to others, processing information and communicating effectively.

Leadership - Covering the abilities of instilling trust, providing direction, and delegating responsibility.

Adaptability - Encompassing the skills of adjusting to circumstances and thinking creatively.

Relationships - Assessing the capabilities to build relationships and build team success.

Task Management - Gauging the level of aptitude for working efficiently and competently.

Production - Appraising the ability to initiate action and achieve results.

Development of Others - Measuring proficiencies in cultivating individual talents and in motivating successfully.

Personal Development - Including the behaviors of displaying commitment and seeking improvement. (p. 4, used with permission)

These eight competencies were divided into two second-order factors, Results, and Relationship by a panel of subject matter experts. The second-order factors were determined by selecting which competencies best define operational Results and leader-follower relationships.

Three studies of reliability have been conducted on the CheckPoint 360° Competency Feedback System. The internal reliability test was conducted between 1994 and 1996. The resulting reliability scores ranged from .12 to .85. A test-retest reliability study for a sample of 49 managers over a time interval of 8 weeks had scores from .21 to .83. The third study was a factor analytic study that resulted in a seven-factor model, with alpha scores ranging from .75 to .90 (Profiles International, 2001).

#### Reliability Analysis

Whenever a test is scored an assumption exists that there is a true score for that test. It is also assumed that any test will have a certain amount of measurement error. Those measurement errors are assumed to be random (Gall, Borg, & Gall, 1996). In classical test theory the reliability of a test is a measure of how much test error there is in the scores of a test or, in other words, how much consistency there is in the scores from test to test (Crocker & Algina, 1986; Gall et al., 1996).

Reliability can be estimated by several means: alternate form, test-retest, or internal consistency. Alternate form requires two different forms of the test. Scores from each administration will be correlated to determine the reliability coefficient. Alternate form is not a common practice because of the time and

expense of developing and administering two forms of a test. The test-retest method overcomes the need for having an alternate form of a given test. The test is administered, then administered again at a later time. The correlation between the two tests, called the coefficient of stability, is the measure of reliability. The time between tests may be a cause of error. Too much time may result in changes in true score or the rater's perception of the true score. Too little time between tests may result in remembered responses that will cause the coefficient of determination to be too high (Crocker & Algina, 1986; Gall et al., 1996).

Alternate form and test-retest methods are not practical for this study. Because the data come from an existing data set, it is impractical to attempt to contact respondents to take an alternate form or to reevaluate the leader. The time between administrations would also be such that changes could have occurred in the leader so that observers would not be measuring the same levels of attributes for the leader. Observer perception may change over time as new information is added to the observer's leadership schema.

Internal consistency may be measured by dividing the test in half and measuring the correlation between the two halves. However, there are many ways to divide the test. These different divisions can result in different reliability scores.

The coefficient obtained from the split half method is likely to underestimate the reliability coefficient for the full-length test. Errors of measurement due to content sampling are reduced in longer tests. Using the Spearman Brown prophecy formula overcomes this problem by estimating a

corrected reliability coefficient for the full-length test (Crocker & Algina, 1986).

The formula is:

$$\rho_{xx'_n} = \frac{2\rho_{AB}}{1 + \rho_{AB}} \quad (3.1)$$

Where  $\rho_{xx'}$  is the reliability projected for the full-length test and  $\rho_{AB}$  is the correlation between the two half tests (Crocker & Algina, 1986).

Another approach is based upon item covariance. Cronbach's coefficient alpha can be used when items have a range of scoring weights. Cronbach's alpha ( $\alpha$ ) is a widely used method for computing test score reliability (Gall et al., 1996). It is calculated using the following formula.

$$\hat{\alpha} = \frac{k}{k-1} \left( 1 - \frac{\sum \hat{\sigma}_i^2}{\hat{\sigma}_x^2} \right) \quad (3.2)$$

In the formula  $k$  is the number of items in the test. The variance of item is shown as  $\hat{\sigma}_i^2$ , and  $\hat{\sigma}_x^2$  is the total variance of the test (Crocker & Algina, 1986).

Because there are multiple ways to split a test and each may have a different result, and because Cronbach alpha is a conservative estimate, the y internal reliability was calculated using the Cronbach alpha method. The internal consistency was measured for each of the eight competency factors for each rater group.

## Data Analysis

### MTMR Matrix Validation Analysis

The multitrait-multirater matrix presents all the intercorrelations that result when each of several traits is measured by each of several methods (Campbell & Fiske, 1959). Checking these intercorrelations can identify validity issues. At this level of analysis, corrections in test assumptions may be made before subjecting the data to higher order analysis. Validation of any data is often the result of measurement by independent means. Convergent validity is a confirmation of the data for a single trait by independent means, and discriminant validation is the distinction between traits. A test may be invalidated if the correlation between traits is too high when they are intended to differ (Campbell & Fiske, 1959).

Campbell and Fiske (1959) provided labels of the MTMM matrix for convenience of identification and discussion. Appendix D contains the MTMR matrix for this study and may be helpful in understanding the following discussion. Reliability diagonals, one for each method, are monotrait-monomethod values. The reliability diagonals, shown in bold, occupy the space generally identified as the unity diagonal in a correlation matrix. The monomethod block consists of the reliability diagonal and the heterotrait-monomethod triangle. The heteromethod block is made up of the validity diagonal and two heterotrait-heteromethod triangles. A validity diagonal, shown in italics, is a monotrait-heteromethod value. The two heterotrait-heteromethod triangles are not identical.

Four aspects of the MTMR matrix bear on the question of validity. First, values in the validity diagonal should be different from zero and large enough to encourage further examination. Second, values in the validity diagonal should be larger than the values in the column and row of the adjacent heterotrait-heteromethod triangles. Third, the same trait across different methods should correlate higher than different traits using the same method. In other words, the value in the validity diagonal for a given trait should correlate higher than cross-trait correlations in the heterotrait-monomethod triangle. Fourth, a similar pattern of trait intercorrelations should occur in each of the heterotrait triangles for both monomethod and heteromethod blocks. The values are not necessarily the same, but their relative values should be similar (Campbell & Fiske, 1959).

Researchers point out three problems with MTMM analysis as proposed by Campbell and Fiske (1959). First, there are no quantifying methods to determine the degree to which correlational criteria are met. Second, the matrix does not facilitate separation of the method variance and random error. And third, the Campbell and Fiske criteria assume that there are no correlations between trait and method factors (Schmitt & Stults, 1986). Because of these issues the data analysis did not stop here. The MTMR matrix was further analyzed using second-order confirmatory factor analysis.

### Structural Equation Modeling

Factor analysis belongs to the field of multivariate analysis. The term *factor analysis* refers to a number of statistical techniques that resolve a set of variables into a smaller set of hypothetical variables. The hypothetical variables,

called factors, attempt to explain the correlations between observed scores on a number of tests. Factor analysis uses underlying factors to explain the correlations in measured variables. The researcher asks whether there is some factor, which if partialled out, would leave no remaining intercorrelations between the tests. This is called the principle of conditional linear independence. This principle says that the factors shall account for all linear relationships among the variables (Jöreskog, 1979a).

Confirmatory factor analysis and exploratory factor analysis are common structural equation modeling processes. A brief explanation of the differences can clarify how the two are related.

Confirmatory factor analysis (CFA) is a theory based process that evaluates latent variables as they are identified by measured factors. The latent variables can not be directly measured but are identified by factor loading. Factors are identified in advance where the experimenter already has certain knowledge of the variables. Theory determines how the variables are expected to relate to the factors (Bandalos, 1996; Jöreskog, 1979b). CFA methods allow researchers to directly test the fit between theories and data structures, rather than allowing the structure to emerge from the data without regard to theoretical expectations. (Kieffer, 1999)

Exploratory factor analysis predates confirmatory factor analysis. It was developed to extract latent variables from underlying data structures. "Exploratory factor analysis (EFA) is an analytic technique in which the primary concern is to reduce a larger set of variables into a smaller and

more manageable set based upon consistency of data" (Kieffer, 1999, p. 76). Correlations in the data suggest how they load on latent factors. EFA is used to generate theories from data sets or to compare emerging structures to existing theories. (Bradley & Knox, 2003, p. 3)

The differences between EFA and CFA primarily concern their use of theory. CFA is theory testing and has a strong empirical base. EFA requires little or no theoretical or literature base; it is theory generating (Bandalos, 1996). The distinctions between EFA and CFA begin to diminish in actual research practice. Most studies to some extent are both exploratory and confirmatory because research involves some variables that are known and others that are unknown. Strict confirmatory factor analysis with one formulated model that is either accepted or rejected is rare. Strict CFA situations are rare because few researchers are content to reject a model without suggesting an alternative model. Another research design using CFA is the alternative model process. In an alternative model process several alternate models may be proposed a priori. One of the models will be selected based upon the analysis of the data set applied to each model. Research using alternative models is also rare because few researchers specify alternative models. The more common research design is the model generating method in which the researcher proposes one tentative model. If the initial model does not fit the data, the model is modified and tested again. This process may result in testing several models. Model generating is by far the most common approach to CFA (Jöreskog & Sörbom, 1993).

This research used model generating confirmatory factor analysis to identify the best model to data fit. The goal was not just to identify the best fit, but also to determine that each parameter has a meaningful interpretation (Jöreskog & Sörbom, 1993). A multitrait-multimethod model was developed that examines the correlational relationship between the raters and the second-order factors defined by the latent factors.

Any CFA model must be carefully formulated before being confronted by empirical data (Jöreskog & Sörbom, 1993). A two-step process is recommended whereby the observed variables are tested against the latent variables they are intended to measure. Once the latent variables are defined, the latent variable relationships in the structural model are examined (Schumacker & Lomax, 1996). All CFA models are developed from a theory; in this case, the theory is that rater attributes measured by the CheckPoint 360<sup>o</sup> Competency Feedback System may subsequently define two second-order latent factors: Results and Relationship. The theoretical assumption should be tested as the model is analyzed through MTMM matrix evaluation for rater group differences. This step in the process allows for thoughtful consideration of the basic theory and permits modification to strengthen the initial assumptions.

#### Model Fit

Confirmatory factor analysis attempts to find the best fit based upon a priori theory. Model fit is a measure of the difference between original matrix ( $S$ ) and the matrix implied by the model known as the reproduced matrix ( $\Sigma$ ) (Schumacker & Lomax, 1996). Chi-square is a frequently used fit statistic. It may

be considered a "badness-of-fit measure" because a small and nonsignificant chi-square indicates a good fit of the model to the data (Jöreskog & Sörbom, 1993). However, chi-square has some limitations, including being sensitive to sample size. A large sample size will tend to cause chi-square to become large and significant. A significant chi-square indicates that there is a difference between the original and reproduced matrices (Schumacker & Lomax, 1996). Other fit indices were designed to avoid some of the associated with chi-square (Bentler & Bonett, 1980) Therefore additional fit statistics are also evaluated to support conclusions of model to data fit.

Root mean square of error approximation (RMSEA) is a measure of discrepancy per degrees of freedom. "Browne and Cudeck (1993) suggest that a value of 0.05 . . . indicates a close fit and that values up to 0.08 represent reasonable errors of approximation in the population" (Jöreskog & Sörbom, 1993, p. 124).

Normed fit index (NFI) is an incremental fit index which measures the improvement of a target model to a more restricted baseline model (Hu & Bentler, 1999). NFI indicates good fit of the model to the data when its value approaches 1.0.

Goodness of fit index (GFI) does not depend upon sample size to measure model to data fit (Jöreskog & Sörbom, 1993). "The GFI is based on a ratio of the sum of squared differences between the observed and reproduced matrices to the observed variances, thus allowing for scale" (Schumacker & Lomax, 1996, p. 125). The goodness of fit index indicates good fit of the model to

the data when its value approaches 1.0. Hu and Bentler (1999) suggest a value of 0.95 as a cut-off value.

### Model Modification

Modification of the model may be suggested from the data. Modification indices are also part of the Lisrel 8 output (Jöreskog & Sörbom, 1993). It is not uncommon to find discrepancies between the observed and reproduced matrices. Assuming the theory is valid, the researcher may wish to make post hoc modifications to the model. Two conditions must be met before considering changes to the model. First, modifications must be defensible based upon theory. Second, modifications must be verified with another sample (Bandalos, 1996). Before considering modifications to the model the researcher should check the parameter estimates to see if they make sense. Do the parameter estimates fall into the expected range and have the expected sign? (Schumacker & Lomax, 1996).

Modification index (MI) estimates how much chi-square will decrease if the parameter is set free and the model is reestimated. The largest MI shows the parameter with the greatest impact on chi-square. However, if from a theoretical sense the modification does not make sense, the next largest MI should be evaluated (Jöreskog & Sörbom, 1993). Once a modification is made, other parameter estimates and their modification indices are likely to change. Changes should be made one at a time, and the model reestimated. Only then should the researcher proceed to the next modification if it is still defensible.

Another possible modification is to constrain a nonsignificant parameter. Parameters with a small  $t$ -value are not statistically different from zero and therefore may be fixed to zero. Again, theoretical interest must be considered. The theoretical assumptions must support the removal of the parameter (Schumacker & Lomax, 1996).

Once all theoretically valid modifications have been made and verified, the model should be tested with another sample. Results based upon a given sample will have been fit to the original data. The model should be reestimated with other data to verify that the changes generalize to other samples (Bandalos, 1996).

#### Model Selection

Multitrait-multimethod (MTMM) matrix is used to study validity across different assessment methods (Wothke, 1996). Multitrait-multirater (MTMR) is a specific application of multitrait-multimethod (MTMM). Several different methods have been evaluated in analyzing MTMR matrices (Becker & Cote, 1994; Conway, 1996; Wothke, 1996).

Three possible modeling approaches were reviewed with the rationale for selecting one approach over the other two. The three models evaluated were direct product (DP), correlated uniqueness (CU), and confirmatory factor analysis (CFA).

Direct product (DP) is a multiplicative model; it assumes that traits and methods may combine multiplicatively. If the variance in individual measures is multiplicative trait and method effects cannot be separated. This means that there are not separate loading factors for the traits and methods (Conway, 1996). The

DP model assumes that interactions exist when the size of the method effects depends on the size of the correlation among the measured traits (Becker & Cote, 1994). True correlations are assumed to be attenuated by using different methods. For example, different raters may have different conceptualizations of the traits being rated (Conway, 1996). Therefore, the raters will not rate the trait in a similar fashion, resulting in a lower than expected correlation between the raters.

Studies using the DP model have produced mixed results (Conway, 1996). Becker and Cote (1994) found the DP model generally to be the weakest of the three models (DP, CU, or CFA). CFA and CU models tended to provide better fits to the data (Becker & Cote, 1994). Conway (1996) found the DP model to more appropriate for MTMR data than the CFA model on 3 out of 20 matrices. Although the DP model has reasonable performance, it was not used here. The difference between rater groups was of primary interest in this study; therefore, the multiplicative assumption that traits and methods cannot be separated was not applicable. A program for analyzing the DP model, called MUTMUM has not been widely distributed (Wothke & Browne, 1990).

The correlated uniqueness (CU) model assumes additive trait and method effects similar to the CFA model. The method factors are not actually included in the model. Instead, the error factors, called *uniqueness factors*, of the observed variables are allowed to correlate. The CU model assumes that the methods are uncorrelated, whereas CFA does not make that assumption (Conway, 1996).

The correlated uniqueness model has been found to be superior to the DP model or the CFA model in evaluating MTMM matrices (Becker & Cote, 1994; Conway, 1996). The superiority of the CU model has been demonstrated as being less prone to ill-defined solutions and a possible confounding of trait and method effects (Becker & Cote, 1994). This robustness has resulted in a following for the CU model. The assumption of correlated error terms may be a weakness of the CU model. Correlation of error terms is not helpful in analyzing the model in a practical sense. Correlation of the error terms may result from measurement error, halo effect, and rater perception. This study attempted to find possible causes of the interrater differences and that is possible only by evaluating the raters as latent factors.

Confirmatory factor analysis (CFA) is the single most widely used means for analyzing MTMM data, and it has maintained an appreciable following in the literature over other models (Becker & Cote, 1994; and Wothke, 1996). Despite its popularity, some controversy has arisen due to under-identification and nonconvergence problems (Becker & Cote, 1994). The CFA model includes a set of correlated trait factors, a set of correlated method factors, and a vector of error factors. Correlations between trait and method factors are generally constrained to zero (Conway, 1996). The CFA model, like the CU model, assumes additive methods effects (Becker & Cote, 1994).

Becker and Cote (1994) concluded that, estimation problems aside, CFA models tend to fit MTMM data better than DP and CU models. Conway (1996) stated that, when appropriate solutions are considered, the results from DP, CU,

and CFA models are similar. The CFA proposed model evaluates the differences between raters through the use of second-order latent factors (Wothke & Browne, 1990). Confirmatory analysis is the only method discussed that will allow the differences between raters to be evaluated.

The solution to overcoming many of the issues of under-identification and nonconvergence lies in understanding the data prior to analysis. Non-normality, outliers, and missing data can all prevent proper analysis of structural equation models. A component of Lisrel 8 analysis software is Prelis 2 (Jöreskog & Sörbom, 1993), which allows for numerous data transformations. Prelis enables selection of cases for analysis, provides univariate and multivariate skewness and kurtosis statistics, provides other common descriptive statistics, and includes a bivariate normal chi-square test (Schumacker & Lomax, 1996). Additionally, the MTMM matrix may be analyzed using the procedures described by Campbell and Fiske (1959). The correlations between raters and across traits identify validity issues that may be used to test model assumptions before the data are processed using confirmatory factor analysis

Careful analysis of the raw data and the use of the MTMM matrix will help reduce the risks associated with CFA model identification and nonconvergence. The complexity of MTMR analysis and the number of possible causes of differences between raters may make the variable of rater perception too small to identify.

## Data Processing

The data were processed using a second-order CFA model. Two second-order latent factors, Relationship and Results, were identified by loading appropriate competency factors on the second-order factors. Covariance between the second-order factors and rater groups determined whether there is a difference in how each rater group perceives the leader. The structural model was written from the following matrix equation (Schumacker & Lomax, 1996).

$$\eta = \mathbf{B}\eta + \mathbf{\Gamma}\xi + \zeta \quad (3.3)$$

The matrix format of the above structural equation begins to show the details of the analysis. The two second-order endogenous or dependent latent variables, Relationship and Results, are identified by Eta ( $\eta$ ). The 12 exogenous or independent latent variables consist of the eight competency factors and four raters. These independent variables are identified by Xi ( $\xi$ ). The factor loadings, which were analyzed, are represented by Lambda ( $\lambda$ ), and the portion of the latent variable attributed to error is Zeta ( $\zeta$ ). The individual structural equations for this matrix are shown in column 1 of Appendix A.

$$\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ B_{2,1} & 0 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \lambda_{1,1} & 0 \\ \lambda_{2,1} & 0 \\ \lambda_{3,1} & 0 \\ \lambda_{4,1} & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & \lambda_{9,2} \\ 0 & \lambda_{10,2} \\ 0 & \lambda_{11,2} \\ 0 & \lambda_{12,2} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \\ 0 \\ 0 \\ 0 \\ 0 \\ \xi_9 \\ \xi_{10} \\ \xi_{11} \\ \xi_{12} \end{bmatrix} + \begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \zeta_4 \\ 0 \\ 0 \\ 0 \\ 0 \\ \zeta_9 \\ \zeta_{10} \\ \zeta_{11} \\ \zeta_{12} \end{bmatrix} \quad (3.4)$$

Only one measurement model was used for this analysis because the observed variables are used to define only the exogenous latent variables no observed variables were used to define the endogenous latent variables. The following matrix equation describes the measurement model.

$$X = \Lambda_x \xi + \delta \quad (3.5)$$

The corresponding multitrait-multirater matrix depicts how the observed variables and raters define the exogenous latent variables. Column two of Appendix A shows the structural equations for the independent variables. A graphical depiction of the whole CFA model is shown in Appendix B.

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \\ x_{10} \\ x_{11} \\ x_{12} \\ x_{13} \\ x_{14} \\ x_{15} \\ x_{16} \\ x_{17} \\ x_{18} \\ x_{19} \\ x_{20} \\ x_{21} \\ x_{22} \\ x_{23} \\ x_{24} \\ x_{25} \\ x_{26} \\ x_{27} \\ x_{28} \\ x_{29} \\ x_{30} \\ x_{31} \\ x_{32} \end{bmatrix} = \begin{bmatrix} \lambda_{1,1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{1,9} & 0 & 0 & 0 \\ 0 & \lambda_{2,2} & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{2,9} & 0 & 0 & 0 \\ 0 & 0 & \lambda_{3,3} & 0 & 0 & 0 & 0 & 0 & \lambda_{3,9} & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_{4,4} & 0 & 0 & 0 & 0 & \lambda_{4,9} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda_{5,5} & 0 & 0 & 0 & \lambda_{5,9} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_{6,6} & 0 & 0 & \lambda_{6,9} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{7,7} & 0 & \lambda_{7,9} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{8,8} & \lambda_{8,9} & 0 & 0 & 0 \\ \lambda_{9,1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{9,19} & 0 & 0 \\ 0 & \lambda_{10,2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{10,10} & 0 & 0 \\ 0 & 0 & \lambda_{11,3} & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{11,10} & 0 & 0 \\ 0 & 0 & 0 & \lambda_{12,4} & 0 & 0 & 0 & 0 & 0 & \lambda_{12,10} & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda_{13,5} & 0 & 0 & 0 & 0 & \lambda_{13,10} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_{14,6} & 0 & 0 & 0 & \lambda_{14,10} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{15,7} & 0 & 0 & \lambda_{15,10} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{16,8} & 0 & \lambda_{16,10} & 0 & 0 \\ \lambda_{17,1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{17,11} & 0 \\ 0 & \lambda_{18,2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{18,11} & 0 \\ 0 & 0 & \lambda_{19,3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{19,11} & 0 \\ 0 & 0 & 0 & \lambda_{20,4} & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{20,11} & 0 \\ 0 & 0 & 0 & 0 & \lambda_{21,5} & 0 & 0 & 0 & 0 & 0 & \lambda_{21,11} & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_{22,6} & 0 & 0 & 0 & 0 & \lambda_{22,11} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{23,7} & 0 & 0 & 0 & \lambda_{23,11} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{24,8} & 0 & 0 & \lambda_{24,11} & 0 \\ \lambda_{25,1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{25,12} \\ 0 & \lambda_{26,2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{26,12} \\ 0 & 0 & \lambda_{27,3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{27,12} \\ 0 & 0 & 0 & \lambda_{28,4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{28,12} \\ 0 & 0 & 0 & 0 & \lambda_{29,5} & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{29,12} \\ 0 & 0 & 0 & 0 & 0 & \lambda_{30,6} & 0 & 0 & 0 & 0 & 0 & \lambda_{30,12} \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{31,7} & 0 & 0 & 0 & 0 & \lambda_{31,12} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{32,8} & 0 & 0 & 0 & \lambda_{32,12} \end{bmatrix} + \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \\ \xi_5 \\ \xi_6 \\ \xi_7 \\ \xi_8 \\ \xi_9 \\ \xi_{10} \\ \xi_{11} \\ \xi_{12} \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \\ \delta_7 \\ \delta_8 \\ \delta_9 \\ \delta_{10} \\ \delta_{11} \\ \delta_{12} \\ \delta_{13} \\ \delta_{14} \\ \delta_{15} \\ \delta_{16} \\ \delta_{17} \\ \delta_{18} \\ \delta_{19} \\ \delta_{20} \\ \delta_{21} \\ \delta_{22} \\ \delta_{23} \\ \delta_{24} \\ \delta_{25} \\ \delta_{26} \\ \delta_{27} \\ \delta_{28} \\ \delta_{29} \\ \delta_{30} \\ \delta_{31} \\ \delta_{32} \end{bmatrix} \quad (3.6)$$

Factor loadings ( $\lambda$ ) identify the extent to which a given observed variable is able to define the latent variable. They serve as a validity coefficient. In confirmatory factor analysis, latent variables are not directly measured; they are inferred from the relationships with the observed variables. Factor loadings

between the observed variables and the latent variables indicate their correlation. How well the observed variables measure each latent variable is determined by the sum of the factor loadings squared ( $h^2$ ), divided by the number of observed variables ( $n$ ) or  $h^2/n$ . The amount of variance not explained by the observed variables is  $1-h^2$  (Schumacker & Lomax, 1996).

Measurement error ( $\delta$ ) is that portion of the observed variable that measures something other than the latent variable. Error terms are similar to residuals on multiple regression in that they are the part of the observed variable not explained by the latent factor. The error term serves as a measure of reliability (Bandalos, 1996; Schumacker & Lomax, 1996).

#### Structural Equation Program

The analysis was done using Lisrel 8.5 (Jöreskog & Sörbom, 2000a). The Simplis programs used in this analysis are included in Appendix C. The observed variables are coded by rater and trait. Subordinates' scores for the first trait are identified as DR1. Peer scores are designated with a P. The boss and self-ratings are designated by B and S, respectively. Statistical data such as the correlation matrix, means, and standard deviation were drawn from a separate file. Abbreviated names identify the latent variables. Equations in Simplis Command Language (Jöreskog & Sörbom, 1993) format define the relationships between the variables. The program provided loading factors for these relationships. These loading factors were used to analyze the relationships between the rater groups and the two dependant latent dependent variables.

## CHAPTER 4

### ANALYSIS OF RESULTS

#### Data Preparation

The total population from the database was 6,021 cases. A case consists of the leader's self-rating, one boss rating, and up to eight ratings from peers and subordinates. Missing values occurred in some of the cases where one of the rating groups did not participate. For example, a staff manager may not have subordinates, then values are not reported for subordinates for this manager.

Missing values occurred for self-ratings in less than 1% of the cases; boss ratings were missing in 4% of the cases; peer ratings were missing in 5% of the cases; and subordinate ratings were missing in 15% of the cases. Rather than eliminate the cases that had missing data, values were imputed using the EM method in Lisrel 8.5 (Jöreskog & Sörbom, 2000a). EM imputation is a two-step iterative process. The first E-step finds the current estimated value. The second M-step maximizes the expected log likelihood as if it were based upon complete data. This maximization results in the next value for the missing parameter. Using the new value, the E-step is repeated (Rubin & Thayer, 1982).

#### Descriptive Statistics

The revised data set including imputed missing values was processed using Prelis 2.30 (Jöreskog & Sörbom, 2000b). The scores were negatively skewed with values ranging from -0.14 to -1.11. Table 3 is a summary of the descriptive statistics.

Table 3

*Mean and Standard Deviation for Raw and Imputed Data*

Raters / Traits	Raw		Imputed	
	Mean	Std Dev.	Mean	Std Dev.
Self				
Communication	3.98	0.40	4.00	0.41
Leadership	4.22	0.39	4.22	0.39
Adaptation	3.95	0.50	3.95	0.50
Relationship	4.11	0.44	4.11	0.44
Task	4.01	0.49	4.02	0.49
Production	4.12	0.49	4.12	0.50
Develop Others	4.08	0.50	4.08	0.51
Personal Development	4.12	0.46	4.13	0.46
Boss				
Communication	3.90	0.53	3.91	0.52
Leadership	4.06	0.50	4.07	0.50
Adaptation	3.81	0.61	3.82	0.60
Relationship	3.99	0.55	3.99	0.55
Task	4.05	0.57	4.06	0.57
Production	4.02	0.61	4.02	0.60
Develop Others	4.02	0.59	4.01	0.59
Personal Development	4.11	0.57	4.12	0.56

*(table continues)*

Rater/Traits	Raw		Imputed	
	Mean	Std Dev	Mean	Std Dev
<b>Subordinates</b>				
Communication	4.04	0.51	4.05	0.48
Leadership	4.14	0.50	4.14	0.46
Adaptation	4.01	0.54	4.01	0.50
Relationship	4.08	0.54	4.08	0.50
Task	4.18	0.50	4.18	0.47
Production	4.14	0.52	4.14	0.49
Develop Others	4.03	0.60	4.03	0.56
Personal Development	4.16	0.51	4.16	0.48
<b>Peers</b>				
Communication	3.92	0.45	3.93	0.45
Leadership	4.05	0.44	4.06	0.43
Adaptation	3.84	0.48	3.85	0.48
Relationship	3.96	0.48	3.97	0.48
Task	4.06	0.45	4.07	0.45
Production	4.00	0.47	4.00	0.47
Develop Others	3.98	0.50	3.98	0.50
Personal Development	4.04	0.45	4.04	0.45

## Reliability

Reliability was calculated using SPSS standard version release 8.0.0<sup>©</sup> (1997). The Cronbach alpha values for each of the variables are shown in Table 4. Alpha values were calculated for each competency by rater, and an overall alpha value was calculated for each competency for all raters. There were a total of 60,940 observer ratings used in the reliability analysis. Each leader was rated by up to 10 observers.

Table 4

*Cronbach Alpha for Observed Variables by Rater (n = 60,940)*

Traits (72 items)	Raters				Overall
	Self	Boss	Peer	Subordinate	
Communication	0.84	0.89	0.91	0.93	0.92
Leadership	0.85	0.88	0.91	0.92	0.91
Adaptability	0.83	0.87	0.90	0.90	0.90
Relationships	0.82	0.86	0.89	0.91	0.89
Task Management	0.79	0.83	0.88	0.88	0.88
Production	0.86	0.90	0.92	0.92	0.92
Development of Others	0.82	0.86	0.88	0.89	0.88
Personal Development	0.79	0.84	0.84	0.85	0.84

## MTMR Matrix

The correlation matrix from the imputed data is shown in Appendix D, along with the reliability values. Campbell and Fiske (1959) described the means

to evaluate a MTMR matrix. First, the values in the validity diagonals should be sufficiently large to encourage further examination. In this case the between-rater validities for the self and all other observers was rather low. The convergent validity between the self and boss ranged from 0.16 to 0.23. Self and subordinate validities ranged from 0.12 to 0.24, and self to peers ranged from 0.16 to 0.24. The other raters had slightly higher convergent validity numbers. Boss and subordinates ranged from 0.31 to 0.38. The boss and peers ranged from 0.41 to 0.47, and the subordinates and peers ranged from 0.43 to 0.47.

The second aspect is that the value in the validity diagonal should be larger than the values in the corresponding row and column in the heterotrait-heteromethod triangles. The validity value should be higher than the correlations between any other variables having neither traits nor raters in common (Campbell & Fiske, 1959). This aspect held true for the most part. However, there were several exceptions. For example: the correlation between the subordinate's rating for Leadership and the boss' rating for Develop Others is 0.35. The validity value for subordinate's rating of Leadership is 0.33. Considering the exceptions and the relative closeness of the heterotrait-heteromethod values to the validity values there is little evidence that the data demonstrate discriminant validity.

The third aspect is that a variable should correlate higher with an independent means to measure it than measures of different traits with the same method (Campbell & Fiske, 1959). This condition was not met (see Appendix D). In all cases the values in the heterotrait-monomethod triangles were larger than

the values in the validity diagonals. This aspect suggests there is little or no discriminant validity between the raters.

The last aspect defined by Campbell and Fiske (1959) is that the same general pattern of trait relationships occurs in each of the monomethod and heteromethod blocks. The pattern of the values does hold for the MTMR matrix discussed. This suggests that there is some evidence of discriminant validity.

### Model Estimation, Fit, and Modification

#### Original Model

The original model failed to converge. The phi and psi matrix were non-positive definite. Phi is the covariance matrix of the latent independent variables. Psi is the covariance matrix of the errors. Psi contains the level of unexplained variance for each equation. The non-positive definite error (NPD) means that the determinant of a matrix is zero. When this occurs several of the statistics related to the matrix cannot be generated or trusted because they are not valid. NPD occurs due to some collinearity or linear dependency among the observed variables. Collinearity occurs when one variable is a linear combination of some other variable (Schumacker & Lomax, 1996).

A simplified model was tested where only the four rater groups were evaluated as latent variables. The model did converge, but the fit statistics were marginal. Chi-square was significant at 5031, RMSEA was .14, and the normed fit index (NFI) was .87. These values indicate a marginal fit of the model to the data.

The conventional rule of thumb for most fit statistics is a value greater than 0.90. suggested a greater cut-off criterion in the range of 0.95 and 0.06 for RMSEA will result in lower Type II errors and acceptable levels of Type I errors.

Several iterations were tested in order to find the fault in the original model and to develop a model that would begin fit the data. One iteration tested rater scores across the eight competencies only. This analysis resulted in low factor loadings and poor model to data fit statistics. Chi-square was significant at 147,742, RMSEA was .33, and NFI was .38. These values indicate a very poor fit.

Another iteration was tested that combined the competencies and raters but without second-order factors. Chi-square was significant at 3,837, RMSEA was .041, and NFI was .98. Two of these fit statistics indicate a good fit of the model to the data. Chi-square did not indicate a good fit, but that may have been due to a large sample size or complexity of the model. The model was tested on the complete data set of 6,021 sets of ratings.

When the second-order factors were reintroduced to the combined rater and competency model, it failed to converge. This suggests that either the model is too complex or there is no second-order factor for the data as analyzed.

In order to verify the theoretical base of the model, a group of 14 business management students participated in a model definition exercise. The group of nontraditional students had an average of 12.8 years of work experience after high school.

The participants were asked to identify the eight competencies with one or the other of the two second-order factors of Relationship or Results. The majority of participants agreed with the original model except for the competency of Personal Development where they chose Relationship. When this modification was tested, the model failed to converge. This modification was rejected.

#### Revised Model

A *Revised Model* was created to further analyze the data. A copy of the Simplis 8 (Jöreskog & Sörbom, 2000) program is included in Appendix C. The *Revised Model* defined the latent variables for each rater based upon the competencies related to the second-order factors, Relationship and Results. For example, the peer ratings for the observed variables, Leadership, Communication, Relationship, and Develop Others were loaded on a factor called PeerRel, for Peer Relationship. This design has the advantage of increased parsimony and allows for direct comparison of the latent factors. The model shown in Figure 2 was produced using Amos 5<sup>©</sup> (Arbuckle, 2003).

The observed variables were coded for ease of programming and display in the figure. The coding consists of a letter designator for the rater group and a number identifies the competency being rated. The letter designators are DR for Subordinates, P for Peer, B for Boss and S for Self. The number designators are 1 for Communication, 2 for Leadership, 3 for Adaptability, 4 for Relationships, 5 for Task Management, 6 for Production, 7 for Develop Others and 8 for Personal Development. For example the code DR3 is the score for adaptability as rated by the subordinates.

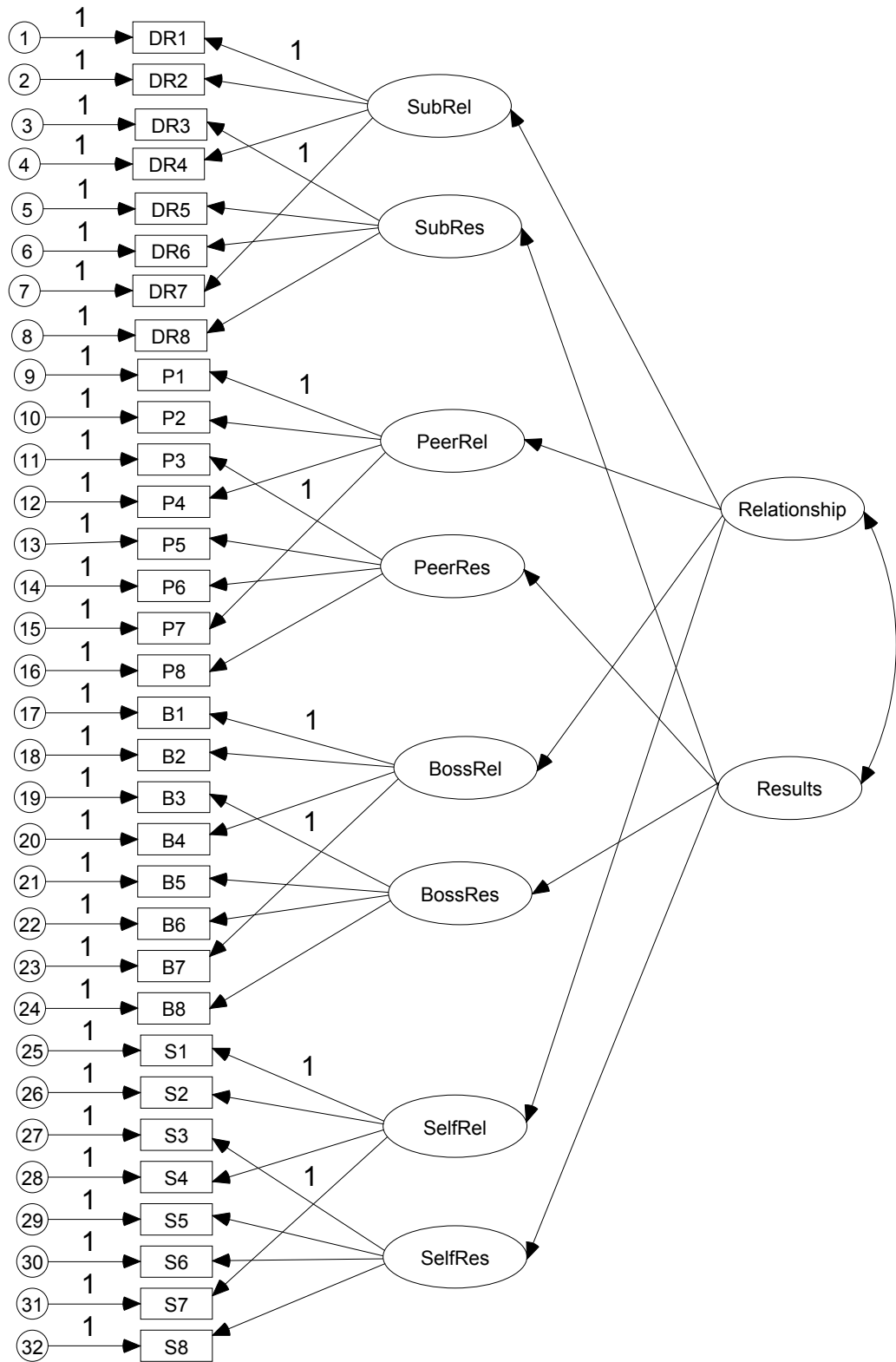


Figure 2. Revised Model with observed variables by rater loading on latent variables representing Results or Relationship by rater.

When the *Revised Model* was tested it did converge although the fit statistics suggested that the data did not fit the model well.

A review of the modification indices for the *Revised Model* suggested that the error terms within rater groups might be correlated to improve data to model fit. Correlating the error terms is called correlated uniqueness (CU). Conway (1996) found that CU models are appropriate for analysis of multitrait-multirater models and that CU models are quite robust. A summary of fit statistics for both the *Revised* and *Modified* model are shown in Table 5.

Table 5

*Comparison of Model Fit Statistics*

Fit statistic	Sample data <sup>1</sup>		Complete data <sup>2</sup>
	<i>Revised</i>	<i>Modified</i>	<i>Modified</i>
Chi Square	5861	1356	13697
RMSEA	0.18	0.089	0.088
RMR	0.066	0.053	0.031
NFI	0.63	0.92	0.94
CFI	0.65	0.92	0.94
GFI	0.49	0.81	0.85

Note. <sup>1</sup>n = 400. <sup>2</sup>n = 6,021.

## Modified Model

Adding the correlated uniqueness resulted in a *Modified Model* that fits the data much better. The path diagram for the *Modified Model* is shown in Figure 3. The fit statistics are included in Table 5. The complete data set including all 6,021 cases was then tested against the *Modified Model*. A summary of these fit statistics is also shown in the third column of Table 5. The large difference in chi-square is due to the larger sample size.

One reason that correlated uniqueness results in better data to model fit is due to covariance of the error terms. Becker and Cote (1994) suggested that measurement effects may result in covariance among competencies when measured by the same method.

Examples of method effects of concern to researchers in our field include socially desirable responding (Arnold, Feldman & Purbhoo, 1985; Rosenkrantz, Luthans & Hennessey, 1983), halo effect in performance appraisal and response sets in self report questionnaires (Spector, 1987). (Becker & Cote, 1994, p. 625)

Method effects are a theoretical possibility within rater groups for this study. Social desirability, halo effect, and response sets may well result in covariance among the error terms for the individual competencies. For this reason the modification was accepted and the model retested with a different sample. Bandalos (1996) stated that, because results obtained from one sample of data will be fitted to idiosyncrasies of those data, they may not generalize to other samples. Therefore, post hoc modification must be replicated on another sample.



Testing for the robustness of the Modified Model included testing the model with 35 random samples of data from the original data set. Each sample consisted of 400 cases with replacement. This bootstrap analysis allowed for the calculation of confidence intervals for the selected fit statistics. The bootstrap samples were taken using SPSS (1997). Table 6 shows the mean standard deviation, standard error of measurement, and 95% confidence interval for selected fit statistics for the *Modified Model*.

Table 6

*Fit Data for Modified Model*

Fit Statistic	Bootstrap			
	M	SD	SEM	95% CI
Chi Square	1132	130	13	(1119; 1144)
RMSEA	0.08	0.01	0.001	(0.076; 0.077)
RMR	0.04	0.01	0.001	(0.037; 0.038)
NFI	0.87	0.01	0.001	(0.87; 0.87)
CFI	0.89	0.01	0.001	(0.89; 0.89)
GFI	0.76	0.13	0.013	(0.75; .077)

*Note:* Cut-off criterion: RMSEA < .06, RMR <.05, NFI > .95 CFI >. 95, GFI > .95 (Hu and Bentler, 1999)

With the final iteration and the *Modified Model* defined, it is prudent to evaluate each of the latent variables independently to examine how each

observed variable defines the associated latent variable (Schumacker & Lomax, 1996). Testing of the structural model, or in this case the second-order factor model, may be meaningless unless the measurement model has demonstrated validity and fit (Jöreskog & Sörbom 1993).

The first step is to examine the individual latent variables that correspond to the Relationship factor. These factor diagrams are shown in Figure 4. The observed variables of Communication, Leadership, Relationships, and Develop Others have been identified as variables that define the factor Relationship. The latent variables are separated into rater groups SubRel for Subordinate Relationship, PeerRel for Peer Relationship, BossRel for Boss Relationship and SelfRel for Self Relationship.

The results of the factor analysis are shown in Table 7. The observed variables are identified, in the table, both by name and corresponding number so that they may be referenced in the figures. It should be noted that the subordinates have a greater consensus on the factor of Relationship than any other group and the self ratings have a lesser consensus. The values in parenthesis are the error terms associated with each factor loading. Error terms indicate measurement error as well as take into consideration the fact that other variables may be used to define the latent construct. The eight models confirm the validity of the four groups in the study. The variance explained shows that the observed variables do measure what they are intended to measure and they define the latent constructs well.

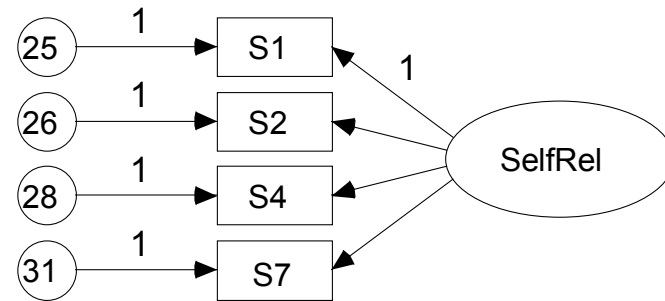
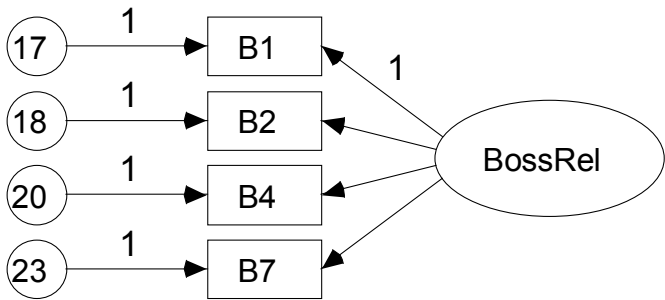
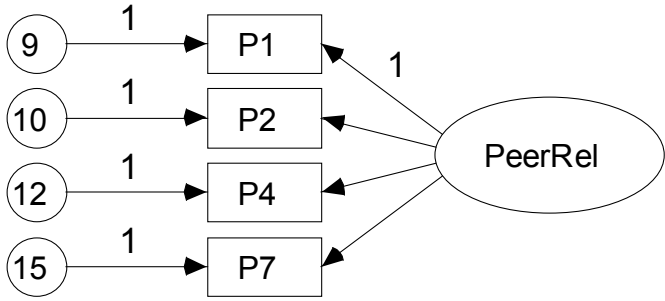
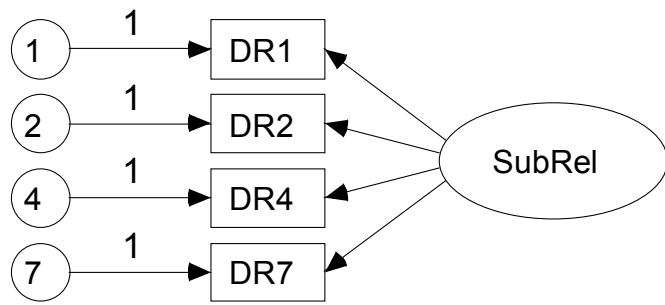


Figure 4. Latent factor path diagrams for Relationship factors. See Table 7 for factor loadings and error term values

Table 7

*Factor Loading and Error Terms for Factor Diagrams.*

Observed Variable	Subordinate		Peer		Boss		Self	
Relationship								
1 Communication	0.951	(0.10)	0.944	(0.11)	0.780	(0.39)	0.834	(0.31)
2 Leadership	0.954	(0.09)	0.906	(0.18)	0.823	(0.32)	0.857	(0.27)
4 Relationships	0.953	(0.09)	0.955	(0.09)	0.942	(0.11)	0.860	(0.26)
7 Develop Others	0.915	(0.16)	0.886	(0.22)	0.880	(0.23)	0.754	(0.43)
Variance Explained	0.890		0.852		0.737		0.685	
Results								
3 Adaptability	0.908	(0.18)	0.879	(0.23)	0.841	(0.29)	0.809	(0.35)
5 Task Management	0.909	(0.17)	0.893	(0.20)	0.808	(0.23)	0.786	(0.38)
6 Production	0.938	(0.12)	0.905	(0.18)	0.882	(0.22)	0.872	(0.24)
8 Personal Develop	0.905	(0.18)	0.861	(0.26)	0.792	(0.37)	0.790	(0.38)
Variance Explained	0.837		0.783		0.691		0.664	

Note. Variance explained ( $h^2$ ) is the percent of variance that the observed variables explain for each of the latent factors. Where  $h^2$  is the sum of the square of the factor loadings divided by the number of factor loadings.

The second step is to examine the individual latent variables that correspond to the Results factor. The factor diagrams are shown in Figure 5. Observed variables are Adaptability, Task Management, Production, and Personal Development. These variables define the factor of Results.

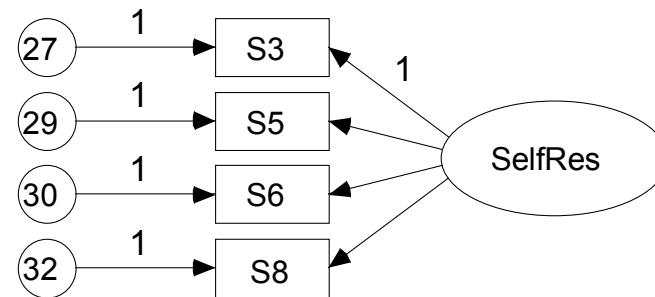
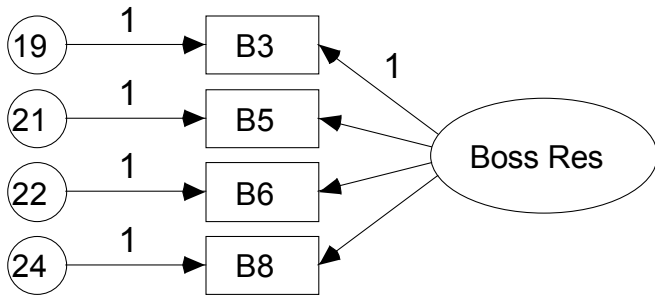
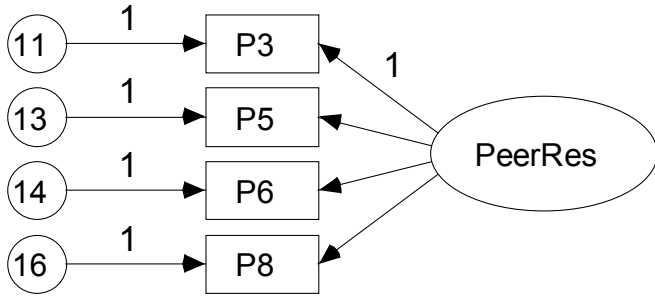
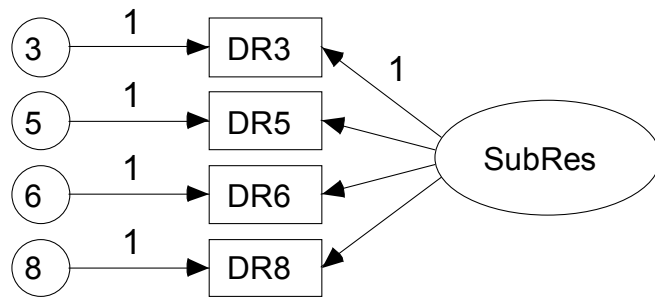


Figure 5. . Latent factor path diagrams for Results factors. See Table 7 for factor loadings and error term values.

The factor loadings and error terms for the Results factors are also shown in Table 7 for convenient comparison between raters and factors. The fit statistics for all of the latent factor path diagrams are shown in Table 8.

Table 8

*Fit Statistics for Factor Diagrams.*

	Chi-Square (df)	RMSEA	RMR	NFI	CFI	GFI
<b>Relationship</b>						
SubRel	17.1 (2)	0.142	0.002	0.992	0.993	0.978
PeerRel	25.5 (2)	0.167	0.003	0.986	0.987	0.971
BossRel	24.3 (2)	0.171	0.007	0.980	0.981	0.969
SelfRel	15.3 (2)	0.129	0.004	0.984	0.986	0.981
<b>Results</b>						
SubRes	54.3 (2)	0.264	0.004	0.969	0.970	0.933
PeerRes	65.1 (2)	0.284	0.005	0.954	0.955	0.923
BossRes	37.2 (2)	0.218	0.010	0.967	0.968	0.952
SelfRes	3.45 (2)	0.044	0.003	0.996	0.998	0.996

*Note.* Cut-off criterion: RMSEA < .06, RMR < .05, NFI > .95, CFI > .95, GFI > .95

(Hu and Bentler, 1999)

These data were generated using a random sample of 400 cases. A sensitivity analysis for chi-square suggests that a sample size of less than fifty would result in a nonsignificant chi-square. The 400 case sample was used

because the larger sample is needed in order to analyze the complete model and it was desirable to maintain consistency..

The latent factors were then combined in the next step to define the individual second-order factors of Relationship and Results. Figures 6 and 7 show the factor diagrams for the two second-order factors.

The second-order factors Relationship and Result are evaluated independently. It is possible to see how the latent factors load on the second-order factor by referring to the factor loadings (gamma) in Table 9. Both second-order factors are shown in the table to facilitate comparisons. Psi contains the variance and covariance among the latent dependent equations. Which is the amount of unexplained variance in each equation (Schumacker & Lomax, 1996). It becomes evident here that Relationship is better defined by the latent factors than Results and that the self ratings are much lower on the Results factor.

Table 9

*Second-order Path Diagram Factor Loading*

Latent Factor	Relationship		Latent Factor	Results	
	Gamma	Psi		Gamma	Psi
SubRel	0.994	0.012	SubRes	0.973	0.052
PeerRel	0.996	0.008	PeerRes	0.962	0.074
BossRel	0.996	0.009	BossRes	0.905	0.180
SelfRel	0.981	0.038	SelfRes	0.465	0.783

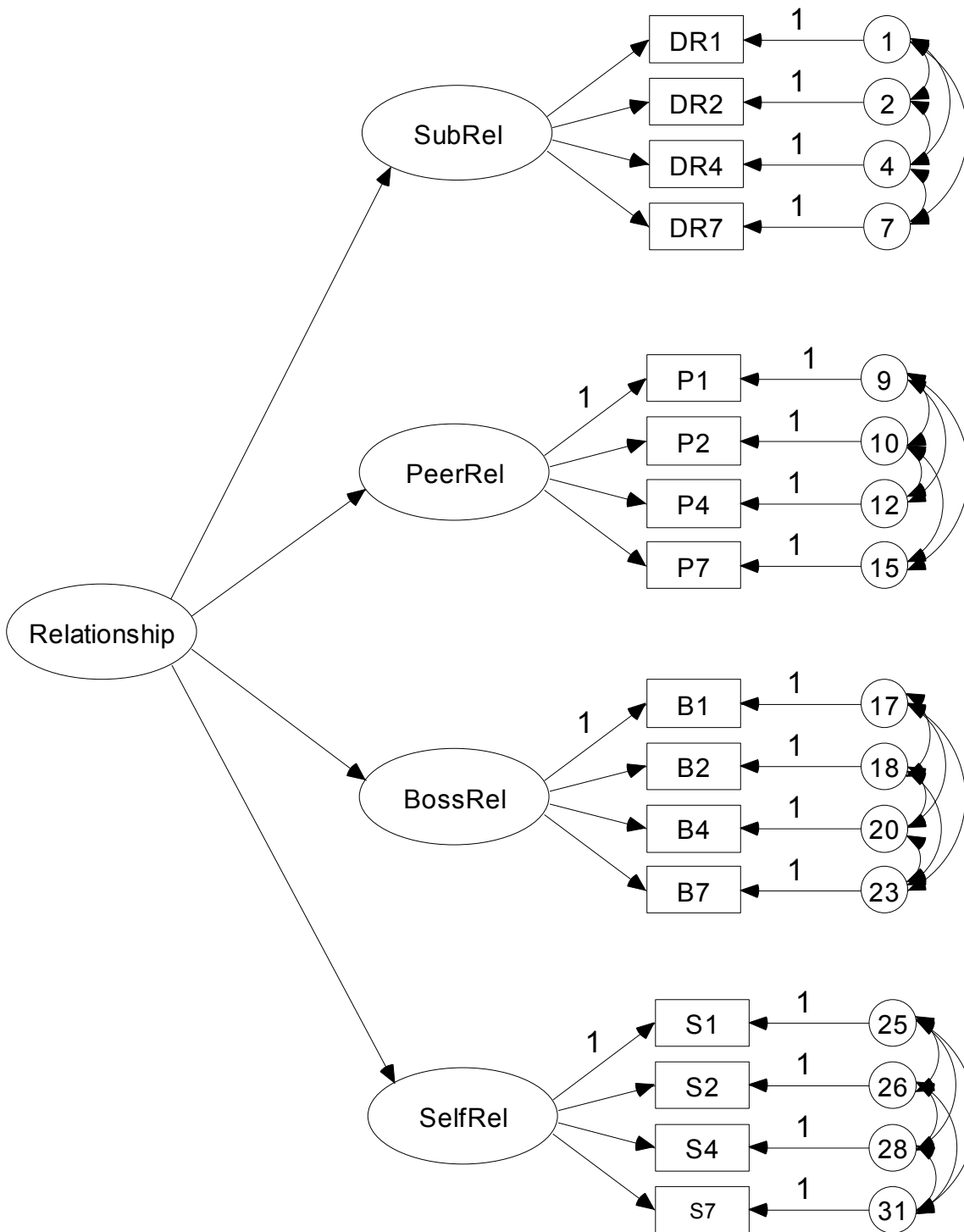


Figure 6. Relationship second-order factor path diagram. See Table 9 for factor loading values.

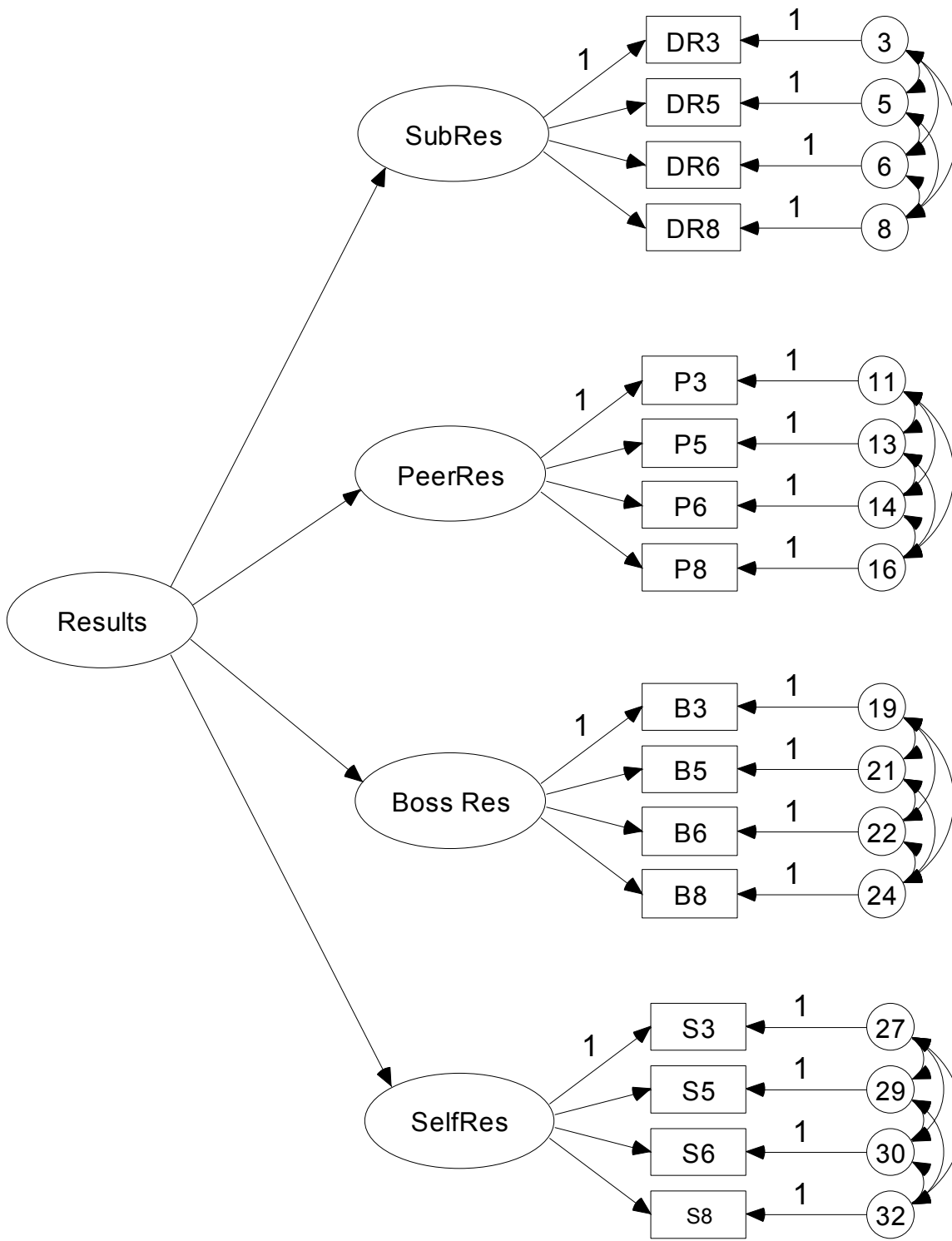


Figure 7. Results second-order factor path diagram. See Table 9 for factor loading values.

Selected fit statistics for the second-order factor models are shown in Table 10. As the model develops and becomes more complex the fit values approach the cut off values. This is expected since as the number of estimated parameters is increased and it becomes more difficult to achieve a desired level of fit.

Table 10

*Fit Statistics for Second-order Factor Diagrams*

	Chi-Square (df)	RMSEA	RMR	NFI	CFI	GFI
Relationship	383.4 (80)	0.094	0.006	0.953	0.964	0.898
Results	635.7 (80)	0.142	0.051	0.900	0.911	0.816

*Note.* Cut-off criterion: RMSEA < .06, RMR < .05, NFI > .95 CFI > .95, GFI > .95 (Hu and Bentler, 1999)

The third and final step is to analyze the complete model. This analysis used entire data set of 6021 cases. The fit statistics, for the *Modified Model*, were previously discussed and shown in Table 5. The path diagram was also previously shown in Figure 3. Because of the complexity of the complete model and for the sake of clarity only the second-order factors and the dependant latent factors are shown in Figure 8. Also shown are the completely standardized factor loadings and the correlation value between the second-order factors. The factor loadings are also shown in Table 11 along with unexplained variance-Psi.

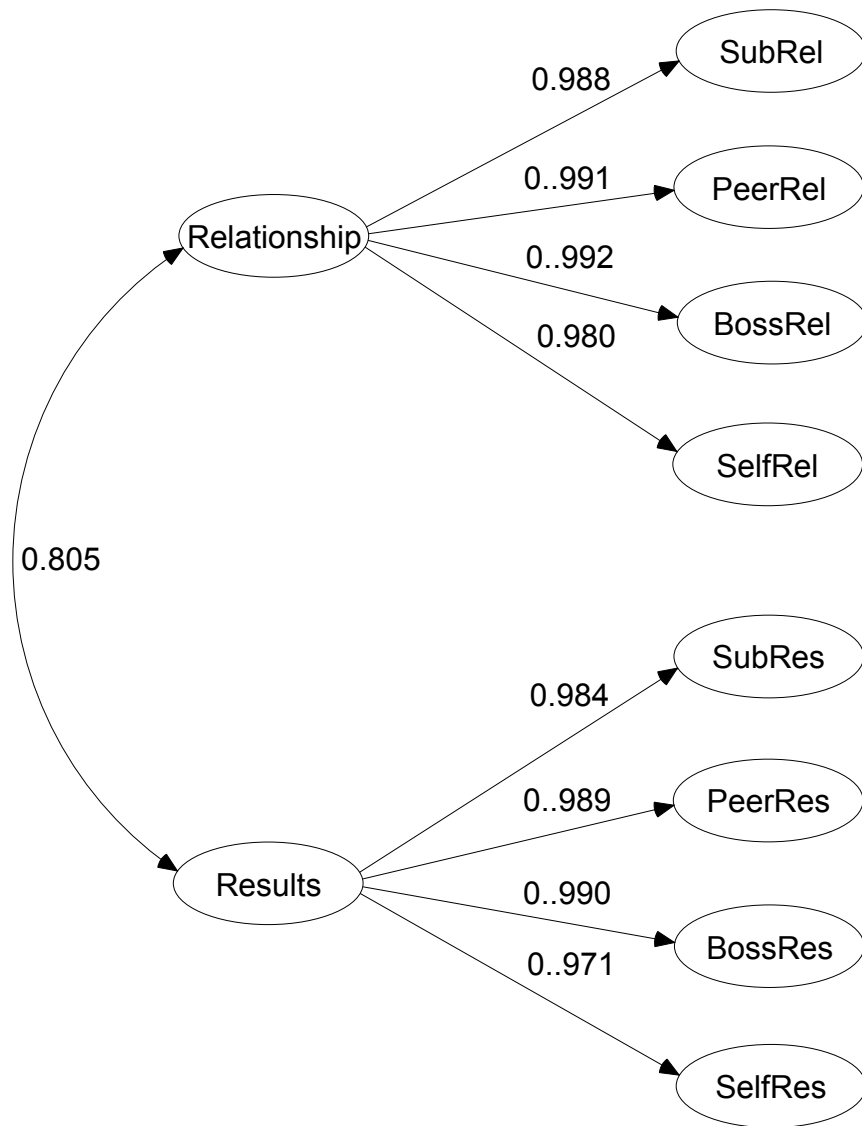


Figure 8. Second-order factor path diagram with factor loading. n = 6021

The factor loadings were all significant at an alpha level of 0.05, which indicates that the observed variables did measure what they were intending to measure. The correlated uniqueness terms indicate relatively high correlation between errors for the observed variables. The standardized values for the eight self-rated variables ranged from .514 to .640. Boss error correlations ranged from .423 to .640. Peer uniqueness correlations ranged from .437 to .733, and Subordinate standardized correlated uniqueness terms ranged from .610 to .785.

Table 11

*Modified Model Factor Loading*

Latent Factor	Relationship		Latent Factor	Results	
	Gamma	Psi		Gamma	Psi
SubRel	0.988	0.023	SubRes	0.984	0.029
PeerRel	0.991	0.021	PeerRes	0.989	0.074
BossRel	0.992	0.025	BossRes	0.990	0.033
SelfRel	0.980	0.032	SelfRes	0.971	0.027

The bootstrap analysis data, shown in Table 12, for the second-order factor loadings indicates the robustness of the factor scores and provide a 95% confidence interval for the factor scores. The analysis also facilitated a comparison of the second-order factor loadings by rater. The bootstrap and t test were conducted using SPSS (1997) with 35 samples of 400 cases.

Table 12

*Bootstrap Analysis of Second-order Factor Loading*

Rater/Factor	Bootstrap				t	p
	M	SD	SEM	95% CI		
Subordinate						
Relationship	0.980	0.008	0.001	(0.979; 0.981)		
Results	0.980	0.017	0.002	(0.978; 0.982)	0.089	>0.05
Peers						
Relationship	0.986	0.005	0.001	(0.986; 0.987)		
Results	0.987	0.007	0.001	(0.986; 0.987)	0.190	>0.05
Boss						
Relationship	0.986	0.006	0.001	(0.986; 0.987)		
Results	0.989	0.006	0.001	(0.988; 0.990)	2.058	<0.05
Self						
Relationship	0.968	0.016	0.002	(0.966; 0.969)		
Results	0.974	0.031	0.003	(0.971; 0.977)	1.076	>0.05

Note. Sample =35, n = 400

A t value greater than 1.96 that indicates there is a statistically significant difference, at an alpha level of .05, between the scores for the two second-order factors for each rater group. The second-order factor loadings are of interest in the research hypotheses for this study. Table 12 allows easy comparison of the

Results and Relationship factor scores. The correlation between the two second-order factors of Relationship and Results is 0.805. This strong correlation indicates that the two factors are related.

The total variance explained for the latent variables defining the second-order factors is very high. Calculating  $h^2$  from the squared multiple correlations, the latent variables of SubRel, PeerRel, BossRel, and SelfRel explain 97.6% of the total variance for the second-order factor Relationship. The latent variables SubRes, PeerRes, BossRes, and SelfRes explain 96.8% of total variance for the second-order factor Results.

### Hypothesis Testing

#### Subordinate Rating Results

Hypothesis 1 states that second-order factor loading for subordinate-ratings will be statistically different, and Relationship factor loading will be greater than the Results factor loading. The null hypothesis, equation 4.1, states that the factor loading for Relationship is less than or equal to the factor loading for Results.

$$H_1: \quad H_0 : \nu_{1,1} - \nu_{2,2} \leq 0 \quad (4.1)$$

Table 12, shows that the difference between the factor loading for Relationship and Results is not statistically significant. Therefore, the analysis failed to reject the null hypothesis.

## Peer Rating Results

Hypothesis 2 states that second-order factor loading for peer-ratings will be statistically different, and Results factor loading will be greater than the Relationship factor loading. The null hypothesis, equation 4.1, states that the factor loading for Relationship is greater than or equal to the factor loading for Results.

$$H_2: \quad H_0 : v_{3,1} - v_{4,2} \geq 0 \quad (4.2)$$

Table 12 shows that the difference between the factor loading for Relationship and Results is not statistically significant. Therefore, the analysis failed to reject the null hypothesis.

## Boss Rating Results

Hypothesis 3 states that second-order factor loading for boss-ratings will be statistically different, and Results factor loading will be greater than the Relationship factor loading. The null hypothesis, equation 4.2, states that the factor loading for Relationship is greater than or equal to the factor loading for Results.

$$H_3: \quad H_0 : v_{5,1} - v_{6,2} \geq 0 \quad (4.3)$$

Table 12, shows that the difference between the factor loading for Relationship and Results is statistically significant. Therefore, the null hypothesis was rejected.

#### Self- rating Results

Hypothesis 4 states that second-order factor loading for self-ratings will be statistically different, and Results factor loading will be greater than the Relationship factor loading. The null hypothesis, equation 4.1, states that the factor loading for Relationship is greater than or equal to the factor loading for Results.

$$H_4: \quad H_0 : \nu_{7,1} - \nu_{8,2} \geq 0 \quad (4.4)$$

Table 12 shows that the difference between the factor loading for Relationship and Results is not statistically significant. Therefore, the analysis failed to reject the null hypothesis.

## CHAPTER 5

### DISCUSSION CONCLUSION AND RECOMMENDATIONS

#### Discussion

##### MTMR Matrix

Review of the MTMR matrix in Appendix D shows good reliability for the values in the reliability diagonal. The consistency of the raters within-groups rating individual competencies is good. Further analysis is dependent upon good reliability. The question of validity is more complicated to evaluate.

The values in the validity diagonals (in italics) were low to moderate at best. These values suggest that observers rating the leadership competencies do not agree on the competencies being measured or on the rating for those competencies. This is not surprising since understanding the differences in observer rating was the purpose of this research. The values provide enough evidence of validity to precede with the analysis.

Discriminant validity is also of interest. It is important that there be a difference in how observers perceive the different competencies. Three aspects of a MTMR matrix may be used to provide evidence of discriminant validity. Again, the matrix in Appendix D failed to show strong evidence of discriminant validity. This suggests that there is not sufficient distinction between the measured leadership constructs being rated by the various observers.

These findings are similar to those found in the literature. For example, Harris and Shaubroeck (1988) found weak correlations between observer ratings. Reasons for the differences range from differences in rater perception

(Lord et al., 2001; Howell & Hall-Merenda, 1999), differences in leader behavior (Fiedler, 1978; Hersey et al., 2001; House, 1971), and measurement error, including response pattern errors, halo effect, and leniency bias.

Campbell and Fiske (1959) argued that the convergence of independent measures requires that agreement between the measures be demonstrated. In 360-degree feedback instruments there is typically no convergence. This lack of convergence was replicated here and provides the motivation to search for the reasons for divergence.

#### Model Fit and Estimation

The next step in the analysis was to evaluate model fit. The *Modified Model*, Figure 3., demonstrates a reasonable fit to the data when model complexity is taken into account. Values 0.95 are considered cut-off values for NFI and other fit indices that range from 0 to 1.0. RMSEA should be less than .06 (Bentler & Bonett, 1980). Traditional cut-off values were in the range of 0.90 for GFI and other fit indices.

Model fit was evaluated for both a small sample of 400 cases and the complete data set of 6,021 cases. The reason for this was to examine the effects of sample size on the chi-square value. In Table 5 the chi-square value for the *Modified Model* with a sample of 400 was 1,356. With the sample of 6,021, chi-square became 13,697. Both values were significant, so chi-square does not indicate a good model to data fit for the given sample sizes.

Model fit was further analyzed using 35 bootstrap samples. The results, summarized in Table 6, suggest that the fit statistics are rather stable due to the small standard error of mean and close confidence intervals.

Model complexity is a mitigating issue, in addition to sample size, when evaluating model fit so it is prudent to evaluate the fit of the latent variable path diagrams from Figures 4 and 5 as well as the second-order path diagrams in Figures 6 and 7. The fit statistics, as shown in Table 8, for the latent factor path diagrams all indicate good data to model fit with the exception of chi-square and RMSEA. As noted earlier a sensitivity analysis suggests that both chi-square and RMSEA would indicate good data to model fit with a sample size of less than 50 cases. A sample of 400 cases was used to maintain consistency throughout the analysis since the larger sample was required in order to analyze the *Revised and Modified Models*.

The second-order path diagrams had good data to model fit for the Relationship factor as shown in Table 10, though the fit statistics were not as strong for the Results factor. Examining the difference between the latent factor fit statistics and the second-order fit statistics it is easy to see how adding a level of complexity to the model impacts the overall impression of model fit.

Though the fit statistics are below the recommended cut off values, the model to data fit should be considered strong enough to encourage further analysis for several reasons. First, there are indications that the second-order factors are worthy of further research. Second, explaining validity issues in 360-degree feedback instruments should not be abandoned too quickly. The level of

fit may prevent making comprehensive generalizations, it does not prevent speculation and recommendations based upon the findings.

#### Model Estimation and Analysis

The latent factor loadings, in Table 7, were statistically significant for all rater groups. The individual factor loadings indicate the correlation between the observed variable and the latent variable they measure. Table 8 further helps to understand the meanings of the factor loadings. The Table 8 shows the total variance explained by the factor loadings. Variance explained ranges from 10.9% for self-ratings defining Results to 29.2% for peer ratings defining Results..

Typically this means that the observed variables are measuring some other variable, the data are unreliable or there is a second-order factor (Schumacker & Lomax, 1996). The reliability data in Table 4 imply that the data are reliable and the model tested includes second-order factors. Therefore, one may conclude that the observed variable includes some other variable. The nature of 360 degree feedback would also lead to this conclusion. Research has postulated that different raters view leadership in different ways (Borman, 1974; Brutus & Fleenor, 1998; Pulakos et al., 1996). Other research contends that there are antecedents that predispose raters to different perceptions of leadership (Antonioni & Park, 2001; Borman, 1974; Brutus & Fleenor, 1998; et al., 2001; Penny, 2001; Pulakos et al., 1996). In either case it is expected that the unexplained variance may be large.

Unexplained variance is one reason that correlating the error terms greatly improved model fit. Additionally, the correlated uniqueness values suggest that

there are method effects influencing rater scores. Examples of method effects include socially desirable response, halo effect, and response patterns (Becker & Cote, 1994). Therefore, high correlated uniqueness and low factor scores are not unexpected. The well-documented variation between raters perceptions, prototypes and observations explains the scores.

What was unexpected was the higher factor loading values for all raters on the second-order factor, Relationship. Table 9 and 11 shows the second-order factor loadings and the standard errors. In each case the factor loadings are statistically significant but the difference between Results and Relationship is not statistically significant. Three of the four hypotheses for this study posited that factor loadings for Results would be greater than Relationship. It was expected that only the subordinates would have higher factor loadings on Relationship due to their direct day-to-day contact with the leader. Boss, self and peers would indicate Results as higher. Each rater group is further discussed individually.

### Hypothesis Discussion

#### Subordinate Rating Results

Hypothesis 1 states that second-order factor loading for subordinate-ratings will be statistically different, and Relationship factor loading will be greater than the Results factor loading. The null hypothesis, equation 5.1, states that the factor loading for Relationship is less than or equal to the factor loading for Results.

$$H_1: \quad H_0 : \nu_{1,1} - \nu_{2,2} \leq 0 \quad (5.1)$$

The analysis failed to reject the null hypothesis. The factor scores for Relationship and Result are not statistically different for subordinates. This suggests that subordinates recognize both Relationship and Result factors as important to the functions of a leader.

### Peer-Rating Results

Hypothesis 2 states that second-order factor loading for peer-ratings will be statistically different, and Results factor loading will be greater than the Relationship factor loading. The null hypothesis, equation 4.1, states that the factor loading for Relationship is greater than or equal to the factor loading for Results.

$$H_2: \quad H_0 : v_{3,1} - v_{4,2} \geq 0 \quad (5.2)$$

The analysis failed to reject the null hypothesis. The factor scores for Relationship and Result are not statistically different for peers. This suggests that peers also recognize both Relationship and Result factors as important to the functions of a leader.

### *Boss Rating Results*

Hypothesis 3 states that second-order factor loading for boss-ratings will be statistically different, and Results factor loading will be greater than the Relationship factor loading. The null hypothesis, equation 4.2, states that the

factor loading for Relationship is greater than or equal to the factor loading for Results.

$$H_3: \quad H_0 : \nu_{5,1} - \nu_{6,2} \geq 0 \quad (5.3)$$

The null hypothesis is rejected for the boss ratings. A small but statistically significant difference was found between the Relationship and Results factors and the boss did identify Results as greater than Relationship as predicted.

#### *Self Rating Results*

Hypothesis 4 states that second-order factor loading for self-ratings will be statistically different, and Results factor loading will be greater than the Relationship factor loading. The null hypothesis, equation 4.1, states that the factor loading for Relationship is greater than or equal to the factor loading for Results.

$$H_4: \quad H_0 : \nu_{7,1} - \nu_{8,2} \geq 0 \quad (4.4)$$

The analysis failed to reject the null hypothesis. The factor scores for Relationship and Result are not statistically different for self raters. This suggests that self raters recognize both Relationship and Result factors as important to their role as a leader.

The only statistically significant difference between the second-order factors was with the boss group of raters. The actual difference was very small and raises the question of practical significance. It was expected that the

differences would be much greater. The fact that the values were close indicates that though there is a difference the two constructs are important to the boss in providing feedback to a leader.

Another interesting observation is that the factor scores for all raters are very close for both factors. It was not the purpose of this analysis to find agreement between rater groups but the data suggest that there is more agreement between raters than expected given 360-degree feedback literature (e.g. Harris & Schaubroeck, 1988; Maurer, Raju, & Collins, 1998; Mount, Judge, Scullen, Sytsma, & Hezlett, 1998; Shapira & Zevulun, 1989; Tsui & Ohlott, 1988). Factor scores for the latent variables showed that the different raters did not agree upon the observed constructs used to define leadership. Yet the agreement between raters does seem to occur on the second-order factors of Relationship and Results.

The two second-order factors Relationship and Results correlate highly in this model with a correlation value of .805. Typically when a high correlation exists between second-order factors the question should be asked if the two factors actually indicate only one factor? In this case that one factor may be called Leadership. If the assumption is made that raters are considering one factor, Leadership, when performing their ratings then this explanation is valid.

However the theoretical foundation for this analysis precludes combining the Relationship and Results factors for two reasons. First, because the purpose of the research was to determine if there is a difference between these factors in the perceptions of the rater groups.

Second, it can be argued that Relationship and Result are distinct constructs that happen to have a strong correlation. Effective leaders must achieve results through the efforts of others. It becomes necessary to develop a relationship with those others in order to exert influence upon their behaviors. Likewise positive results achieved by a workgroup will have a positive impact upon that group's perception of the leader and therefore result in a stronger perceived relationship.

### Conclusions

An attempt was made in this study to contribute to the knowledge of between-group differences by suggesting that the differences are the result of differences in observer perception based upon organization level and perception is influenced both by differences between rater groups and leader behaviors toward members of a group. It was further proposed that the organizational level of the observer influences the prototype or schema that the rater holds about leadership and that prototype influences how the leader is perceived and rated by the observer (Lord et al., 2001).

These findings agree that between-group raters do not agree on the dimensions of the latent variables used in the 360-degree feedback. More agreement was found in the correlated uniqueness terms. These facts are supported by previous research (e.g. Harris & Schaubroeck, 1988; Maurer et al., 1998; Mount et al., 1998; Shapira & Zevulun, 1989; Tsui & Ohlott, 1988) and the data analysis discussed here.

Rather than finding the differences between groups, a fundamental construct was found that forms the foundation for agreement. There seems to be agreement that Relationship and Results underlie the dimensions in leader feedback and that the two dimensions are highly correlated. This model does not attempt to identify the direction of that correlation and for good reason. Human relations are necessary to achieve results and good results, strengthen relationships.

Raters appear to agree on the second-order factors of Relationship and Results. The high percentage of variance explained reinforces the strength of the second-order factor loadings. Results and Relationship are important underlying factors in leadership feedback. It also appears that all rater groups see Relationships as an important contributor to Results.

Raters' perception of Relationship was found to be highly correlated to the perception of Results. The factor scores were found not to be significantly different within rater groups. The leader has different relationships with members of each group. Within group correlated uniqueness suggests a bias that tends to level the scores for the observed dimensions of leader performance. Results are a more objective measure but were not as well defined. It can be concluded that the relationships leaders develop with all organization members affect feedback ratings, workgroup results, and effective leadership performance.

The model fit statistics imply that generalizable conclusions can be made only with caution. The data suggest that there is some underlying factor that raters agree upon. Several reasons explain the moderate model fit . First, the

high correlations in the error factors within rater groups suggest method effects such as leniency bias, response pattern error, and halo effect. Second, model complexity has a negative impact upon goodness of fit statistics. Third, the instruments used in 360-degree feedback may elicit different responses from different rater groups. And fourth, there may be another second-order factor that the observed variables are attempting to measure. Given these issues the model fit is sufficient to encourage further research.

### Recommendations

The analysis presented here suggests the need for further research. Other models are possible and perhaps some model exists where there is a better fit to the data. Recognition of this fact presents several questions and recommendations.

How would separate instruments, designed specifically for each rater group, improve the factor scores on the latent variables? If the differences between groups are the result of perception and situational leadership behavior, then separate instruments may be in order.

Are there other observed variables that would better relate to the second-order factors of Relationship and Results? Perhaps different measures or methods would result in stronger factor loadings on the latent factors used to define the second-order factors.

Are there other second-order factors that would result in better model to data fit? Potentially other second-order factors exist. Identifying and testing those factors may strengthen the model.

Will other 360-degree instruments have similar relationships with the second-order factors of Results and Relationship? Other instruments, by their very nature, will measure different competencies and define different factors. Replicating this model with other instruments may improve model fit and identify the most appropriate competency measures.

How well do the second-order factors load on objective measures of leader-member relationship and workgroup results? Relationship would relate to citizenship behaviors of work-group members in adherence to rules and procedures, attendance, turnover, and altruism (Organ, 1997). Results would relate to workgroup performance measures of throughput, cost control, quality, and safety.

APPENDIX A  
Structural Equations

### Dependent latent variables

$$\eta_1 = \lambda_{1,1}\xi_1 + \zeta_1$$

$$\eta_1 = \lambda_{2,1}\xi_2 + \zeta_2$$

$$\eta_1 = \lambda_{3,1}\xi_3 + \zeta_3$$

$$\eta_1 = \lambda_{4,1}\xi_4 + \zeta_4$$

$$\eta_2 = \lambda_{9,2}\xi_9 + \zeta_9$$

$$\eta_2 = \lambda_{10,2}\xi_{10} + \zeta_{10}$$

$$\eta_2 = \lambda_{11,1}\xi_{11} + \zeta_{11}$$

$$\eta_2 = \lambda_{12,2}\xi_{12} + \zeta_{12}$$

### Independent variables

$$x_1 = \lambda_{1,1}\xi_1 + \lambda_{1,9}\xi_9 + \delta_1$$

$$x_2 = \lambda_{2,2}\xi_2 + \lambda_{2,9}\xi_9 + \delta_2$$

$$x_3 = \lambda_{3,3}\xi_3 + \lambda_{3,9}\xi_9 + \delta_3$$

$$x_4 = \lambda_{4,4}\xi_4 + \lambda_{4,9}\xi_9 + \delta_4$$

$$x_5 = \lambda_{5,5}\xi_5 + \lambda_{5,9}\xi_9 + \delta_5$$

$$x_6 = \lambda_{6,6}\xi_6 + \lambda_{6,9}\xi_9 + \delta_6$$

$$x_7 = \lambda_{7,7}\xi_7 + \lambda_{7,9}\xi_9 + \delta_7$$

$$x_8 = \lambda_{8,8}\xi_8 + \lambda_{8,9}\xi_9 + \delta_8$$

$$x_9 = \lambda_{9,1}\xi_1 + \lambda_{9,10}\xi_{10} + \delta_9$$

$$x_{10} = \lambda_{10,2}\xi_2 + \lambda_{10,10}\xi_{10} + \delta_{10}$$

$$x_{11} = \lambda_{11,3}\xi_3 + \lambda_{11,10}\xi_{10} + \delta_{11}$$

$$x_{12} = \lambda_{12,4}\xi_4 + \lambda_{12,10}\xi_{10} + \delta_{12}$$

$$x_{13} = \lambda_{13,5}\xi_5 + \lambda_{13,10}\xi_{10} + \delta_{13}$$

$$x_{14} = \lambda_{14,6}\xi_6 + \lambda_{14,10}\xi_{10} + \delta_{14}$$

$$x_{15} = \lambda_{15,7}\xi_7 + \lambda_{15,10}\xi_{10} + \delta_{15}$$

$$x_{16} = \lambda_{16,8}\xi_8 + \lambda_{16,10}\xi_{10} + \delta_{16}$$

$$x_{17} = \lambda_{17,1}\xi_1 + \lambda_{17,11}\xi_{11} + \delta_{17}$$

$$x_{18} = \lambda_{18,2}\xi_2 + \lambda_{18,11}\xi_{11} + \delta_{18}$$

$$x_{19} = \lambda_{19,3}\xi_3 + \lambda_{19,11}\xi_{11} + \delta_{19}$$

$$x_{20} = \lambda_{20,4}\xi_4 + \lambda_{20,11}\xi_{11} + \delta_{20}$$

$$x_{21} = \lambda_{21,5}\xi_5 + \lambda_{21,11}\xi_{11} + \delta_{21}$$

$$x_{22} = \lambda_{22,6}\xi_6 + \lambda_{22,11}\xi_{11} + \delta_{22}$$

$$x_{23} = \lambda_{23,7}\xi_7 + \lambda_{23,11}\xi_{11} + \delta_{23}$$

$$x_{24} = \lambda_{24,8}\xi_8 + \lambda_{24,11}\xi_{11} + \delta_{24}$$

$$x_{25} = \lambda_{25,1}\xi_1 + \lambda_{25,12}\xi_{12} + \delta_{25}$$

$$x_{26} = \lambda_{26,2}\xi_2 + \lambda_{26,12}\xi_{12} + \delta_{26}$$

$$x_{27} = \lambda_{27,3}\xi_3 + \lambda_{28,12}\xi_{12} + \delta_{27}$$

$$x_{28} = \lambda_{28,4}\xi_4 + \lambda_{28,12}\xi_{12} + \delta_{28}$$

$$x_{29} = \lambda_{29,5}\xi_5 + \lambda_{29,12}\xi_{12} + \delta_{29}$$

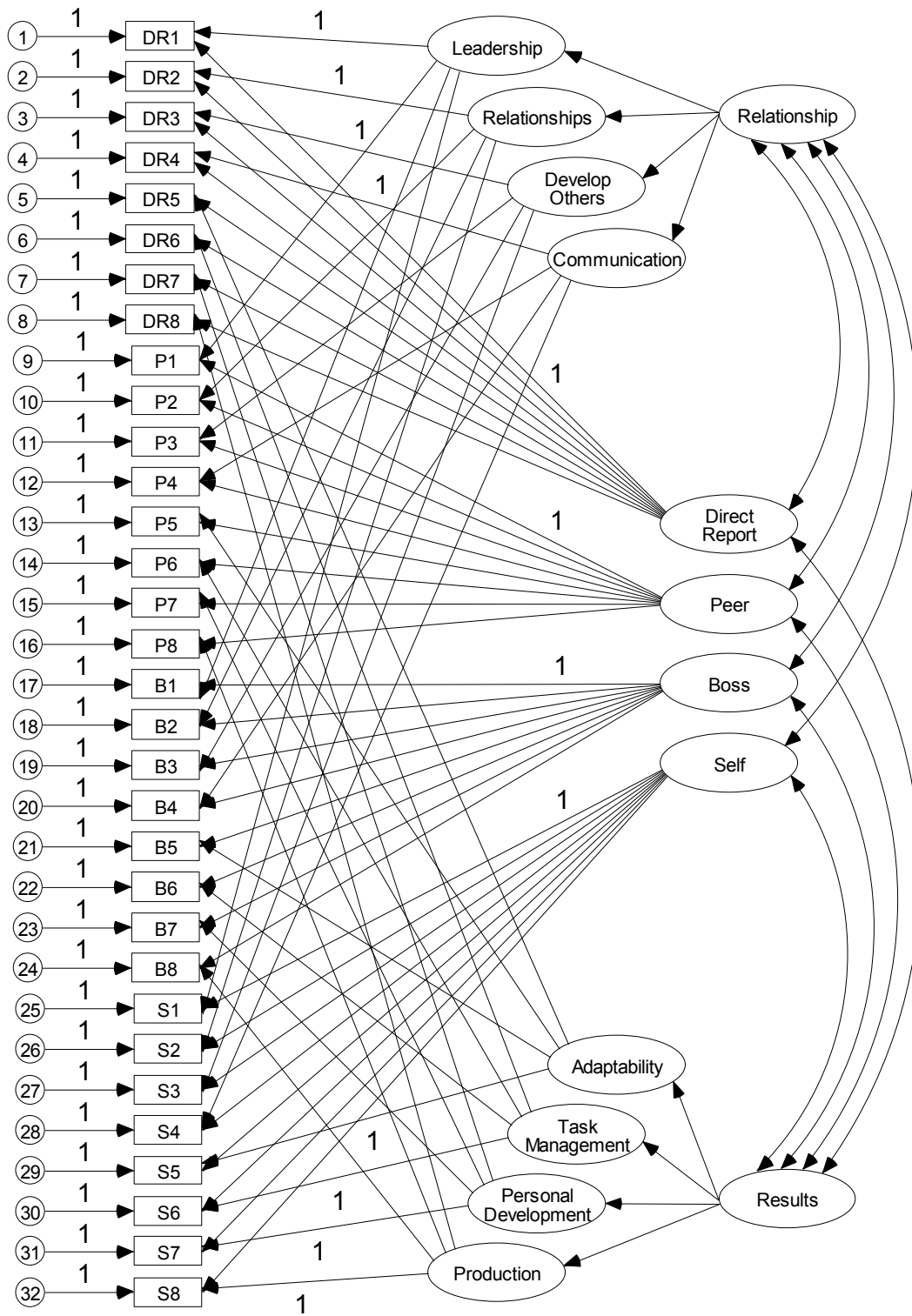
$$x_{30} = \lambda_{30,6}\xi_6 + \lambda_{30,12}\xi_{12} + \delta_{30}$$

$$x_{31} = \lambda_{31,7}\xi_7 + \lambda_{31,12}\xi_{12} + \delta_{31}$$

$$x_{32} = \lambda_{32,8}\xi_8 + \lambda_{32,12}\xi_{12} + \delta_{32}$$

## APPENDIX B

### Model of Rater Perception



## APPENDIX C

### Structural Equation Programs

Revised Second-order Factor Program

Observed variables: S1 S2 S3 S4 S5 S6 S7 S8 B1 B2 B3 B4 B5 B6 B7 B8 DR1

DR2 DR3 DR4 DR5 DR6 DR7 DR8 P1 P2 P3 P4 P5 P6 P7 P8

Correlation Matrix from file: Sample1.cor

Means from file: Sample1.mn

Standard Deviations from file: Sample1.sd

Sample size: 6021

Latent variables: SELFREL BOSSREL DRREL PEEREL Relation SELFRES

BOSSRES DRRES PEERES Results

Relationships:

DR1 = 1\*DRREL

DR2 DR4 DR7 = DRREL

P1 = 1\*PEEREL

P2 P4 P7 = PEEREL

B1 = 1\*BOSSREL

B2 B4 B7 = BOSSREL

S1 = 1\*SELFREL

S2 S4 S7 = SELFREL

DR3 = 1\*DRRES

DR5 DR6 DR8 = DRRES

P3 = 1\*PEERES

P5 P6 P8 = PEERES

B3 = 1\*BOSSRES

B5 B6 B8 = BOSSRES

S3 = 1\*SELFRES

S5 S6 S8 = SELFRES

Paths:

Relation -> DRREL PEEREL BOSSREL SELFREL

Results -> DRRES PEERES BOSSRES SELFRES

Set the variance of Relation equal to 1.0

Set the variance of Results equal to 1.0

Set the error variance of DRREL to .001

Set the error variance of PEEREL to .001

Set the error variance of BOSSREL to .001

Set the error variance of SELFREL to .001

Set the error variance of DRRES to .001

Set the error variance of PEERES to .001

Set the error variance of BOSSRES to .001

Set the error variance of SELFRES to .001

Options: SC AD=150

Lisrel output:

End of problem

Modified Second-order Factor Program

Observed variables: S1 S2 S3 S4 S5 S6 S7 S8 B1 B2 B3 B4 B5 B6 B7 B8 DR1

DR2 DR3 DR4 DR5 DR6 DR7 DR8 P1 P2 P3 P4 P5 P6 P7 P8

Correlation Matrix from file: *filename*

Means from file: *filename*

Standard Deviations from file: *filename*

Sample size: 6021

Latent variables: SELFREL BOSSREL DRREL PEEREL Relation SELFRES

BOSSRES DRRES PEERES Results

Relationships:

DR1 = 1\*DRREL

DR2 DR4 DR7 = DRREL

P1 = 1\*PEEREL

P2 P4 P7 = PEEREL

B1 = 1\*BOSSREL

B2 B4 B7 = BOSSREL

S1 = 1\*SELFREL

S2 S4 S7 = SELFREL

DR3 = 1\*DRRES

DR5 DR6 DR8 = DRRES

P3 = 1\*PEERES

P5 P6 P8 = PEERES

B3 = 1\*BOSSRES

B5 B6 B8 = BOSSRES

S3 = 1\*SELFRES

S5 S6 S8 = SELFRES

Paths:

Relation -> DRREL PEEREL BOSSREL SELFREL

Results -> DRRES PEERES BOSSRES SELFRES

Let the Errors between S1 - S8 Correlate

Let the Errors between B1 - B8 Correlate

Let the Errors between DR1 - DR8 Correlate

Let the Errors between P1 - P8 Correlate

Set the variance of Relation equal to 1.0

Set the variance of Results equal to 1.0

Set the error variance of DRREL to .001

Set the error variance of PEEREL to .001

Set the error variance of BOSSREL to .001

Set the error variance of SELFREL to .001

Set the error variance of DRRES to .001

Set the error variance of PEERES to .001

Set the error variance of BOSSRES to .001

Set the error variance of SELFRES to .001

Options: SC AD=150 ND=3

Lisrel output:

End of problem

## Factor Program for Subordinate Relationship

Observed variables: S1 S2 S3 S4 S5 S6 S7 S8 B1 B2 B3 B4 B5 B6 B7 B8 DR1

DR2 DR3 DR4 DR5 DR6 DR7 DR8 P1 P2 P3 P4 P5 P6 P7 P8

Correlation Matrix from file: *filename*

Means from file: *filename*

Standard Deviations from file: *filename*

Sample size: 400

Latent variables: SELFREL BOSSREL DRREL PEEREL Relation SELFRES

BOSSRES DRRES PEERES Results

Relationships:

DR1 = 1\*DRREL

DR2 DR4 DR7 = DRREL

Options: SC AD=150 ND=3

Lisrel output:

End of problem

Factor Program for Subordinate Results

Observed variables: S1 S2 S3 S4 S5 S6 S7 S8 B1 B2 B3 B4 B5 B6 B7 B8 DR1

DR2 DR3 DR4 DR5 DR6 DR7 DR8 P1 P2 P3 P4 P5 P6 P7 P8

Correlation Matrix from file: *filename*

Means from file: *filename*

Standard Deviations from file: *filename*

Sample size: 6021

Latent variables: SELFREL BOSSREL DRREL PEEREL Relation SELFRES

BOSSRES DRRES PEERES Results

Relationships:

DR3 = 1\*DRRES

DR5 DR6 DR8 = DRRES

Options: SC AD=150 ND=3

Lisrel output:

End of problem

## Relationship Second-order Program

Observed variables: S1 S2 S3 S4 S5 S6 S7 S8 B1 B2 B3 B4 B5 B6 B7 B8 DR1

DR2 DR3 DR4 DR5 DR6 DR7 DR8 P1 P2 P3 P4 P5 P6 P7 P8

Correlation Matrix from file: *filename*

Means from file: *filename*

Standard Deviations from file: *filename*

Sample size: 400

Latent variables: SELFREL BOSSREL DRREL PEEREL Relation SELFRES

BOSSRES DRRES PEERES Results

Relationships:

DR1 = 1\*DRREL

DR2 DR4 DR7 = DRREL

P1 = 1\*PEEREL

P2 P4 P7 = PEEREL

B1 = 1\*BOSSREL

B2 B4 B7 = BOSSREL

S1 = 1\*SELFREL

S2 S4 S7 = SELFREL

Paths:

Relation -> DRREL PEEREL BOSSREL SELFREL

Let the Errors between S1 S2 S4 S7 Correlate

Let the Errors between S2 S4 S7 Correlate

Let the Errors between S4 S7 Correlate

Let the Errors between B1 B2 B4 B7 Correlate

Let the Errors between B2 B4 B7 Correlate

Let the Errors between B4 B7 Correlate

Let the Errors between DR1 DR2 DR4 DR7 Correlate

Let the Errors between DR2 DR4 DR7 Correlate

Let the Errors between DR4 DR7 Correlate

Let the Errors between P1 P2 P4 P7 Correlate

Let the Errors between P2 P4 P7 Correlate

Let the Errors between P4 P7 Correlate

Set the variance of Results equal to 1.0

Set the error variance of DRREL to .001

Set the error variance of PEEREL to .001

Set the error variance of BOSSREL to .001

Set the error variance of SELFREL to .001

Options: SC AD=150 ND=3

Lisrel output:

End of problem

## Results Second-order Program

Observed variables: S1 S2 S3 S4 S5 S6 S7 S8 B1 B2 B3 B4 B5 B6 B7 B8 DR1

DR2 DR3 DR4 DR5 DR6 DR7 DR8 P1 P2 P3 P4 P5 P6 P7 P8

Correlation Matrix from file: *filename*

Means from file: *filename*

Standard Deviations from file: *filename*

Sample size: 6021

Latent variables: SELFREL BOSSREL DRREL PEEREL Relation SELFRES

BOSSRES DRRES PEERES Results

Relationships:

DR3 = 1\*DRRES

DR5 DR6 DR8 = DRRES

P3 = 1\*PEERES

P5 P6 P8 = PEERES

B3 = 1\*BOSSRES

B5 B6 B8 = BOSSRES

S3 = 1\*SELFRES

S5 S6 S8 = SELFRES

Paths:

Results -> DRRES PEERES BOSSRES SELFRES

Let the Errors between S3 S5 S6 S8 Correlate

Let the Errors between S5 S6 S8 Correlate

Let the Errors between S6 S8 Correlate

Let the Errors between B3 B5 B6 B8 Correlate

Let the Errors between B5 B6 B8 Correlate

Let the Errors between B6 B8 Correlate

Let the Errors between DR3 DR5 DR6 DR8 Correlate

Let the Errors between DR5 DR6 DR8 Correlate

Let the Errors between DR6 DR8 Correlate

Let the Errors between P3 P5 P6 P8 Correlate

Let the Errors between P5 P6 P8 Correlate

Let the Errors between P6 P8 Correlate

Set the variance of Results equal to 1.0

Set the error variance of DRRES to .001

Set the error variance of PEERES to .001

Set the error variance of BOSSRES to .001

Set the error variance of SELFRES to .001

Options: SC AD=150 ND=3

Lisrel output:

End of problem

Appendix D  
Multitrait-Multirater Matrix

Raters/Traits	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
<b>Self</b>																
1 Communication	<b>0.84</b>															
2 Leadership	0.75	<b>0.85</b>														
3 Adaptation	0.66	0.64	<b>0.83</b>													
4 Relationship	0.75	0.73	0.66	<b>0.82</b>												
5 Task	0.65	0.67	0.62	0.60	<b>0.79</b>											
6 Production	0.64	0.69	0.68	0.63	0.71	<b>0.86</b>										
7 Develop Others	0.65	0.65	0.56	0.72	0.56	0.56	<b>0.82</b>									
8 Personal Development	0.64	0.63	0.66	0.68	0.61	0.65	0.64	<b>0.79</b>								
<b>Boss</b>																
9 Communication	<del>0.17</del>	<del>0.11</del>	<del>0.09</del>	<del>0.14</del>	<del>0.09</del>	<del>0.09</del>	<del>0.09</del>	<del>0.10</del>	<b>0.89</b>							
10 Leadership	<del>0.15</del>	<del>0.16</del>	<del>0.09</del>	<del>0.14</del>	<del>0.13</del>	<del>0.15</del>	<del>0.10</del>	<del>0.10</del>	0.79	<b>0.88</b>						
11 Adaptation	<del>0.14</del>	<del>0.11</del>	<del>0.22</del>	<del>0.13</del>	<del>0.12</del>	<del>0.17</del>	<del>0.10</del>	<del>0.15</del>	0.74	0.70	<b>0.87</b>					
12 Relationship	<del>0.15</del>	<del>0.11</del>	<del>0.10</del>	<del>0.19</del>	<del>0.05</del>	<del>0.07</del>	<del>0.14</del>	<del>0.12</del>	0.79	0.77	0.73	<b>0.86</b>				
13 Task	<del>0.12</del>	<del>0.12</del>	<del>0.12</del>	<del>0.08</del>	<del>0.22</del>	<del>0.18</del>	<del>0.06</del>	<del>0.09</del>	0.68	0.71	0.69	0.60	<b>0.83</b>			
14 Production	<del>0.11</del>	<del>0.13</del>	<del>0.13</del>	<del>0.09</del>	<del>0.16</del>	<del>0.23</del>	<del>0.07</del>	<del>0.10</del>	0.67	0.73	0.71	0.62	0.79	<b>0.90</b>		
15 Develop Others	<del>0.12</del>	<del>0.11</del>	<del>0.10</del>	<del>0.17</del>	<del>0.07</del>	<del>0.10</del>	<del>0.17</del>	<del>0.13</del>	0.69	0.71	0.66	0.79	0.58	0.59	<b>0.86</b>	
16 Personal Development	<del>0.10</del>	<del>0.07</del>	<del>0.12</del>	<del>0.11</del>	<del>0.08</del>	<del>0.11</del>	<del>0.08</del>	<del>0.16</del>	0.71	0.69	0.73	0.72	0.67	0.70	0.69	<b>0.84</b>
<b>Subordinates</b>																
17 Communication	<del>0.16</del>	<del>0.11</del>	<del>0.07</del>	<del>0.19</del>	<del>0.07</del>	<del>0.07</del>	<del>0.15</del>	<del>0.05</del>	<del>0.35</del>	<del>0.31</del>	<del>0.25</del>	<del>0.36</del>	<del>0.24</del>	<del>0.22</del>	<del>0.35</del>	<del>0.24</del>
18 Leadership	<del>0.14</del>	<del>0.15</del>	<del>0.07</del>	<del>0.17</del>	<del>0.08</del>	<del>0.10</del>	<del>0.14</del>	<del>0.05</del>	0.32	0.33	0.25	0.35	0.25	0.25	0.35	0.24
19 Adaptation	<del>0.13</del>	<del>0.12</del>	<del>0.19</del>	<del>0.17</del>	<del>0.09</del>	<del>0.14</del>	<del>0.14</del>	<del>0.10</del>	0.30	0.29	0.34	0.33	0.25	0.26	0.35	0.26
20 Relationship	<del>0.15</del>	<del>0.12</del>	<del>0.09</del>	<del>0.22</del>	<del>0.04</del>	<del>0.06</del>	<del>0.17</del>	<del>0.07</del>	0.31	0.29	0.25	0.38	0.20	0.19	0.36	0.24
21 Task	<del>0.12</del>	<del>0.12</del>	<del>0.10</del>	<del>0.12</del>	<del>0.17</del>	<del>0.14</del>	<del>0.11</del>	<del>0.06</del>	0.32	0.32	0.28	0.30	0.36	0.31	0.32	0.25
22 Production	<del>0.12</del>	<del>0.14</del>	<del>0.11</del>	<del>0.13</del>	<del>0.13</del>	<del>0.18</del>	<del>0.12</del>	<del>0.06</del>	0.30	0.32	0.29	0.30	0.31	0.34	0.33	0.25
23 Develop Others	<del>0.14</del>	<del>0.14</del>	<del>0.09</del>	<del>0.20</del>	<del>0.06</del>	<del>0.08</del>	<del>0.24</del>	<del>0.09</del>	0.28	0.28	0.24	0.34	0.19	0.20	0.38	0.24
24 Personal Development	<del>0.12</del>	<del>0.10</del>	<del>0.12</del>	<del>0.18</del>	<del>0.05</del>	<del>0.09</del>	<del>0.15</del>	<del>0.12</del>	0.31	0.29	0.30	0.36	0.24	0.25	0.37	0.37
<b>Peers</b>																
25 Communication	<del>0.20</del>	<del>0.12</del>	<del>0.07</del>	<del>0.19</del>	<del>0.08</del>	<del>0.08</del>	<del>0.15</del>	<del>0.09</del>	<del>0.45</del>	<del>0.37</del>	<del>0.33</del>	<del>0.43</del>	<del>0.32</del>	<del>0.29</del>	<del>0.38</del>	<del>0.32</del>
26 Leadership	<del>0.17</del>	<del>0.16</del>	<del>0.06</del>	<del>0.17</del>	<del>0.10</del>	<del>0.11</del>	<del>0.13</del>	<del>0.09</del>	0.40	0.41	0.31	0.41	0.35	0.33	0.37	0.33
27 Adaptation	<del>0.18</del>	<del>0.14</del>	<del>0.22</del>	<del>0.19</del>	<del>0.12</del>	<del>0.17</del>	<del>0.15</del>	<del>0.15</del>	0.38	0.35	0.43	0.39	0.34	0.34	0.36	0.35
28 Relationship	<del>0.18</del>	<del>0.12</del>	<del>0.08</del>	<del>0.24</del>	<del>0.05</del>	<del>0.06</del>	<del>0.18</del>	<del>0.11</del>	0.40	0.35	0.31	0.47	0.25	0.24	0.40	0.33
29 Task	<del>0.15</del>	<del>0.13</del>	<del>0.10</del>	<del>0.12</del>	<del>0.20</del>	<del>0.17</del>	<del>0.08</del>	<del>0.08</del>	0.37	0.38	0.34	0.32	0.46	0.40	0.32	0.32
30 Production	<del>0.15</del>	<del>0.15</del>	<del>0.14</del>	<del>0.14</del>	<del>0.15</del>	<del>0.22</del>	<del>0.11</del>	<del>0.10</del>	0.37	0.39	0.37	0.34	0.41	0.45	0.33	0.33
31 Develop Others	<del>0.16</del>	<del>0.13</del>	<del>0.09</del>	<del>0.22</del>	<del>0.05</del>	<del>0.08</del>	<del>0.22</del>	<del>0.11</del>	0.37	0.35	0.31	0.43	0.25	0.26	0.44	0.32
32 Personal Development	<del>0.15</del>	<del>0.10</del>	<del>0.12</del>	<del>0.18</del>	<del>0.07</del>	<del>0.11</del>	<del>0.14</del>	<del>0.17</del>	0.38	0.35	0.36	0.41	0.31	0.31	0.39	0.41

Raters/Traits	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
<b>Subordinates</b>																
17 Communication	<b>0.93</b>															
18 Leadership	0.91	<b>0.92</b>														
19 Adaptation	0.85	0.86	<b>0.90</b>													
20 Relationship	0.90	0.90	0.87	<b>0.91</b>												
21 Task	0.82	0.84	0.83	0.77	<b>0.88</b>											
22 Production	0.83	0.86	0.85	0.79	0.89	<b>0.92</b>										
23 Develop Others	0.85	0.85	0.82	0.88	0.74	0.77	<b>0.89</b>									
24 Personal Development	0.86	0.86	0.87	0.88	0.80	0.82	0.86	<b>0.85</b>								
<b>Peers</b>																
25 Communication	0.47	0.42	0.39	0.42	0.40	0.38	0.40	0.40	<b>0.91</b>							
26 Leadership	0.44	0.44	0.38	0.41	0.41	0.40	0.40	0.39	0.87	<b>0.91</b>						
27 Adaptation	0.39	0.36	0.46	0.38	0.38	0.39	0.37	0.41	0.82	0.78	<b>0.90</b>					
28 Relationship	0.45	0.41	0.40	0.47	0.35	0.34	0.43	0.43	0.88	0.85	0.80	<b>0.89</b>				
29 Task	0.38	0.36	0.37	0.32	0.46	0.42	0.32	0.34	0.77	0.80	0.77	0.69	<b>0.88</b>			
30 Production	0.37	0.37	0.39	0.32	0.43	0.46	0.33	0.36	0.77	0.81	0.81	0.71	0.86	<b>0.92</b>		
31 Develop Others	0.45	0.42	0.41	0.46	0.36	0.37	0.47	0.44	0.81	0.80	0.77	0.87	0.67	0.70	<b>0.88</b>	
32 Personal Development	0.38	0.36	0.38	0.38	0.36	0.36	0.37	0.43	0.81	0.80	0.83	0.84	0.75	0.77	0.81	<b>0.84</b>

Notes:  
 Monotrait-monomethod values indicate reliability. (shown in bold)  
 Monotrait-heteromethod values indicate convergent validity. (shown in italics)  
 Monotrait-heteromethod triangle correlates traits within rater group. (solid triangle)  
 Heterotrait-heteromethod triangle correlates traits between groups. (dashed triangle)

## Appendix E

Permssion to use Copyright Material.



International, Inc., The Assessment Co.

January 28, 2004

Thomas Bradley  
7 Greenhollow  
Wichita Falls, TX 76308

Dear Thomas,


Profiles International, Inc. hereby grants to Thomas Bradley the right to use the trademarks and copyrighted materials from the CheckPoint 3600 Competency Feedback System™ for the specific purposes of research pertaining to his dissertation for the degree of Doctor of Philosophy at the University of North Texas.

The copyrighted materials include rating data, examples of the questions in the text of the dissertation and a sample of the questionnaire in the, appendix.

Also, the trademarked names of Profiles International and CheckPoint 3600 Competency Feedback System™ may be used in reference to the instrument and data used in the dissertation research.

Limitations to the use of the data are governed by Profiles International's agreements with clients to maintain confidentiality. Limits also apply to the security of the CheckPoint 3600 instrument pertaining to the business of Profiles International.

Prior written permission is required before the trademarked names and copyrighted materials of Profiles International, Inc. are used for any publication.

Sincerely,  
  
Scott F. Haney, Esq.  
Corporate Counsel  
Profiles International, Inc

Always Seeking Excellence

5205 LAKE SHORE DR. . WACO, TX 76710 . 254.751.1644 . FAX 254.772.8155 . [www.profilesinternational.com](http://www.profilesinternational.com)

## REFERENCES

- Antonioni, D., & Park, H. (2001). The relationship between rater affect and three sources of 360-degree feedback ratings. *Journal of Management, 27*, 479-495.
- Arbuckle, J. (2003). Amos 5 [Computer software]. Chicago: SmallWaters Corp.
- Atkins, P. W., & Wood, R. E. (2002). Self-versus others' ratings as predictors of assessment center ratings: Validation evidence for 360-degree feedback programs. *Personnel Psychology, 55*(4), 871-905.
- Atwater, L. E., Ostroff, C., Yammarino, F. J., & Fleenor, J. W. (1998). Self-other agreement: Does it really matter? *Personnel Psychology, 51*, 577-598.
- Atwater, L., Roush, P., & Fischthal, A. (1995). The influence on upward feedback on self and follower ratings of leadership. *Personnel Psychology, 48*, 35-59.
- Avolio, B. J., & Bass, B. M. (1988). Transformational leadership, charisma and beyond. In J. G. Hunt, B. R. Baliga, H. P. Dachler, & C. A. Schriesheim (Eds.), *Emerging leadership vistas* (pp. 29-49). Lexington, MA: DC Heath.
- Bandalos, B. (1996). Confirmatory factor analysis. In J. Stevens (Ed.), *Applied multivariate statistics for the social sciences* (3rd ed., pp. 389-420). Mahwah, NJ: Erlbaum.
- Becker, T. E., & Cote, J. A. (1994). Additive and multiplicative method effects on applied psychological research: An empirical assessment of three models. *Journal of Management, 20*(3), 625-641.

- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606.
- Borman, W. C. (1974). The rating of individuals in organizations: An alternative approach. *Organizational Behavior and Human Performance*, 15, 105-124.
- Bradley, T. P., & Knox, D. W. (2003, February). *An investigation of confirmatory factor analysis*. Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio, TX.
- Brutus, S., & Fleenor, J. W. (1998). Does 360-degree feedback work in different organizations? A between industry comparison. *Journal of Management Development*, 17(2/3), 177-191.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81-105.
- Cellar, D. F., Sidle, S., Gourdy, K., & O'Brien, D. (2001). Effects of leader style, leader sex, and subordinate personality on leader evaluations and future subordinate motivation. *Journal of Business and Psychology*, 16(1), 61-72.
- Chemers, M. M. (1997). *An integrative theory of leadership*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Conway, J. M. (1996). Analysis and design of multitrait-multirater performance appraisal studies. *Journal of Management*, 22(1), 139-162.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Fort Worth TX: Harcourt Brace Jovanovich.

- DeNisi, A. S., & Kluger, N. A. (2000). Feedback effectiveness: Can 360-degree appraisals be improved? *Academy of Management Executive*, 14(1), 129-139.
- Fiedler, F. E. (1978). The contingency model and the dynamics of the leadership process. In L. Berowitz (Ed.), *Advances in experimental social psychology* (New York: Academic Press).
- Gall, M. D., Borg, W. R., & Gall, J. P. (1996). *Educational research: An introduction* (6th. ed.). White Plains, NY: Longman.
- Gerstner, C. R., & Day, D. V. (1997). Meta-analytic review of leader-member exchange theory: Correlates and construct issues. *Journal of Applied Psychology*, 82(6), 827-844.
- Harris, M. M., & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology*, 41, 43-62.
- Hersey, P., Blanchard, K. H., & Johnson, D. E. (2001). *Management of organizational behavior: Leading human resources*. Upper Saddle River, NJ: Prentice-Hall.
- Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1998). *Applied statistics for the behavioral sciences*. Boston: Houghton Mufflin.
- House, R. J. (1971). A path goal theory of leader effectiveness. *Administrative Sciences Quarterly*, 16, 321-338.
- Howell, J. M., & Hall-Merenda K. E. (1999). The ties that bind: The impact of leader-member exchange, transformational and transactional leadership,

- and distance on predicting follower performance. *Journal of Applied Psychology*, 84(5), 680-694.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1-56.
- Johnson, J. W., & Ferstl, K. L. (1999). The effects of interrater and self-other agreement on performance improvement following upward feedback. *Personnel Psychology*, 52, 271-303.
- Jöreskog, K. G. (1979a). Basic ideas of factor and component analysis. In K. G. Jöreskog and D. Sörbom (Eds.), *Advances in factor analysis and structural equation models* (pp.21-44). Cambridge, MA: Abt Books.
- Jöreskog, K. G. (1979b). A general approach to confirmatory maximum likelihood factor analysis with addendum. In K. G. Jöreskog & D. Sörbom, *Advances in factor analysis and structural equation models*. (pp.21-44). Cambridge, MA: Abt Books
- Jöreskog, K. G., & Sörbom, D. (1993). *Lisrel 8: Structural equation modeling with the Simplis command language*. Hillsdale NJ: Lawrence Erlbaum Associates.
- Jöreskog, K. G., & Sörbom, D. (2000a). Lisrel 8.50 [Computer software]. Lincolnwood, IL: Scientific Software International.
- Jöreskog, K. G., & Sörbom D. (2000b). Prelis 2.50 [Computer software]. Lincolnwood, IL: Scientific Software International.

- Kieffer, K. M. (1999). An introductory primer on the appropriate use of exploratory and confirmatory factor analysis. *Research in the Schools, 5* (2), 75-92.
- Kluger, N. A., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*, 254-284.
- Lawler, E. E. (1967). The multitrait-multirater approach to measuring managerial job performance. *Journal of Applied Psychology, 51*(5), 369-381.
- Lord, R. G., Brown, D. J., & Freiberg, S. J. (1999). Understanding the dynamics of leadership: The role of follower self-concepts in the leader/follower relationship. *Organizational Behavior and Human Decision Processes, 78*(3), 167-203.
- Lord, R. G., Brown, D. J., Harvey, J. L., & Hall, R. J. (2001). Contextual constraints on prototype generation and their multilevel consequences for leadership perceptions. *Leadership Quarterly, 12*, 311-338.
- Lord, R. G., & Maher, K. J., (1991). *Leadership and information processing: Linking perceptions and performance*. London: Routledge.
- Maurer, T. J., Mitchell, R. D., & Barbeite, F. G. (2002). Predictors of attitude toward 360-degree feedback system and involvement in post-feedback management development activity. *Journal of Occupational and Organizational Psychology, 75*, 87-107.
- Maurer, T. J., Raju, N. S., & Collins, W. C. (1998). Peer and subordinate performance measurement equivalence. *Journal of Applied Psychology, 83*(5), 693-702.

- Meindl, J. R. (1995). The romance of leadership as a follower-centric theory: A social constructionist approach. *Leadership Quarterly*, 6(3), 329-341.
- Mount, M. K., Judge, T. A., Scullen, S. E., Sytsma, M. R., & Hezlett, S. A. (1998). Trait, rater and level effects in 360-degree performance ratings. *Personnel Psychology*, 51, 557-576.
- Nickols, F. W. (1989). Making work productive. In D. B. Gradous (Ed.), *Systems theory applied to human resource development: Theory to practice monograph* (pp 91-116). Alexandria, VA: American Society of Training and Development.
- Organ, D. W. (1997). Organizational citizenship behavior: It's construct clean-up time. *Human Performance*, 10(2), 85-97.
- Penny, J. (2001). Differential item functioning an international 360-degree assessment: Evidence of gender stereotype, environmental complexity, and organizational contingency. *European Journal of Work and Organizational Psychology*, 10 (3), 245-271.
- Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research. *Journal of Management*, 12, 531-544.
- Profiles International. (2001). *Profiles CheckPoint 360° Competency Feedback System™: History, development, and research information*. (Available from Profiles International, 5205 Lakeshore Drive, Waco, TX, 76710)
- Pulakos, E. D., Schmitt, N., & Chan, D. (1996). Models of job performance ratings: An examination of leader race, leader gender, and rater level effects. *Human Performance*, 9(2), 103-119.

- Rosenkrantz, S. A., Lathans, F., & Hennessey, H. W. (1983). Role conflict and ambiguity scales: An evaluation of psychometric properties and the role of social desirability response bias. *Educational and Psychological Measurement, 43*, 957-970.
- Rossett, A. (1999). Analysis for human performance technology. In H. D. Stolovitch & E. J. Keeps (Eds.), *Handbook of human performance technology: Improving individual and organizational performance worldwide*. (2nd ed., pp 139-162). San Francisco: Jossey-Bass Pfeiffer.
- Rubin, D. B., & Thayer, D. T. (1982). EM algorithms for ML factor analysis. *Psychometrika, 47*(1), 69-76.
- Schmitt, N., & Stults, D. M. (1986). Methodology review: Analysis of multitrait-multimethod matrices. *Applied Psychological Measurement, 10*(1), 1-22.
- Schumacker, R. E., & Lomax, R. G. (1996). A beginner's guide to structural equation modeling. Mahwah NJ: Lawrence Erlbaum Associates.
- Shapira, Z., & Zevulun, E. (1989). Performance ratings in organizations: A facet analysis interpretation of the multitrait-multirater approach. *Multivariate Behavioral Research, 24*(2), 209-232.
- Smith, P. C. (1976). Behaviors, Results and organizational effectiveness: The problem of the criterion. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational behavior*. Chicago: Rand McNally.
- Spector, P. E. (1987). Method variance as an artifact on self-reported affect and perceptions at work: Myth or significant problem? *Journal of Applied Psychology, 72*, 438-443.

- SPSS for Windows 8.0.0 (Standard Version) [Computer software]. (1997).  
Chicago: SPSS Inc.
- Toegel, G., & Conger, J. A. (2003). 360-degree assessment: Time for reinvention. *Academy of Management Learning & Education*, 2(3), 297-311.
- Tsui, A. S., & Ohlott, P. (1988). Multiple assessment of managerial effectiveness: Interrater agreement and consensus in effectiveness models. *Personnel Psychology*, 41, 779-803.
- Ulrich, D., Zenger, J., & Smallwood, N. (1999). *Results-based leadership*. Boston: Harvard Business School Press.
- Walker, A. G., & Smither, J. W. (1999). A five-year study of upward feedback: What managers do with their Results matters. *Personnel Psychology*, 52, 393-423.
- Wothke, W. (1996). Models for multitrait-multirater matrix analysis. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 7-56). Mahwah, NJ: Lawrence Erlbaum Associates.
- Wothke, W., & Browne, M. W. (1990). The direct product model for the MTMM matrix parameterized as a second-order factor analysis model. *Psychometrika*, 55(2), 255-262.
- Yukl, G., & Tracey, J. B. (1992). Consequences of influence tactics used with subordinates, peers, and the boss. *Journal of Applied Psychology*, 77(4), 525-535.

Zazanis, M. M., Zaccaro, S. J., & Kilcullen, R. N. (2001). Identifying motivation and interpersonal performance using peer evaluations. *Military Psychology, 13*(2), 73-88.